Information Spectrum Approach to Second-Order Coding Rate in Channel Coding

Masahito Hayashi

***Abstract*—In this paper, second-order coding rate of channel coding is discussed for general sequence of channels. The optimum second-order transmission rate with a constant error constraint ϵ is obtained by using the information spectrum method. We apply this result to the discrete memoryless case, the discrete memoryless case with a cost constraint, the additive Markovian case, and the Gaussian channel case with an energy constraint. We also clarify that the Gallager bound does not give the optimum evaluation in the second-order coding rate.**

***Index Terms*—Additive Markovian channel, central limit theorem, channel coding, Gallager bound, information spectrum, second-order coding rate.**

I. Introduction

BASED on the channel coding theorem, there exists a sequence of codes for the given channel W such that the average error probability goes to 0 when the transmission rate R is less than C_W^{DM}. That is, if the number n of applications of the channel W is sufficiently large, the average error probability of a good code goes to 0. In order to evaluate the average error probability with finite n, we often use the exponential rate of decrease, which depends on the transmission rate R. However, such an exponential evaluation ignores the constant factor. Therefore, it is not clear whether exponential evaluation provides a good evaluation for the average error probability when the transmission rate R is close to the capacity. In fact, many researchers believe that, out of the known evaluations, the Gallager bound [1] gives the best upper bound of average error probability in the channel coding when the transmission rate is greater than the critical rate. This is because the Gallager bound provides the optimal exponential rate of decrease. In order to clarify this point, we focus on the second-order coding rate in channel coding, in which we describe the transmission length by $C_W^{\mathrm{DM}}n + R_2\sqrt{n}$. From a practical viewpoint, when the coding length is close to $C_W^{\mathrm{DM}}n$, the second-order coding rate gives a better evaluation of average error probability than the first-order coding rate. In fact, the second error coding rate has been applied for evaluation of the average error probability of random coding concerning the phase basis, which is essential to the security of quantum key distribution [2]. Therefore, it is appropriate to treat the second-order coding rate from the applied viewpoint as well as the theoretical viewpoint. In the case of the discrete memoryless case, Strassen [3] derived the optimum rate R_2 for an arbitrary average error probability $0 < \epsilon < 1$ using the Gaussian distribution. In this paper, we extend his result to more general cases, i.e., the discrete memoryless case with cost constraint, the Gaussian additive noise case with the energy constraint, and the additive Markovian case. Further, our proof for the discrete memoryless case is much simpler than the original one. Indeed, Strassen's proof is difficult to follow because his proof is not so simple. In this paper, in order to treat this problem from a unified viewpoint, we employ the method of information spectrum, which was initiated by Han and Verdú [4], and was mainly formulated by Han [5]. The second-order coding rate is closely related to the method of information spectrum because Hayashi [6] treated this problem of fixed-length source coding and intrinsic randomness using the method of information spectrum. Hayashi [6] discussed the error probability when the compressed size is $H(P)n + a\sqrt{n}$, where n is the size of input system and $H(P)$ is the entropy of the distribution P of the input system. In the method of information spectrum, we treat the general asymptotic formula, which gives the relationship between the asymptotic optimal performance and the normalized logarithm of the likelihood of the probability distribution. In order to treat a special case, we apply the general asymptotic formula to the respective information source and calculate the asymptotic stochastic behavior of the normalized logarithm of the likelihood. That is, in the information spectrum method, we have two steps, deriving the general asymptotic formula and applying the general asymptotic formula. With respect to fixed-length source coding and intrinsic randomness, the same relation holds concerning the general asymptotic formula in the second-order coding rate. However, there is a difference concerning the application of the general asymptotic formula to the independent and identical distributions. That is, while the normalized logarithm of the likelihood approaches the entropy $H(P)$ in probability in the first-order coding rate, the stochastic behavior is asymptotically described by the Gaussian distribution in the first-order coding rate. In other words, in the second step, the first-order coding rate corresponds to the law of large numbers, and the second-order coding rate corresponds to the central limit theorem.

In this paper, we treat the channel coding in the second-order coding rate, i.e., the case in which the transmission length is $C_W^{\mathrm{DM}}n + a\sqrt{n}$. Similar to the aforementioned case, we employ the method of information spectrum. That is, we treat the general channel, which is the general sequence

Manuscript received August 22, 2008; revised March 23, 2009. Current version published October 21, 2009. This work was supported by MEXT under Grant-in-Aid for Scientific Research on Priority Area "Deepening and Expansion of Statistical Mechanical Informatics (DEX-SMI)" 18079014 and by MEXT under Grant-in-Aid for Young Scientists (A) 20686026. The material in this paper was presented in part at the IEICE Technical Committee Conference on Information Theory, Wakayama, Japan, May 2008.

The author is with the Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan (e-mail: hayashi@math.is.tohoku.ac.jp).

Communicated by H. Yamamoto, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2009.2030478

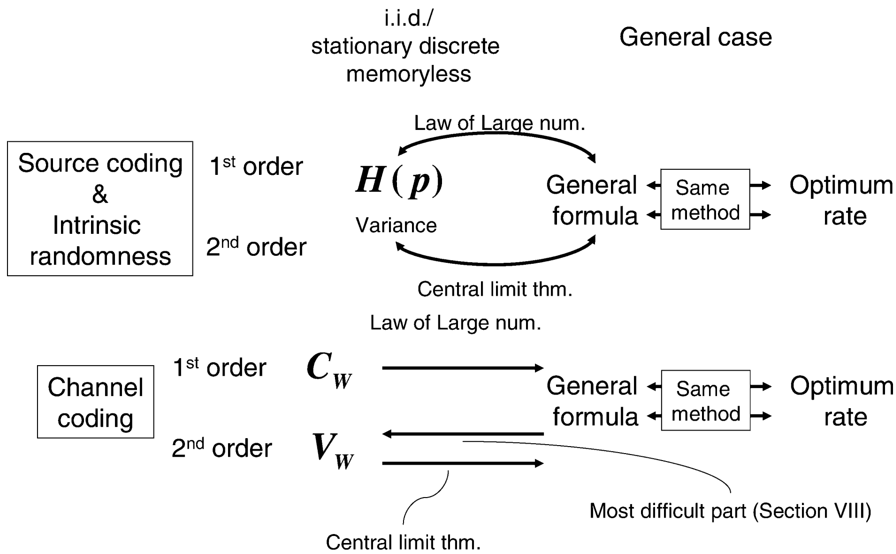


Fig. 1. Relationship between the present result and fixed-length source coding/intrinsic randomness. The $\rightarrow$ arrow describes the direct part, and the $\leftarrow$ arrow describes the converse part.

$\{W^n(y\,|\,x)\}$ of probability distributions without structure. As shown by Verdú and Han [14], this method enables us to characterize the asymptotic performance with only the random variable $\frac{1}{n}\log\frac{W^n(y\,|\,x)}{W^n_{P^n}(y)}$ (the normalized logarithm of the likelihood ratio between the conditional distribution and the nonconditional distribution) without any further assumption, where $W^n_{P^n}(y) \stackrel{\rm def}{=} \sum_x P^n(x)W^n(y\,|\,x)$. Concerning this general asymptotic formula, if we can suitably formulate theorems in the second-order coding rate and establish an appropriate relationship between the first-order coding rate and the second-order coding rate, we can easily extend proofs concerning the first-order coding rate to those of the second-order coding rate. Therefore, there is no serious difficulty in establishing the general asymptotic formula in the second-order coding rate. In order to clarify this point, we present proofs of some relevant theorems in the first-order coding rate, even though they are known.

In order to treat the special cases, it is sufficient to apply the general asymptotic formula, i.e., to calculate the asymptotic behavior of the random variable $\frac{1}{n}\log\frac{W^n(y\,|\,x)}{W^n_{P^n}(y)}$. The additive Markovian case can be treated in the same way as fixed-length source coding and intrinsic randomness. However, other special cases have other difficulties, which do not appear in fixed-length source coding or intrinsic randomness. The first difficulty is the optimization concerning the input distribution in the converse part of the channel coding. This problem commonly appears among three cases, i.e., the discrete memoryless case, the discrete memoryless case with cost constraint, and the Gaussian additive noise case with the energy constraint. In the discrete memoryless case, the second-order coding rate corresponds to simple application of the central limit theorem, while the first-order coding rate corresponds to the law of large numbers. Hence, the performance in second-order coding rate is characterized by the variance of the logarithmic likelihood ratio, and the direct part can be easily obtained in this case. This relationship is summarized in Fig. 1.

However, there is another difficulty in the direct part for the discrete memoryless case with cost constraint and the Gaussian additive noise case with the energy constraint. In these cases, all of the encoded signals have to satisfy cost constraint. This kind of difficulty does not appear in the case of first-order coding rate of both the discrete memoryless case with cost constraint and the Gaussian additive noise case with the energy constraint. This is because it is sufficient to construct the code whose average error probability goes to zero in the case of the first-order coding rate while it is required to construct the code whose average error probability goes to a given threshold ϵ in the case of the second-order coding rate. We find a code satisfying the following: its average error probability goes to zero and its average cost is less than the constraint.

Then, there exists a subcode satisfying the following: its average error probability goes to zero and the costs of all encoded signals are less than the constraint. However, the same method cannot be applied when we find a code satisfying the following: its average error probability goes to ϵ and its average cost is less than the constraint. In this paper, we directly construct a code, in which the costs of all encoded signals are less than the constraint.

Here, we describe the meaning of the second-order coding rate. When the transmission length is described by $nC^{\rm DM}_W + \sqrt{n}R_2$, as shown in Section IX-A, the optimal error can be approximately attained by random coding. Since it seems that random coding cannot be realized, our evaluation seems to be related to only the theoretical best performance. However, in the quantum key distribution, it can be realized concerning the phase bases [7], [2]. In such a setting, the coding length is on the order of 10 000 or 100 000 [8]. In the quantum key distribution, Hayashi [2] has applied the second-order coding rate to evaluate the phase error probability, which is directly linked to the security of the final key.

The remainder of this paper is organized as follows. In Section II, we revisit the second-order coding rate in the stationary discrete memoryless case, and discuss the second-order coding rate in the stationary discrete memoryless case with cost constraint. In Section III, the Markovian additive channel is treated. In Section IV, the Gaussian additive noise case with the energy constraint is considered. These results are shown in Section X by employing the method of information spectrum. In the present result, the performance of information transmission is discussed in terms of second-order coding rate using two important quantities V_W^+ and V_W^- instead of the capacity in the discrete memoryless case. In other cases, similar quantities play the same role.

In Section V, we compare our evaluation with the Gallager bound [1] in the second-order setting. In Section VI, the properties of V_W^+ and V_W^- are discussed. In Section VI-A, we discuss a typical example such that V_W^+ is different from V_W^-. In Section VI-B, the additive properties concerning V_W^+ and V_W^- are proved. In Section VII, the notations of the information spectrum are explained. In Section VIII, the performance of the information transmission is discussed in terms of the second-order coding rate using the information spectrum in the general case. That is, we present general formulas for the second-order coding rate. In Section IX, the theorem presented in the previous section is proved. In Section X, using general formulas for the second-order coding rate, we demonstrate our proof of the second-order coding rate in the stationary discrete memoryless case using our general result concerning the second-order coding rate. In this proof, the direct part is immediate. The converse part is the most difficult considered herein because we must treat the information spectrum for the general input distributions in the sense of the second-order coding rate.

II. Second-Order Coding Rate in Stationary Discrete Memoryless Channels

As the most typical case, we revisit the second-order coding rate of stationary discrete memoryless channels, in which we use an n-multiple application of the discrete channel $W(y\,|\,x)$, which transmits the information from the input system $\mathcal{X}$ to the output system $\mathcal{Y}$. That is, the channel considered here is given as the stationary discrete memoryless channel $W^{\times n}(\vec{y}_n\,|\,\vec{x}_n) \stackrel{\rm def}{=} \prod_{i=1}^n W(y_i\,|\,x_i)$, where $\vec{x}_n=(x_1,\dots,x_n)$ and $\vec{y}_n=(y_1,\dots,y_n)$. Note that, in this paper, $P\times P'(W\times W')$ denotes the product of two distributions P and P' (two channels W and W'), and $P^{\times n}(W^{\times n})$ denotes the product of n uses of the distribution P (the channel W), i.e., the nth independent and identical distribution (i.i.d.) of P (the nth stationary memoryless channel of W). In this case, when the transmission rate is less than the capacity $C_W^{\rm DM}$, the average error probability goes to 0 exponentially, if we use a suitable encoder and the maximum-likelihood decoder.

Let N_n be the size of the transmitted information. The encoder is a map ϕ from $\{1,\dots,N_n\}$ to $\mathcal{X}^n$, and the decoder is given by the set of subsets $\{\mathcal{D}_i\}_{i=1}^{N_n}$ of $\mathcal{Y}^n$, where $\mathcal{D}_i$ corresponds to the decoding region of $i\in\{1,\dots,N_n\}$. Then, the code is given by the triple $(N_n,\phi,\{\mathcal{D}_i\}_{i=1}^{N_n})$ and is denoted by Φ_n. The average error probability $P_{e,W^{\times n}}(\Phi_n)$ is described as

$$P_{e,W^{\times n}}(\Phi_n) \stackrel{\rm def}{=} \frac{1}{N_n}\sum_{i=1}^{N_n}\left(1-W_{\phi(i)}^{\times n}(\mathcal{D}_i)\right)$$

where $W_x(y) \stackrel{\rm def}{=} W(y\,|\,x)$. For simplicity, the size N_n is denoted by $|\Phi_n|$. The performance of the code Φ_n is given by the pair of $P_e(\Phi_n)$ and $|\Phi_n|$. As stated by the channel coding theorem [9], the capacity is given by

$$C_W^{\rm DM} = \max_{P\in\mathcal{P}(\mathcal{X})} I(P,W) = \min_{Q\in\mathcal{P}(\mathcal{Y})}\max_x D(W_x\,\|\,Q) \quad (1)$$

where $\mathcal{P}(\mathcal{Y})$ is the set of distributions on $\mathcal{Y}$, and

$$W_P(y) \stackrel{\rm def}{=} \sum_x P(x)W(y\,|\,x)$$

$$I(P,W) \stackrel{\rm def}{=} \sum_x P(x)D(W_x\,\|\,W_P)$$

$$D(P\,\|\,P') \stackrel{\rm def}{=} \sum_x P(x)\log\frac{P(x)}{P'(x)}.$$

The second equation of (1) was shown in Ohya–Petz–Watanabe [15] and Hayashi–Nagaoka [16] with the quantum setting. For a reader's convenience, its proof is given in Appendix I with the nonquantum setting. Thus, $Q_M \stackrel{\rm def}{=} \operatorname{argmin}_Q \max_x D(W_x\,\|\,Q)$ satisfies

$$D(W_x\,\|\,Q_M) \le C_W^{\rm DM}. \quad (2)$$

Throughout this paper, we choose the base of the logarithm to be e.

Although the above channel coding theorem concerns only the first-order coding rate of the transmission length $\log N_n$, our main focus is the analysis of the second-order coding rate. When the transmission length $\log N_n$ asymptotically behaves as $nC_W^{\rm DM}+a\sqrt{n}$, the optimal average error is given by

$$C_p^{\rm DM}\left(a, C_W^{\rm DM}\,|\,W\right) \stackrel{\rm def}{=} \inf_{\{\Phi_n\}_{n=1}^\infty}\left\{ \overline{\lim_{n\to\infty}}\, P_{e,W^{\times n}}(\Phi_n) \,\middle|\, \underline{\lim_{n\to\infty}}\,\frac{1}{\sqrt{n}}\left(\log|\Phi_n| - nC_W^{\rm DM}\right)\ge a\right\}. \quad (3)$$

Fixing the average error probability, we obtain the following quantity:

$$C^{\rm DM}\left(\epsilon, C_W^{\rm DM}\,|\,W\right) \stackrel{\rm def}{=} \sup_{\{\Phi_n\}_{n=1}^\infty}\left\{ \underline{\lim_{n\to\infty}}\,\frac{1}{\sqrt{n}}\left(\log|\Phi_n| - nC_W^{\rm DM}\right) \,\middle|\, \overline{\lim_{n\to\infty}}\, P_{e,W^{\times n}}(\Phi_n)\le\epsilon\right\}. \quad (4)$$

We refer to this value as the optimum second-order transmission rate with the error probability ϵ. In order to treat the second-order coding rate, we need the distribution function G for the standard Gaussian distribution (with expectation 0 and variance 1), which is defined by

$$G(x) \stackrel{\rm def}{=} \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

In this problem, the quantity $V_{P,W}$

$$V_{P,W} \stackrel{\rm def}{=} \sum_x P(x)\sum_y W_x(y)\left(\log\frac{W_x(y)}{W_P(y)} - D(W_x\,\|\,W_P)\right)^2$$

plays an important role. By using these quantities, $C_p^{\rm DM}(a, C_W^{\rm DM} | W)$ and $C^{\rm DM}(\epsilon, C_W^{\rm DM} | W)$ are calculated in the stationary discrete memoryless case as follows.

Theorem 1 [3]: When the cardinality $|\mathcal{X}|$ is finite and $P_M \stackrel{\rm def}{=} {\rm argmax}_P I(P,W)$ exists uniquely, then

$$C_p^{\rm DM}\left(a, C_W^{\rm DM} \,|\, W\right) = G\left(a/\sqrt{V_{P_M,W}}\right) \tag{5}$$

$$C^{\rm DM}\left(\epsilon, C_W^{\rm DM} \,|\, W\right) = \sqrt{V_{P_M,W}}G^{-1}(\epsilon). \tag{6}$$

When $\{W_x\}$ is linearly independent by regarding distributions as positive vectors, the map $P \mapsto W_P$ is a one-to-one map. Then, $P_M \stackrel{\rm def}{=} {\rm argmax}_P I(P,W)$ exists uniquely. However, when $\{W_x\}$ is not linearly independent, ${\rm argmax}_P I(P,W)$ is not necessarily unique. In order to treat such a case, we introduce two quantities V_W^+ and V_W^- and two distributions P_{M+} and P_{M-}

$$V_W^+ \stackrel{\rm def}{=} \max_{P\in\mathcal{V}} V_{P,W}$$

$$V_W^- \stackrel{\rm def}{=} \min_{P\in\mathcal{V}} V_{P,W}$$

$$P_{M+} \stackrel{\rm def}{=} \mathop{\rm argmax}_{P\in\mathcal{V}} V_{P,W}$$

$$P_{M-} \stackrel{\rm def}{=} \mathop{\rm argmin}_{P\in\mathcal{V}} V_{P,W}$$

where $\mathcal{V} \stackrel{\rm def}{=} \{P \,|\, I(P,W) = C_W^{\rm DM}\}$. In order to treat such a case, Theorem 1 is generalized as follows.

Theorem 2 [3]: When the cardinality $|\mathcal{X}|$ is finite and the set $\mathcal{V}$ has multiple elements, (5) and (6) are generalized as

$$C_p^{\rm DM}\left(a, C_W^{\rm DM} \,|\, W\right) = \begin{cases} G\left(a/\sqrt{V_W^+}\right), & a \ge 0 \\ G\left(a/\sqrt{V_W^-}\right), & a < 0 \end{cases}$$

$$C^{\rm DM}\left(\epsilon, C_W^{\rm DM} \,|\, W\right) = \begin{cases} \sqrt{V_W^+}G^{-1}(\epsilon), & \epsilon \ge 1/2 \\ \sqrt{V_W^-}G^{-1}(\epsilon), & \epsilon < 1/2. \end{cases}$$

More precisely, the direct part

$$C_p^{\rm DM}\left(a, C_W^{\rm DM} \,|\, W\right) \le \begin{cases} G\left(a/\sqrt{V_W^+}\right), & a \ge 0 \\ G\left(a/\sqrt{V_W^-}\right), & a < 0 \end{cases} \tag{7}$$

$$C^{\rm DM}\left(\epsilon, C_W^{\rm DM} \,|\, W\right) \ge \begin{cases} \sqrt{V_W^+}G^{-1}(\epsilon), & \epsilon \ge 1/2 \\ \sqrt{V_W^-}G^{-1}(\epsilon), & \epsilon < 1/2 \end{cases} \tag{8}$$

holds without any assumption, and the converse part

$$C_p^{\rm DM}\left(a, C_W^{\rm DM} \,|\, W\right) \ge \begin{cases} G\left(a/\sqrt{V_W^+}\right), & a \ge 0 \\ G\left(a/\sqrt{V_W^-}\right), & a < 0 \end{cases}$$

$$C^{\rm DM}\left(\epsilon, C_W^{\rm DM} \,|\, W\right) \le \begin{cases} \sqrt{V_W^+}G^{-1}(\epsilon), & \epsilon \ge 1/2 \\ \sqrt{V_W^-}G^{-1}(\epsilon), & \epsilon < 1/2 \end{cases}$$

holds with the assumption $|\mathcal{X}| < \infty$.

Next, consider the cost function $c: \mathcal{X} \mapsto \mathbb{R}$. In this case, we often assume that all codewords $\phi(i)$ of the code Φ_n belong to the set

$$\mathcal{X}_{c,K}^n \stackrel{\rm def}{=} \left\{ \vec{x}_n \in \mathcal{X}^n \,\middle|\, \sum_{i=1}^n c(x_i) \le nK \right\}.$$

The maximum coding rate with the above condition is called the capacity with the cost constraint, and is given by [10]

$$\begin{aligned} C_{W,c,K}^{\rm DM} &= \max_{P:{\rm E}_P c(x) \le K} I(P,W) \\ &= \min_{Q\in\mathcal{P}(\mathcal{Y})} \max_{P:{\rm E}_P c(x)\le K} J(P,Q,W) \end{aligned}$$

where

$$J(P,Q,W) \stackrel{\rm def}{=} \sum_{x\in\mathcal{X}} P(x) D(W_x \,\|\, Q).$$

In the same way to (3) and (4), we define the following values with the cost constraint:

$$\begin{aligned} &C_p^{\rm DM}\left(a, C_W^{\rm DM} \,|\, W, c, K\right) \\ &\stackrel{\rm def}{=} \inf_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varlimsup_{n\to\infty} P_{e,W^{\times n}}(\Phi_n) \right| \\ &\qquad \varliminf_{n\to\infty} \frac{1}{\sqrt{n}}\left(\log|\Phi_n| - nC_W^{\rm DM}\right) \ge a, \\ &\qquad \left. {\rm supp}(\Phi_n) \subset \mathcal{X}_{c,K}^n \right\}. \end{aligned} \tag{9}$$

$$\begin{aligned} &C^{\rm DM}\left(\epsilon, C_W^{\rm DM} \,|\, W, c, K\right) \\ &\stackrel{\rm def}{=} \sup_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varliminf_{n\to\infty} \frac{1}{\sqrt{n}}\left(\log|\Phi_n| - nC_W^{\rm DM}\right) \right| \\ &\qquad \varlimsup_{n\to\infty} P_{e,W^{\times n}}(\Phi_n) \le \epsilon, \\ &\qquad \left. {\rm supp}(\Phi_n) \subset \mathcal{X}_{c,K}^n \right\} \end{aligned} \tag{10}$$

where ${\rm supp}(\Phi_n)$ expresses the set $\{\phi(1), \ldots, \phi(N)\}$ for a code $\Phi = (N, \phi, \{\mathcal{D}_i\}_{i=1}^N)$. We introduce two quantities $V_{W,c,K}^+$ and $V_{W,c,K}^-$ and two distributions $P_{M+,c,K}$ and $P_{M-,c,K}$

$$V_{W,c,K}^+ \stackrel{\rm def}{=} \max_{P\in\mathcal{V}_{c,K}} V_{P,W}$$

$$V_{W,c,K}^- \stackrel{\rm def}{=} \min_{P\in\mathcal{V}_{c,K}} V_{P,W}$$

$$P_{M+,c,K} \stackrel{\rm def}{=} \mathop{\rm argmax}_{P\in\mathcal{V}_{c,K}} V_{P,W}$$

$$P_{M-,c,K} \stackrel{\rm def}{=} \mathop{\rm argmin}_{P\in\mathcal{V}_{c,K}} V_{P,W}$$

where $\mathcal{V}_{c,K} \stackrel{\rm def}{=} \{P \,|\, I(P,W) = C_{W,c,K}^{\rm DM}, {\rm E}_P c(x) \le K\}$.

Theorem 3: When the cardinality $|\mathcal{X}|$ is finite

$$C_p^{\rm DM}\left(a, C_{W,c,K}^{\rm DM} \,|\, W, c, K\right) = \begin{cases} G\left(a/\sqrt{V_{W,c,K}^+}\right), & a \ge 0 \\ G\left(a/\sqrt{V_{W,c,K}^-}\right), & a < 0 \end{cases}$$

$$C^{\rm DM}\left(\epsilon, C_{W,c,K}^{\rm DM} \,|\, W, c, K\right) = \begin{cases} \sqrt{V_{W,c,K}^+}G^{-1}(\epsilon), & \epsilon \ge 1/2 \\ \sqrt{V_{W,c,K}^-}G^{-1}(\epsilon), & \epsilon < 1/2. \end{cases}$$

More precisely, the direct part

$$C_p^{\rm DM}\left(a, C_{W,c,K}^{\rm DM} \,|\, W,c,K\right) \le \begin{cases} G\left(a/\sqrt{V_{W,c,K}^{+}}\right), & a \ge 0 \\ G\left(a/\sqrt{V_{W,c,K}^{-}}\right), & a < 0 \end{cases} \tag{11}$$

$$C^{\rm DM}\left(\epsilon, C_{W,c,K}^{\rm DM} \,|\, W,c,K\right) \ge \begin{cases} \sqrt{V_{W,c,K}^{+}}G^{-1}(\epsilon), & \epsilon \ge 1/2 \\ \sqrt{V_{W,c,K}^{-}}G^{-1}(\epsilon), & \epsilon < 1/2. \end{cases} \tag{12}$$

holds without any assumption, and the converse part

$$C_p^{\rm DM}\left(a, C_{W,c,K}^{\rm DM} \,|\, W,c,K\right) \ge \begin{cases} G\left(a/\sqrt{V_{W,c,K}^{+}}\right), & a \ge 0 \\ G\left(a/\sqrt{V_{W,c,K}^{-}}\right), & a < 0 \end{cases}$$

$$C^{\rm DM}\left(\epsilon, C_{W,c,K}^{\rm DM} \,|\, W,c,K\right) \le \begin{cases} \sqrt{V_{W,c,K}^{+}}G^{-1}(\epsilon), & \epsilon \ge 1/2 \\ \sqrt{V_{W,c,K}^{-}}G^{-1}(\epsilon), & \epsilon < 1/2. \end{cases}$$

holds with the assumption $|\mathcal{X}| < \infty$.

Remark 1: When the sets $\mathcal{X}$ and $\mathcal{Y}$ are given as general probability spaces with general σ-fields $\sigma(\mathcal{X})$ and $\sigma(\mathcal{Y})$, the above formulation can be extended with the following definition. The channel W is given by the real-valued function from $\mathcal{X}$ and $\sigma(\mathcal{Y})$ satisfying the following conditions: i) for any $x \in \mathcal{X}$, W is a probability measure on $\mathcal{Y}$; and ii) for any $F \in \sigma(\mathcal{Y})$, the function $x \mapsto W_x(F)$ is a measurable function on $\mathcal{X}$. The probability measure P takes on values in $\mathcal{X}$. Then, the summands $\sum_{x\in\mathcal{X}} P(x)$ and $\sum_{y\in\mathcal{Y}} W_x(y)$ are replaced by $\int_{\mathcal{X}} P(dx)$ and $\int_{\mathcal{Y}} W_x(dy)$, respectively. For any distribution Q on $\mathcal{Y}$, the function $\frac{W_x(y)}{W_P(y)}$ is replaced by the inverse of Radon–Nikodym derivative $\frac{dW_P}{dW_x}(y)$ of W_P with respect to W_x. In this extension, the direct parts (7), (8), (11), and (12) are valid.

III. Second-Order Coding Rate in Additive Markovian Channel

Next, we focus on the additive Markovian channel, in which we assume that the additive noise obeys the transition matrix $Q(y\,|\,x)$ on the set $\mathcal{X} = \{1,\ldots,d\}$. Then, the channel $W(Q)^n(\vec{y}_n\,|\,\vec{x}_n)$ has the form $\prod_{i=1}^n Q(y_i - x_i\,|\,y_{i-1}-x_{i-1})$, where $y_0 - x_0$ is the initial state s_0 and the arithmetic is based on mod d. For simplicity, we assume that the transition matrix $Q(y\,|\,x)$ is irreducible. Then, the nth marginal distribution $Q^{(n)}(x_n) \stackrel{\rm def}{=} \sum_{x_0,\ldots,x_{n-1}}\prod_{i=1}^n Q(x_i\,|\,x_{i-1})$ approaches the stationary distribution $P_Q(x)$, which is given as the eigenvector of $Q(y\,|\,x)$ associated with the eigenvalue 1 [12]. When the conditional distribution $Q(y\,|\,x)$ is denoted by $Q_x(y)$, the normalized entropy of the distribution $Q^n(\vec{x}_n) \stackrel{\rm def}{=} \prod_{i=1}^n Q(x_i\,|\,x_{i-1})$ goes to $H(Q) \stackrel{\rm def}{=} \sum_x P_Q(x)H(Q_x)$. Then, by defining the capacity $C_W^{\rm AM}$ in the same way as $C_W^{\rm DM}$, the channel capacity $C_W^{\rm AM}$ is calculated as

$$C_Q^{\rm AM} = \log d - H(Q). \tag{13}$$

Similarly to $C_p^{\rm DM}(a, C_W^{\rm DM}\,|\,W)$ and $C^{\rm DM}(\epsilon, C_W^{\rm DM}\,|\,W)$, the second-order quantities $C_p^{\rm AM}(a, C_Q^{\rm AM}\,|\,W)$ and $C^{\rm AM}(\epsilon, C_Q^{\rm AM}\,|\,W)$ are defined for the additive Markovian case. Then, the following theorem holds. In this problem, as is shown later, the mutual information attains the capacity when the input distribution is the uniform distribution. So, when the input distribution is the uniform distribution, the additive Markovian version is the variance $V(Q)$

$$\begin{aligned} V(Q) \stackrel{\rm def}{=} &\sum_{y,x} Q(y\,|\,x)P_Q(x)(-\log Q(y\,|\,x) - H(Q))^2 \\ &+ 2\sum_{z,y,x} Q(z\,|\,y)Q(y\,|\,x)P_Q(x) \\ &\times(-\log Q(z\,|\,y) - H(Q))(-\log Q(y\,|\,x) - H(Q)) \end{aligned}$$

which plays an important role. By using these quantities, $C_p^{\rm AM}(a, C_Q^{\rm AM}\,|\,W)$ and $C^{\rm AM}(\epsilon, C_Q^{\rm AM}\,|\,W)$ are calculated in the additive Markovian case as follows.

Theorem 4: The relations

$$C_p^{\rm AM}\left(a, C_Q^{\rm AM}\,|\,W\right) = G(a/\sqrt{V(Q)})$$
$$C^{\rm AM}\left(\epsilon, C_Q^{\rm AM}\,|\,W\right) = \sqrt{V(Q)}G^{-1}(\epsilon).$$

hold.

Remark 2: It seems strange that Theorems 2 and 3 have separate treatment for $a \ge 0$ and $a < 0$ while Theorem 4 has no such treatment. The separate treatment is caused by the nonuniqueness of the distribution P realizing the capacity. However, in the additive memoryless channel case, the uniqueness holds. That is, only the uniform distribution realizes the capacity. A similar fact holds for the additive Markovian case.

IV. Second-Order Coding Rate in Gaussian Channel

In this section, we consider the case of additive Gaussian noise. In this case, both of the input system and the output system are given by $\mathbb{R}$, and the output distribution $W_x(y)$ is given by $\frac{1}{\sqrt{2\pi N}}e^{-\frac{(y-x)^2}{2N}}$ for a given variance N of noise. If there is no restriction for input signal, the capacity diverges. Hence, it is natural to consider the cost constraint. Consider the cost function $c(x) \stackrel{\rm def}{=} x^2$ and the maximum cost S. Then, the maximum mutual information $\max_{P:{\rm E}_P x^2 \le S} I(P,W)$ is attained when P is equal to $P_M(x) \stackrel{\rm def}{=} \frac{1}{\sqrt{2\pi S}}e^{-\frac{x^2}{2S}}$. In this case

$$D(W_x\,\|\,W_{P_M}) = \frac{1}{2}\log\left(1+\frac{S}{N}\right) + \frac{\frac{x^2}{N} - \frac{S}{N}}{2\left(1+\frac{S}{N}\right)}. \tag{14}$$

Then, the capacity is known to be [9], [11]

$$C_{N,S}^G = \max_{P:{\rm E}_P x^2 \le E} I(P,W) = \frac{1}{2}\log\left(1+\frac{S}{N}\right).$$

Since

$$\int_{-\infty}^{\infty}\left(\log\frac{W_x(y)}{W_{P_M}(y)} - D(W_x\,\|\,W_{P_M})\right)^2 W_x(y)dy = \frac{\frac{S^2}{N^2} + 2\frac{x^2}{N}}{2\left(1+\frac{S}{N}\right)^2}$$

$V_{P_M,W}$ is calculated as

$$V_{P_M,W} = \frac{\frac{S^2}{N^2} + 2\frac{S}{N}}{2\left(1+\frac{S}{N}\right)^2}.$$

Since the cardinality of $\mathbb{R}$ is infinite, the assumption of Section II does not hold. That is, we cannot apply Theorem 3. However, the following theorem holds.

Theorem 5: Define the quantities $C_p^G(a, C_{N,S}^G \mid N, S)$ and $C^G(\epsilon, C_{N,S}^G \mid N, S)$ in the same way as (9) and (10). Then

$$C_p^G\left(a, C_{N,S}^G \mid N, S\right) = G\left(a/\sqrt{V_{P_M,W}}\right)$$
$$C^G\left(\epsilon, C_{N,S}^G \mid N, S\right) = \sqrt{V_{P_M,W}}G^{-1}(\epsilon).$$

V. Comparison With the Gallager Bound

At first glance, the Gallager bound [1] seems to work well for evaluating the average error probability, even when the transmission length is close to nC_W^{DM}. This is because this bound gives the optimal exponential rate when the coding rate is greater than the critical rate. In this section, we clarify whether the present evaluation or the Gallager bound [1] provides a better evaluation when the transmission length is close to nC_W^{DM}. For this analysis, we describe the transmission length by $nC_W^{\mathrm{DM}} + \sqrt{n}R_2$. Let us compare the present evaluation with the Gallager bound, which is given by

$$\min_{\Phi_n : |\Phi_n| \le e^{nR}} P_{e,W^{\times n}}(\Phi_n) \le \min_{P \in \mathcal{P}(\mathcal{Y})} \min_{0 \le s \le 1} e^{n(Rs + \psi_P(s))} \quad (15)$$

where

$$\psi_P(s) \stackrel{\mathrm{def}}{=} \log \sum_y \left(\sum_x P(x) W_x(y)^{\frac{1}{1+s}} \right)^{1+s}.$$

Since the present evaluation is essentially based on Verdú and Han's method [14], this comparison can be regarded as a comparison between Verdú and Han's evaluation and the Gallager bound. Next, we substitute $nC_W^{\mathrm{DM}} + \sqrt{n}R_2$ into nR. Then

$$\min_{0 \le s \le 1} e^{n(Rs + \psi_P(s))} = e^{n \min_{0 \le s \le 1}(C_W^{\mathrm{DM}} s + \frac{R_2}{\sqrt{n}} s + \psi_P(s))}.$$

Taking the derivatives of $\psi_P(s)$, we obtain

$$\left.\frac{d\psi_P(s)}{ds}\right|_{s=0} = -I(P, W)$$
$$\left.\frac{d^2\psi_P(s)}{ds^2}\right|_{s=0} = V_{P,W}.$$

When $C_W^{\mathrm{DM}} = I(P, W)$

$$\begin{aligned} &C_W^{\mathrm{DM}} s + \frac{R_2}{\sqrt{n}} s + \psi_P(s) \\ &\cong C_W^{\mathrm{DM}} s + \frac{R_2}{\sqrt{n}} s - I(P, W)s + \frac{V_{P,W}}{2} s^2 \\ &= \frac{R_2}{\sqrt{n}} s + \frac{V_{P,W}}{2} s^2 \\ &= \frac{V_{P,W}}{2}\left(s + \frac{R_2}{\sqrt{n}V_{P,W}}\right)^2 - \frac{R_2^2}{2nV_{P,W}}. \end{aligned}$$

Therefore, as is rigorously shown in Appendix II, when $R_2 < 0$

$$\lim_{n \to \infty} n \min_{0 \le s \le 1} \left(C_W^{\mathrm{DM}} s + \frac{R_2}{\sqrt{n}} s + \psi_P(s) \right) = -\frac{R_2^2}{2V_{P,W}}. \quad (16)$$

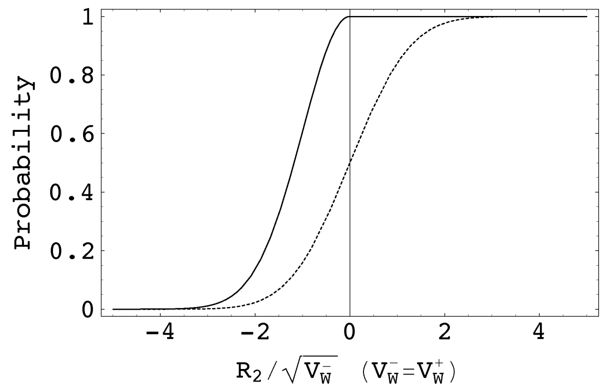


Fig. 2. Comparison between the present evaluation and the Gallager bound. The solid line indicates the Gallager bound, and the dotted line indicates the present evaluation.

Next, we set P as P_{M-}. Then, the Gallager bound yields

$$C_p^{\mathrm{DM}}\left(R_2, C_W^{\mathrm{DM}} \mid W\right) \le e^{-\frac{R_2^2}{2V_W^-}} \quad (17)$$

for any $R_2 < 0$. That is, the gap between our evaluation and the Gallager bound is equal to the difference between

$$F\left(\frac{R_2}{\sqrt{V_W^-}}\right) = \int_{-\infty}^{\frac{R_2}{\sqrt{V_W^-}}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad \text{and} \quad e^{-\frac{R_2^2}{2V_W^-}}.$$

Although the former is smaller than the latter, both exponential rates coincide in the limit $R_2 \to \infty$. Since we can consider that the Gallager bound gives the trivial bound for $R_2 > 0$, both evaluations are illustrated in Fig. 2.

Next, we consider the same comparison for the additive Markovian case. Substitute 1 to n in (15). Then, the Gallager's inequality

$$\min_{\Phi : |\Phi| \le e^{R'}} P_{e,W}(\Phi) \le \min_{0 \le s \le 1} e^{(R's + \psi_P(s))} \quad (18)$$

holds for any channel W, any input distribution P, and any real positive number R'. Substituting $\mathcal{X}^n$ and the additive Markovian channel $Q^n(\vec{y}_n - \vec{x}_n)$, the uniform distribution into $\mathcal{X}$, $W(y \mid x)$, and P, we obtain

$$\psi_P(s) = n\psi_{Q,n}(s)$$

where

$$\psi_{Q,n}(s) \stackrel{\mathrm{def}}{=} -s \log d + \frac{1+s}{n} \log \left(\sum_{\vec{x}_n} Q^n(\vec{x}_n)^{\frac{1}{1+s}} \right).$$

Thus, by substituting nR into R', inequality (18) yields that

$$\min_{\Phi_n : |\Phi_n| \le e^{nR}} P_{e,W(Q)^n}(\Phi_n) \le \min_{0 \le s \le 1} e^{n(Rs + \psi_{Q,n}(s))}. \quad (19)$$

Since the asymptotic first and second cummulants of the random variable $\log Q^n(\vec{x}_n)$ are $-H(Q)$ and $V(Q)$, we have

$$\log \left(\sum_{\vec{x}_n} Q^n(\vec{x}_n)^{1+\frac{t'}{\sqrt{n}}} \right) = -H(Q)t'\sqrt{n} + \frac{V(Q)}{2} t'^2 + o(t'^2)$$

as $t' \to 0$. Since $\frac{1}{1+\frac{t}{\sqrt{n}}} = 1 - \frac{t}{\sqrt{n}} + \frac{t^2}{n} + o(\frac{t^2}{n})$, substituting $-t + \frac{t^2}{\sqrt{n}} + o(\frac{t^2}{\sqrt{n}})$ into t', we obtain

$$\begin{aligned} & n\psi_{Q,n}\left(\frac{t}{\sqrt{n}}\right) \\ &= -t\sqrt{n}\log d + \left(1+\frac{t}{\sqrt{n}}\right) \\ &\quad \times\left(-H(Q)\left(-t+\frac{t^2}{\sqrt{n}}+o\left(\frac{t^2}{\sqrt{n}}\right)\right)\sqrt{n}+\frac{V(Q)}{2}t^2+o(t^2)\right) \\ &= (-\log d + H(Q))t\sqrt{n} + \frac{V(Q)}{2}t^2 + o(t^2). \end{aligned}$$

Substituting $nC_Q^{\mathrm{AM}} + \sqrt{n}R_2$ and $\frac{t}{\sqrt{n}}$ into nR and s in the exponent of (19), we have

$$\begin{aligned} & n\left(C_Q^{\mathrm{AM}}s + \frac{R_2}{\sqrt{n}}s + \psi_{Q,n}(s)\right) \\ &= n\left(C_Q^{\mathrm{AM}}\frac{t}{\sqrt{n}} + \frac{R_2}{\sqrt{n}}\frac{t}{\sqrt{n}} + \psi_{Q,n}\left(\frac{t}{\sqrt{n}}\right)\right) \\ &= R_2 t + \frac{V(Q)}{2}t^2 + o(t^2) \\ &= \frac{V(Q)}{2}\left(t + \frac{R_2}{V(Q)}\right)^2 - \frac{R_2^2}{2V(Q)} \cdot + o(t^2). \end{aligned}$$

Therefore, when $R_2 < 0$, choosing $s = \frac{-R_2}{V(Q)\sqrt{n}}$, we obtain

$$\min_{\Phi^n : |\Phi^n| \le e^{nC_Q^{\mathrm{AM}} + \sqrt{n}R_2}} P_{e,W(Q)^n}(\Phi^n) \le e^{-\frac{R_2^2}{2V(Q)}}$$

which has the same form as (17).

In both cases, when $-3 \le R_2 \le 2$, the difference is not so small. In such a case, it is better to use the present evaluation. That is, the Gallager bound does not give the best evaluation in this case. This conclusion is opposite to the exponential evaluation when the rate is greater than the critical rate. Han [5] calculated the exponential rate of the present bound, and found that it is worse than that of the Gallager bound.[1]

Moreover, a similar conclusion was obtained in the low-density parity-check (LDPC) case. Kabashima and Saad [13] compared the Gallager upper bound of the average error probability and the approximation of the average error probability by the replica method. That is, they compared both thresholds of the rate, i.e., both maximum transmission rates at which the respective error probability goes to zero. In their study [13, Table I], they pointed out that there exists a nonnegligible difference between these two thresholds in the LDPC case. This information may be helpful for discussing the performance of the Gallager bound.

VI. Properties of V_w^+ and V_w^-

A. Example

In this section, we consider a typical example, in which V_W^+ is different from V_W^-. For this purpose, we choose two parameters $q_1, q_2 \in [0,1]$ satisfying

$$\begin{aligned} 0 \le 2q_1 - q_2 \le 1 \\ h(q_1) - \frac{h(q_2)+h(2q_1-q_2)}{2} \le -\log\max\{q_1, 1-q_1\} \end{aligned} \tag{20}$$

[1]This description was provided in the original Japanese version, but not in the English translation.

where $h(x) \stackrel{\mathrm{def}}{=} -x\log x - (1-x)\log(1-x)$. According to the following three conditions i), ii), and iii), we define the five joint distributions W_1, W_2, W_3, W_4, and W_5 on two random variables $A = 0,1$ and $B = 0,1$. In the following, $Q^A(Q^B)$ denotes the marginal distribution of A concerning $A(B)$.

i) Uniformity on A
All distributions are assumed to satisfy

$$W_i^A(0) = 1/2.$$

ii) Same marginal distribution on B for $i = 1,2$
Two random variables $A = 0,1$ and $B = 0,1$ are not independent in W_1 and W_2, but W_1 and W_2 have the same marginal distribution on B. That is

$$\begin{aligned} W_1^{B|A}(0|0) = W_2^{B|A}(0|1) = q_2 \\ W_1^{B|A}(0|1) = W_2^{B|A}(0|0) = 2q_1 - q_2 \end{aligned}$$

where $Q^{B|A}(i|j)$ is the conditional distribution of $B = i$ given $A = j$ when the joint distribution is given by Q. Thus, W_1 and W_2 satisfy

$$W_1^B(0) = W_2^B(0) = q_1.$$

iii) Independence between A and B for $i = 3,4,5$.
Due to condition (20), there exist two solutions for x in the following equation because $d(x\|q_1)$ is monotone increasing in $(q_1, 1)$ and is monotone decreasing in $(0, q_1)$:

$$h(q_1) - \frac{h(q_2)+h(2q_1-q_2)}{2} = d(x\|q_1)$$

where

$$d(x\|y) \stackrel{\mathrm{def}}{=} x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y}.$$

Letting p_1 and p_2 be these two solutions, we define three distributions W_3, W_4, and W_5, in which two random variables $A = 0,1$ and $B = 0,1$ are independent, by

$$W_3^B(0) = p_1, W_4^B(0) = p_2, W_5^B(0) = q_1.$$

From the construction, we can check that

$$D(W_i\|W_5) = h(q_1) - \frac{h(q_2)+h(2q_1-q_2)}{2} \tag{21}$$

for $i = 1,2,3,4$. Consider the subsets of joint distributions

$$\begin{aligned} \mathcal{Z}_0 &\stackrel{\mathrm{def}}{=} \{Q \mid Q^A(0) = 1/2\} \\ \mathcal{Z}_1 &\stackrel{\mathrm{def}}{=} \{Q \in \mathcal{Z}_0 \mid Q^B(0) = q_1\} \\ \mathcal{Z}_2 &\stackrel{\mathrm{def}}{=} \{Q \in \mathcal{Z}_0 \mid Q^{B|A}(0|0) = Q^{B|A}(0|1)\}. \end{aligned}$$

Then, $\mathcal{Z}_1 \cap \mathcal{Z}_2 = \{W_5\}$. Hence, the relationship among $\mathcal{Z}_0$, $\mathcal{Z}_1$, $\mathcal{Z}_2$, W_1, W_2, W_3, W_4, and W_5 is illustrated in Fig. 3.

Then, the following lemma holds.

Lemma 1:

$$\operatorname*{argmax}_Q \min_{x=1,2} D(W_x\|Q) = \operatorname*{argmax}_{Q\in\mathcal{Z}_1} \min_{x=1,2} D(W_x\|Q) \tag{22}$$

$$\operatorname*{argmax}_Q \min_{x=3,4} D(W_x\|Q) = \operatorname*{argmax}_{Q\in\mathcal{Z}_2} \min_{x=3,4} D(W_x\|Q). \tag{23}$$

Its proof is given in Appendix III.

Therefore, (21) implies that

$$\operatorname*{argmax}_{Q} \min_{x=1,2,3,4} D(W_x \,\|\, Q) = W_5$$

and

$$\begin{aligned}\max_Q \min_{x=1,2} D(W_x \,\|\, Q) &= \max_{Q \in \mathcal{Z}_1} \min_{x=1,2} D(W_x \,\|\, Q) \\ &= h(q_1) - \frac{h(q_2) + h(2q_1 - q_2)}{2} \\ \max_Q \min_{x=3,4} D(W_x \,\|\, Q) &= \max_{Q \in \mathcal{Z}_2} \min_{x=3,4} D(W_x \,\|\, Q) \\ &= h(q_1) - \frac{h(q_2) + h(2q_1 - q_2)}{2}.\end{aligned}$$

That is, the capacity of the channel $x = 1,2,3,4 \mapsto W_x$ is calculated as

$$\begin{aligned}C_W^{\mathrm{DM}} &= \max_Q \min_{x=1,2,3,4} D(W_x \,\|\, Q) \\ &= h(q_1) - \frac{h(q_2) + h(2q_1 - q_2)}{2}.\end{aligned}$$

Then, the set $\mathcal{V}$ is given by the convex hull of $P = (1/2, 1/2, 0, 0)$ and $P' = (0, 0, \frac{q_1 - p_2}{p_1 - p_2}, \frac{q_1 - p_1}{p_2 - p_1})$. Thus, $V_{\lambda P + (1-\lambda)P', W} = \lambda V_{P,W} + (1-\lambda) V_{P',W}$. When $V_{P,W} \le V_{P',W}$

$$V_W^+ = V_{P',W}, V_W^- = V_{P,W}.$$

Otherwise

$$V_W^+ = V_{P,W}, V_W^- = V_{P',W}.$$

Our numerical analysis (Fig. 4) suggests the relation $V_{P,W} \le V_{P',W}$.

B. Additivity

The capacity satisfies the additivity condition. That is, for any two channels $\{W_x(y)\}$ and $\{W'_{x'}(y')\}$, the combined channel $\{(W \times W')_{x,x'}(y,y') = W_x(y) W'_{x'}(y')\}$ satisfies the following:

$$C_{W \times W'}^{\mathrm{DM}} = C_W^{\mathrm{DM}} + C_{W'}^{\mathrm{DM}}.$$

Similarly, as mentioned in the following lemma, V_W^+ and V_W^- satisfy the additivity condition.

Lemma 2: The equations

$$V_{W \times W'}^+ = V_W^+ + V_{W'}^+ \tag{24}$$
$$V_{W \times W'}^- = V_W^- + V_{W'}^- \tag{25}$$

hold.

Proof of Lemma 2: We choose the distributions Q and Q' as

$$Q \stackrel{\mathrm{def}}{=} \operatorname*{argmin}_{Q} \max_x D(W_x \,\|\, Q)$$
$$Q' \stackrel{\mathrm{def}}{=} \operatorname*{argmin}_{Q'} \max_{x'} D(W'_{x'} \,\|\, Q').$$

Then

$$Q \times Q' = \operatorname*{argmin}_{Q''} \max_{x,x'} D(W_x \times W'_{x'} \,\|\, Q'').$$

Assume that a distribution P with the random variables x and x' satisfies the following:

$$\sum_{x,x'} P(x,x') W_x \times W'_{x'} = Q \times Q' \tag{26}$$

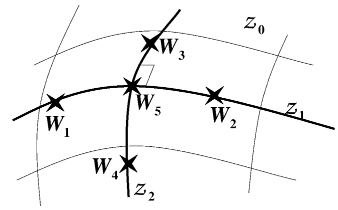


Fig. 3. $\mathcal{Z}_0$, $\mathcal{Z}_1$, $\mathcal{Z}_2$, W_1, W_2, W_3, W_4, and W_5.

$$I(P, W \times W') = C_W^{\mathrm{DM}} + C_{W'}^{\mathrm{DM}}. \tag{27}$$

Then, the marginal distributions P_1 and P_2 of P concerning x and x' satisfy

$$I(P_1, W) = C_W^{\mathrm{DM}}, I(P_2, W') = C_{W'}^{\mathrm{DM}}$$

which implies

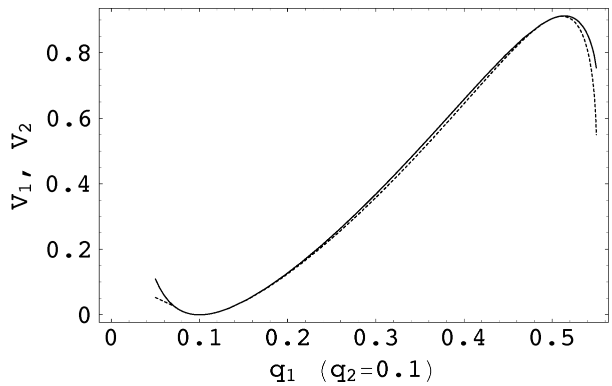
$$D(W_x \,\|\, Q) = C_W^{\mathrm{DM}}, D(W'_{x'} \,\|\, Q') = C_{W'}^{\mathrm{DM}}$$

for $x \in \operatorname{supp}(P_1)$ and $x' \in \operatorname{supp}(P_2)$, where $\operatorname{supp}(P)$ denotes the support of the distribution P. Hence

$$\begin{aligned}V_{P, W \times W'} &= \sum_{x,x'} P(x,x') \sum_{y,y'} W_x(y) W'_{x'}(y') \\ &\quad \times \left(\log \frac{W_x(y)}{Q(y)} + \log \frac{W'_{x'}(y')}{Q'(y')} \right)^2 \\ &\quad - \sum_{x,x'} P(x,x') (D(W_x \,\|\, Q) + D(W'_{x'} \,\|\, Q'))^2 \\ &= \sum_{x,x'} P(x,x') \sum_{y,y'} W_x(y) W'_{x'}(y') \\ &\quad \times \left(\left(\log \frac{W_x(y)}{Q(y)} \right)^2 + \left(\log \frac{W'_{x'}(y')}{Q'(y')} \right)^2 \right. \\ &\qquad \left. + 2 \log \frac{W_x(y)}{Q(y)} \log \frac{W'_{x'}(y')}{Q'(y')} \right) \\ &\quad - \sum_{x,x'} P(x,x') \Big(D(W_x \,\|\, Q)^2 + D(W'_{x'} \,\|\, Q')^2 \\ &\qquad + 2 D(W_x \,\|\, Q) D(W'_{x'} \,\|\, Q') \Big) \\ &= \sum_{x,x'} P(x,x') \sum_{y,y'} W_x(y) W'_{x'}(y') \\ &\quad \times \left(\left(\log \frac{W_x(y)}{Q(y)} \right)^2 + \left(\log \frac{W'_{x'}(y')}{Q'(y')} \right)^2 \right) \\ &\quad - D(W_x \,\|\, Q)^2 - D(W'_{x'} \,\|\, Q')^2 \\ &= V_{P_1, W} + V_{P_2, W'}.\end{aligned}$$

Therefore, when the conditions (26) and (27) are satisfied, the maximum of $V_{P, W \times W'}$ is equal to $V_W^+ + V_{W'}^+$, which implies (24). Similarly, we obtain (25). □

The same fact holds with the cost constraint. The capacity with the cost constraint satisfies the additivity condition. That is, for any two cost functions c and c' for channels $\{W_x(y)\}$ and



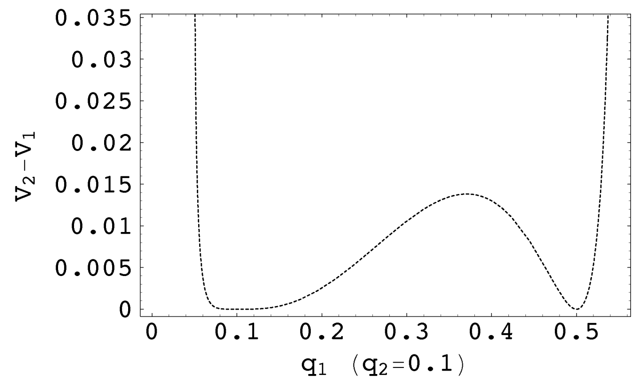


Fig. 4. Comparison between $V_1 = V_{P,W}$ (dotted line) and $V_2 = V_{P',W}$ (solid line).

$\{W'_{x'}(y')\}$, the combined cost $(c+c')(x,x') \stackrel{\rm def}{=} c(x)+c'(x')$ satisfies the following:

$$C^{\rm DM}_{W\times W',c+c',K+K'} = C^{\rm DM}_{W,c,K} + C^{\rm DM}_{W',c',K'}.$$

The quantities $V^+_{W,c,K}$ and $V^-_{W,c,K}$ satisfy the additivity condition.

Lemma 3: The equations

$$V^+_{W\times W',c+c',K+K'} = V^+_{W,c,K} + V^+_{W',c',K'} \tag{28}$$

$$V^-_{W\times W',c+c',K+K'} = V^-_{W,c,K} + V^-_{W',c',K'} \tag{29}$$

hold.

This lemma can be proven in the same way as Lemma 2 by replacing the definitions of Q and Q' by

$$Q \stackrel{\rm def}{=} \mathop{\rm argmin}_{Q} \max_{P:{\rm E}_P c(x)\le K} \sum_x P(x) D(W_x \| Q)$$

$$Q' \stackrel{\rm def}{=} \mathop{\rm argmin}_{Q'} \max_{P':{\rm E}_{P'} c'(x')\le K'} \sum_{x'} P'(x') D(W'_{x'} \| Q').$$

VII. Notations of the Information Spectrum

A. Information Spectrum

In this paper, we treat general channels. First, we focus on two sequences of probability spaces $\{\mathcal{X}_n\}_{n=1}^\infty$ of the input signal and those $\{\mathcal{Y}_n\}_{n=1}^\infty$ of the output signal, and a sequence of probability transition matrixes $\boldsymbol{W} \stackrel{\rm def}{=} \{W^n(y\,|\,x)\}_{n=1}^\infty$. We also focus on a sequence of distributions on input systems $\boldsymbol{P} \stackrel{\rm def}{=} \{P^n\}_{n=1}^\infty$. The asymptotic behavior of the logarithmic likelihood ratio between $W^n_x(y) \stackrel{\rm def}{=} W^n(y\,|\,x)$ and $W^n_{P^n}(y) \stackrel{\rm def}{=} \sum_{x\in\mathcal{X}_n} P^n(x) W^n(y\,|\,x)$ can be characterized by the following quantities:

$$I_p(R\,|\,\boldsymbol{P},\boldsymbol{W}) \stackrel{\rm def}{=} \overline{\lim_{n\to\infty}} \sum_{x\in\mathcal{X}_n} P^n(x) W^n_x \left\{ \frac{1}{n}\log\frac{W^n_x(y)}{W^n_{P^n}(y)} < R \right\}$$

$$I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W}) \stackrel{\rm def}{=} \sup\{R\,|\,I_p(R\,|\,\boldsymbol{P},\boldsymbol{W}) \le \epsilon\} = \inf\{R\,|\,I_p(R\,|\,\boldsymbol{P},\boldsymbol{W}) \ge \epsilon\}$$

for $0\le\epsilon<1$. Focusing on a sequence of distributions on output systems $\boldsymbol{Q} \stackrel{\rm def}{=} \{Q^n\}_{n=1}^\infty$, we can define

$$J_p(R\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \stackrel{\rm def}{=} \overline{\lim_{n\to\infty}} \sum_{x\in\mathcal{X}_n} P^n(x) W^n_x \left\{ \frac{1}{n}\log\frac{W^n_x(y)}{Q^n(y)} < R \right\}$$

$$J(\epsilon\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \stackrel{\rm def}{=} \sup\{R\,|\,J_p(R\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \le \epsilon\} = \inf\{R\,|\,J_p(R\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \ge \epsilon\}$$

for $0\le\epsilon<1$.

When the channel W^n is the nth stationary discrete memoryless channel $W^{\times n}$ of $W(y\,|\,x)$ and the probability distribution $\boldsymbol{P} = \{P^n\}$ is the nth i.i.d. $P^{\times n}$ of P, the law of large numbers guarantees that $I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W})$ coincides with the mutual information $I(P,W) = \sum_{x,y} P(x) W_x(y) \log\frac{W_x(y)}{W_P(y)}$. For a more detailed description of asymptotic behavior, we focus on the second-order n^β for $\beta<1$ concerning the coding length. In order to characterize the coefficient of the second-order n^β, we introduce the following quantities:

$$I^\beta_p(R_2,R_1\,|\,\boldsymbol{P},\boldsymbol{W}) \stackrel{\rm def}{=} \overline{\lim_{n\to\infty}} \sum_{x\in\mathcal{X}_n} P^n(x) W^n_x \left\{ \frac{1}{n^\beta}\left(\log\frac{W^n_x(y)}{W^n_{P^n}(y)} - nR_1\right) < R_2 \right\}$$

$$I^\beta(\epsilon,R_1\,|\,\boldsymbol{P},\boldsymbol{W}) \stackrel{\rm def}{=} \sup\left\{R_2\,|\,I^\beta_p(R_2,R_1\,|\,\boldsymbol{P},\boldsymbol{W}) \le \epsilon\right\} = \inf\left\{R_2\,\middle|\,I^\beta_p(R_2,R_1\,|\,\boldsymbol{P},\boldsymbol{W}) \ge \epsilon\right\}$$

for $0\le\epsilon<1$. When $\beta=\frac{1}{2}$, the superscript β is abbreviated. Similarly, $J^\beta_p(R_2,R_1\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$ and $J^\beta(\epsilon,R_1\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$ are defined for $0\le\epsilon<1$. When $\boldsymbol{W}$ is $\boldsymbol{W}^\times = \{W^{\times n}\}$ and $\boldsymbol{P}$ is $\boldsymbol{P}^\times = \{P^{\times n}\}$, the second order of the coding length is $n^{\frac{1}{2}}$ and the central limit theorem guarantees that $\frac{1}{n^{\frac{1}{2}}}(\log\frac{W^n_x(y)}{W^n_{P^n}(y)} - nI(P,W))$ asymptotically obeys the Gaussian distribution with expectation 0 and variance

$$V_{P,W} \stackrel{\rm def}{=} \sum_x P(x) \sum_y W_x(y) \left(\log\frac{W_x(y)}{W_P(y)} - I(P,W)\right)^2.$$

Therefore, when $\beta=\frac{1}{2}$, using the distribution function F for the standard Gaussian distribution, we can express the above quantities as follows:

$$I(\epsilon,I(P,W)\,|\,\boldsymbol{P}^\times,\boldsymbol{W}^\times) = \sqrt{V_{P,W}}G^{-1}(\epsilon). \tag{30}$$

In the case of additive channels, we focus on the limiting behavior of the entropy rate of the distributions $\boldsymbol{Q} = \{Q^n\}_{n=1}^\infty$ describing the additive noise. Similarly to the above, we define the following:

$$H_p(R\,|\,\boldsymbol{Q}) \stackrel{\rm def}{=} \underline{\lim_{n\to\infty}} \sum_{x\in\mathcal{X}_n} Q^n \left\{ \frac{-1}{n}\log Q^n(x) < R \right\}$$

$$H(\epsilon\,|\,\boldsymbol{Q}) \stackrel{\text{def}}{=} \sup\{R\,|\,H_p(R\,|\,\boldsymbol{Q}) \le \epsilon\} = \inf\{R|H_p(R\,|\,\boldsymbol{Q}) \ge \epsilon\}$$
$$H_p(R_2, R_1\,|\,\boldsymbol{Q}) \stackrel{\text{def}}{=} \underline{\lim}_{n\to\infty} \sum_{x\in\mathcal{X}_n} Q^n\left\{\frac{1}{n^\beta}(-\log Q^n(x) - nR_1) < R_2\right\}$$
$$H^\beta(\epsilon, R_1\,|\,\boldsymbol{Q}) \stackrel{\text{def}}{=} \sup\left\{R_2 \,\middle|\, H_p^\beta(R_2, R_1\,|\,\boldsymbol{Q}) \le \epsilon\right\} = \inf\left\{R_2 \,\middle|\, H_p^\beta(R_2, R_1\,|\,\boldsymbol{Q}) \ge \epsilon\right\}$$

for $0 \le \epsilon < 1$. As is discussed in [6, Sec. VII], when $\boldsymbol{Q}$ is given by a Markovian process $Q(y\,|\,x)$ and $\beta = \frac{1}{2}$, the relationships

$$H(\epsilon\,|\,\boldsymbol{Q}) = H(Q) \tag{31}$$
$$H(\epsilon, H(Q)\,|\,\boldsymbol{Q}) = \sqrt{V(Q)}G^{-1}(\epsilon) \tag{32}$$
$$H_p(R_2, H(Q)\,|\,\boldsymbol{Q}) = G\left(R_2/\sqrt{V(Q)}\right) \tag{33}$$

hold.

B. Stochastic Limits

In order to treat the relationship between the above quantities, we consider the limit superior in probability $\text{p}-\overline{\lim}_{n\to\infty}$ and the limit inferior in probability $\text{p}-\underline{\lim}_{n\to\infty}$, which are defined by

$$\text{p}-\overline{\lim_{n\to\infty}} Z_n\,|_{P_n} \stackrel{\text{def}}{=} \inf\{a\,|\,\lim_{n\to\infty} P_n\{Z_n > a\} = 0\}$$
$$\text{p}-\underline{\lim_{n\to\infty}} Z_n\,|_{P_n} \stackrel{\text{def}}{=} \sup\{a\,|\,\lim_{n\to\infty} P_n\{Z_n < a\} = 0\}.$$

In particular, when $\text{p}-\overline{\lim}_{n\to\infty} Z_n\,|_{P_n} = \text{p}-\underline{\lim}_{n\to\infty} Z_n\,|_{P_n} = a$, we write

$$\text{p}-\lim_{n\to\infty} Z_n\,|_{P_n} = a.$$

The concept $\text{p}-\underline{\lim}_{n\to\infty}$ can be generalized as

$$\epsilon-\text{p}-\underline{\lim_{n\to\infty}} Z_n\,|_{P_n} \stackrel{\text{def}}{=} \sup\{a\,|\,\overline{\lim_{n\to\infty}} P_n\{Z_n < a\} \le \epsilon\}.$$

From the definitions, we can check the following properties:

$$\epsilon-\text{p}-\underline{\lim_{n\to\infty}} Z_n + Y_n|_{P_n} \ge \epsilon-\text{p}-\underline{\lim_{n\to\infty}} Z_n|_{P_n} + \text{p}-\underline{\lim_{n\to\infty}} Y_n|_{P_n} \tag{34}$$
$$\epsilon-\text{p}-\underline{\lim_{n\to\infty}} Z_n + Y_n|_{P_n} \le \epsilon-\text{p}-\underline{\lim_{n\to\infty}} Z_n|_{P_n} + \text{p}-\overline{\lim_{n\to\infty}} Y_n|_{P_n} \tag{35}$$

where P_n is a joint distribution of two random variables Z_n and Y_n. As shown by Han [5], the relation

$$\text{p}-\underline{\lim_{n\to\infty}} \frac{1}{n^\alpha} \log \frac{P^n(x)}{P^{n\prime}(x)}\bigg|_{P^n} \ge 0 \tag{36}$$

holds for $\alpha > 0$ and any two sequences $\boldsymbol{P} = \{P^n\}$ and $\boldsymbol{P}' = \{P^{n\prime}\}$ of distributions with the variable x.

By using this concept, $I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W})$, $J(\epsilon\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$, $I^\beta(\epsilon, R_1\,|\,\boldsymbol{P},\boldsymbol{W})$, and $J^\beta(\epsilon, R_1\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$ are characterized by

$$I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W}) = \epsilon-\text{p}-\underline{\lim_{n\to\infty}} \frac{1}{n} \log \frac{W_x^n(y)}{W_{P^n}^n(y)}\bigg|_{\mathrm{P}_{P^n,W^n}}$$
$$J(\epsilon\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) = \epsilon-\text{p}-\underline{\lim_{n\to\infty}} \frac{1}{n} \log \frac{W_x^n(y)}{Q^n(y)}\bigg|_{\mathrm{P}_{P^n,W^n}}$$
$$I^\beta(\epsilon, R_1\,|\,\boldsymbol{P},\boldsymbol{W}) = \epsilon-\text{p}-\underline{\lim_{n\to\infty}} \frac{1}{n^\beta}\left(\log \frac{W_x^n(y)}{W_{P^n}^n(y)} - nR_1\right)\bigg|_{\mathrm{P}_{P^n,W^n}}$$
$$J^\beta(\epsilon, R_1\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) = \epsilon-\text{p}-\underline{\lim_{n\to\infty}} \frac{1}{n^\beta}\left(\log \frac{W_x^n(y)}{Q^n(y)} - nR_1\right)\bigg|_{\mathrm{P}_{P^n,W^n}}.$$

Substituting $W_{P^n}^n$ and Q^n into P^n and $P^{n\prime}$ in (36), and using (34), we obtain

$$I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W}) \le J(\epsilon\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$$
$$I^\beta(\epsilon, R_1\,|\,\boldsymbol{P},\boldsymbol{W}) \le J^\beta(\epsilon, R_1\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}).$$

Since $1 - H_p(R\,|\,\boldsymbol{Q}) = \underline{\lim}_{n\to\infty} Q^n\{\frac{1}{n}\log Q^n(x) < -R\}$, $H(\epsilon\,|\,\boldsymbol{Q})$ is characterized as

$$\begin{aligned} -H(\epsilon\,|\,\boldsymbol{Q}) &= -\inf\{R\,|\,H_p(R\,|\,\boldsymbol{Q}) \ge \epsilon\} \\ &= \sup\{-R\,|\,1 - H_p(R\,|\,\boldsymbol{Q}) \le 1-\epsilon\} \\ &= (1-\epsilon)-\text{p}-\underline{\lim_{n\to\infty}} \frac{1}{n}\log Q^n(x)\ \,|_{Q^n}. \end{aligned} \tag{37}$$

Similarly

$$-H^\beta(\epsilon, R_1\,|\,\boldsymbol{Q}) = (1-\epsilon)-\text{p}-\underline{\lim_{n\to\infty}} \frac{1}{n^\beta}(\log Q^n(x) + nR_1)\bigg|_{Q^n}. \tag{38}$$

In the following, we discuss the relationship between the aforementioned quantities and channel capacities.

VIII. General Asymptotic Formulas

A. General Case

Next, we consider the ϵ capacity and its related quantity, which are defined by

$$C_p(R\,|\,\boldsymbol{W}) \stackrel{\text{def}}{=} \inf_{\{\Phi_n\}_{n=1}^\infty}\left\{\overline{\lim_{n\to\infty}} P_{e,W^n}(\Phi_n)\,\middle|\, \underline{\lim_{n\to\infty}} \frac{1}{n}\log|\Phi_n| \ge R\right\}$$
$$C(\epsilon\,|\,\boldsymbol{W}) \stackrel{\text{def}}{=} \sup_{\{\Phi_n\}_{n=1}^\infty}\left\{\underline{\lim_{n\to\infty}} \frac{1}{n}\log|\Phi_n|\,\middle|\, \overline{\lim_{n\to\infty}} P_{e,W^n}(\Phi_n) \le \epsilon\right\}.$$

Concerning these quantities, the following general asymptotic formulas hold.

Theorem 6 [14], [16]: The relations

$$\begin{aligned} C_p(R\,|\,\boldsymbol{W}) &= \inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I_p(R-\gamma\,|\,\boldsymbol{P},\boldsymbol{W}) \\ &= \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p(R-\gamma\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \end{aligned} \tag{39}$$
$$\begin{aligned} C(\epsilon\,|\,\boldsymbol{W}) &= \sup_{\boldsymbol{P}} I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W}) \\ &= \sup_{\boldsymbol{P}} \inf_{\boldsymbol{Q}} J(\epsilon\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \end{aligned} \tag{40}$$

hold for $0 \le \epsilon < 1$ and $0 < \beta < 1$.

Remark 3: Historically, Verdú and Han [14] proved the first equation in (40). Hayashi and Nagaoka [16] established the second equation in (40) with $\epsilon = 0$ for the first time, even for the classical case, although their main topic was the quantum case. The relation (39) is proven for the first time in this paper.

Next, we proceed to the second-order coding rate. As a generalization of (3) and (4), we define the following:

$$C_p^\beta(R_2, R_1 \,|\, \boldsymbol{W}) \stackrel{\mathrm{def}}{=} \inf_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varlimsup_{n\to\infty} P_{e,W^n}(\Phi_n) \middle| \varliminf_{n\to\infty} \frac{1}{n^\beta}(\log|\Phi_n| - nR_1) \geq R_2 \right\} \tag{41}$$

$$C^\beta(\epsilon, R_1 \,|\, \boldsymbol{W}) \stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varliminf_{n\to\infty} \frac{1}{n^\beta}(\log|\Phi_n| - nR_1) \middle| \varlimsup_{n\to\infty} P_{e,W^n}(\Phi_n) \leq \epsilon \right\}. \tag{42}$$

Similar to Theorem 6, the following general formulas for the second-order coding rate hold.

Theorem 7: The relations

$$\begin{aligned} C_p^\beta(R_2, R_1 \,|\, \boldsymbol{W}) &= \inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I_p^\beta(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W}) \\ &= \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p^\beta(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \end{aligned} \tag{43}$$

$$\begin{aligned} C^\beta(\epsilon, R_1 \,|\, \boldsymbol{W}) &= \sup_{\boldsymbol{P}} I^\beta(\epsilon, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W}) \\ &= \sup_{\boldsymbol{P}} \inf_{\boldsymbol{Q}} J^\beta(\epsilon, R_1 \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \end{aligned} \tag{44}$$

hold for $0 \leq \epsilon < 1$ and $0 < \beta < 1$.

Indeed, Theorem 7 has greater significance than generalization. This theorem provides a unified viewpoint concerning the second-order asymptotic rate in channel coding and the following merits. First, it shortens the proof of Theorem 3. Second it enables us to extend Theorem 3 to the case of cost constraint. Third, it yields the extension to Gaussian noise case, which has continuous input signals. Fourth, it allows us to extend the same treatment to the Markovian case with the additive noise.

B. Cost Constraint

We focus on a sequence of cost function $\boldsymbol{c} = \{c_n\}_{n=1}^\infty$ where c_n is a function from $\mathcal{X}_n$ to $\mathbb{R}$. In this case, all alphabets are assumed to belong to the set

$$\mathcal{X}_{n,c,K} \stackrel{\mathrm{def}}{=} \left\{ x \in \mathcal{X}_n \middle| \sum_{i=1}^n c_n(x) \leq nK \right\}.$$

That is, our code $\{\Phi_n\}$ is assumed to satisfy that $\mathrm{supp}(\Phi_n) \subset \mathcal{X}_{n,c,K}$. Then, the capacities with cost constraint are given by

$$C_p(R \,|\, \boldsymbol{W}, \boldsymbol{c}, K) \stackrel{\mathrm{def}}{=} \inf_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varlimsup_{n\to\infty} P_{e,W^n}(\Phi_n) \middle| \varliminf_{n\to\infty} \frac{1}{n}\log|\Phi_n| \geq R, \ \mathrm{supp}(\Phi_n) \subset \mathcal{X}_{n,c,K} \right\}$$

$$C(\epsilon \,|\, \boldsymbol{W}, \boldsymbol{c}, K) \stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varliminf_{n\to\infty} \frac{1}{n}\log|\Phi_n| \middle| \varlimsup_{n\to\infty} P_{e,W^n}(\Phi_n) \leq \epsilon, \ \mathrm{supp}(\Phi_n) \subset \mathcal{X}_{n,c,K} \right\}$$

$$C_p^\beta(R_2, R_1 \,|\, \boldsymbol{W}, \boldsymbol{c}, K) \stackrel{\mathrm{def}}{=} \inf_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varlimsup_{n\to\infty} P_{e,W^n}(\Phi_n) \middle| \varliminf_{n\to\infty} \frac{1}{n^\beta}(\log|\Phi_n| - nR_1) \geq R_2, \ \mathrm{supp}(\Phi_n) \subset \mathcal{X}_{n,c,K} \right\}. \tag{45}$$

$$C^\beta(\epsilon, R_1 \,|\, \boldsymbol{W}, \boldsymbol{c}, K) \stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}_{n=1}^\infty} \left\{ \varliminf_{n\to\infty} \frac{1}{n^\beta}(\log|\Phi_n| - nR_1) \middle| \varlimsup_{n\to\infty} P_{e,W^n}(\Phi_n) \leq \epsilon, \ \mathrm{supp}(\Phi_n) \subset \mathcal{X}_{n,c,K} \right\}. \tag{46}$$

Concerning these quantities, the following general asymptotic formulas hold.

Theorem 8: [5], [16]: The relations

$$\begin{aligned} C_p(R \,|\, \boldsymbol{W}, \boldsymbol{c}, K) &= \inf_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,K}} \lim_{\gamma\downarrow 0} I_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{W}) \\ &= \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \end{aligned} \tag{47}$$

$$\begin{aligned} C(\epsilon \,|\, \boldsymbol{W}, \boldsymbol{c}, K) &= \sup_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,K}} I(\epsilon \,|\, \boldsymbol{P}, \boldsymbol{W}) \\ &= \sup_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,K}} \inf_{\boldsymbol{Q}} J(\epsilon \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \end{aligned} \tag{48}$$

hold for $0 \leq \epsilon < 1$ and $0 < \beta < 1$.

Remark 4: Historically, Han [5] proved the first equation in (48). Hayashi and Nagaoka [16] established the second equation in (48) with $\epsilon = 0$ for the first time, even for the classical case, although their main topic was the quantum case. The relation (47) is proven for the first time in this paper.

Similar to Theorem 7, the following general formulas for the second-order coding rate hold.

Theorem 9: The relations

$$\begin{aligned} C_p^\beta&(R_2, R_1 \,|\, \boldsymbol{W}, \boldsymbol{c}, K) \\ &= \inf_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,K}} \lim_{\gamma\downarrow 0} I_p^\beta(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W}) \\ &= \inf_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,K}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p^\beta(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \end{aligned} \tag{49}$$

$$C^\beta(\epsilon, R_1 \,|\, \boldsymbol{W}, \boldsymbol{c}, K)$$

$$
\begin{aligned}
&= \sup_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,K}} I^{\beta}(\epsilon, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W}) \\
&= \sup_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,K}} \inf_{\boldsymbol{Q}} J^{\beta}(\epsilon, R_1 \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \qquad (50)
\end{aligned}
$$

hold for $0 \le \epsilon < 1$.

The above theorems can be regarded as special cases of Theorems 6 and 7 by substituting the set $\mathcal{X}_{n,c,K}$ into the set $\mathcal{X}_n$. Hence, it is sufficient to show Theorems 6 and 7.

C. Additive Case

Next, we consider the case where the channel is given as a sequence of additive channel $\boldsymbol{W}(\boldsymbol{Q}) = \{W^n(Q^n)(y\,|\,x) = Q^n(y-x)\}$ on the set $\mathcal{X}^n$ with the cardinality d. Verdú and Han proved the following theorem.

Theorem 10: [14]: The relations

$$
C_p(R\,|\,\boldsymbol{W}(\boldsymbol{Q})) = 1 - \lim_{\gamma\downarrow 0} H_p(\log d - R + \gamma\,|\,\boldsymbol{Q}) \qquad (51)
$$

$$
C(\epsilon\,|\,\boldsymbol{W}(\boldsymbol{Q})) = \log d - H(1-\epsilon\,|\,\boldsymbol{Q}) \qquad (52)
$$

hold for $0 \le \epsilon < 1$. This theorem and (55) imply (54).

Remark 5: Verdú and Han proved (52) in the case of $\epsilon = 0$ at [14, eq. (7.2)]. Other cases are proven at the first time in this paper.

Similar to Theorem 10, the following formulas for the second-order coding rate hold for general additive channels.

Theorem 11: The relations

$$
C_p^{\beta}(R_2, R_1\,|\,\boldsymbol{W}) = 1 - \lim_{\gamma\downarrow 0} H_p^{\beta}(-R_2+\gamma, \log d - R_1\,|\,\boldsymbol{Q}) \qquad (53)
$$

$$
C^{\beta}(\epsilon, R_1\,|\,\boldsymbol{W}) = -H^{\beta}(1-\epsilon, \log d - R_1\,|\,\boldsymbol{Q}) \qquad (54)
$$

hold for $0 \le \epsilon < 1$. Hence, we obtain Theorem 4 from (32) and (33).

Now, using Theorems 6 and 7, we prove Theorems 10 and 11. Since $W_x^n(y) = Q^n(y-x)$, we have

$$
\begin{aligned}
I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W}) &= \epsilon\text{-p-}\varliminf_{n\to\infty} \frac{1}{n}\log\frac{W_x^n(y)}{W_{P^n}^n(y)}\bigg|_{\mathrm{P}_{P^n,W^n}} \\
&\le \epsilon\text{-p-}\varliminf_{n\to\infty} \frac{1}{n}\log W_x^n(y)\bigg|_{\mathrm{P}_{P^n,W^n}} \\
&\quad + \text{p-}\varlimsup_{n\to\infty} \frac{-1}{n}\log W_{P^n}^n(y)\bigg|_{\mathrm{P}_{P^n,W^n}} \qquad (55)\\
&\le \epsilon\text{-p-}\varliminf_{n\to\infty} \frac{1}{n}\log Q^n(x)\bigg|_{Q^n} + \log d \\
&= \log d - H(1-\epsilon\,|\,\boldsymbol{Q}) \qquad (56)
\end{aligned}
$$

where (55) and (56) follow from (35) and (37), respectively. Since the equality holds when P^n is the uniform distribution, we obtain

$$
\sup_{\boldsymbol{P}} I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W}) = \log d - H(1-\epsilon\,|\,\boldsymbol{Q})
$$

which implies (52). Similarly, we can show (54).

Since $\text{p-}\varlimsup_{n\to\infty}\frac{-1}{n}\log W_{P^n}^n(y)\,|_{W_{P^n}^n} \le d$, we have

$$
\begin{aligned}
&\varlimsup_{n\to\infty} \sum_{x\in\mathcal{X}_n} P^n(x) W_x^n\left\{\frac{1}{n}\log\frac{W_x^n(y)}{W_{P^n}^n(y)} < R\right\} \\
&\ge \varlimsup_{n\to\infty} \sum_{x\in\mathcal{X}_n} P^n(x) W_x^n\left\{\frac{1}{n}\log W_x^n(y) + \log d < R\right\} \\
&= \varlimsup_{n\to\infty} Q^n\left\{\frac{1}{n}\log Q^n(x) < R - \log d\right\} \\
&= 1 - \varliminf_{n\to\infty} Q^n\left\{\frac{-1}{n}\log Q^n(x) < \log d - R\right\}
\end{aligned}
$$

which implies that

$$
I_p(R\,|\,\boldsymbol{P},\boldsymbol{W}) \ge 1 - H_p(\log d - R\,|\,\boldsymbol{Q}).
$$

Thus, we obtain (51). Similarly, we obtain (55).

Remark 6: When the sets $\mathcal{X}_n$ and $\mathcal{Y}_n$ are given as general probability spaces with general σ-fields $\sigma(\mathcal{X}_n)$ and $\sigma(\mathcal{Y}_n)$, the above formulation can be extended with the following definition. The nth channel W^n is given by the real-valued function from $\mathcal{X}_n$ and $\sigma(\mathcal{Y}_n)$ satisfying the following conditions: i) for any $x \in \mathcal{X}_n$, W_x^n is a probability measure on $\mathcal{Y}_n$, and ii) for any $F \in \sigma(\mathcal{Y}_n)$, $W_{\cdot}^n(F)$ is a measurable function on $\mathcal{X}_n$. $\boldsymbol{P}$ and $\boldsymbol{Q}$ take values in sequence of probability measures on $\mathcal{X}_n$ and $\mathcal{Y}_n$, respectively. Then, the summands $\sum_{x\in\mathcal{X}_n} P^n(x)$ and $\sum_{y\in\mathcal{Y}_n} W_x^n(y)$ are replaced by $\int_{\mathcal{X}_n} P^n(dx)$ and $\int_{\mathcal{Y}_n} W_x^n(dy)$, respectively. For any distribution Q on $\mathcal{Y}_n$, the function $\frac{W_x^n(y)}{Q(y)}$ is replaced by the inverse of Radon–Nikodym derivative $\frac{dQ}{dW_x^n}(y)$ of Q with respect to W_x^n. In the above definitions, $\inf_{\boldsymbol{P}}$, $\sup_{\boldsymbol{P}}$, $\inf_{\boldsymbol{Q}}$, and $\sup_{\boldsymbol{Q}}$ are given as the infimum and supremum among all sequences of probability measures on $\{\mathcal{X}_n\}_{n=1}^{\infty}$ and $\{\mathcal{Y}_n\}_{n=1}^{\infty}$. The following proof is also valid in this extension.

IX. Proof of the General Formulas for the Second-Order Coding Rate

In this section, we prove Theorems 6 and 7. That is, for the reader's convenience, we present a proof for the first-order coding rate, as well as that for the second-order coding rate.

A. Direct Part

We prove the direct part, i.e., the inequalities

$$
C_p(R\,|\,\boldsymbol{W}) \le \inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I_p(R-\gamma\,|\,\boldsymbol{P},\boldsymbol{W}) \qquad (57)
$$

$$
C(\epsilon\,|\,\boldsymbol{W}) \ge \sup_{\boldsymbol{P}} I(\epsilon\,|\,\boldsymbol{P},\boldsymbol{W}) \qquad (58)
$$

$$
C_p^{\beta}(R_2, R_1\,|\,\boldsymbol{W}) \le \inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I_p^{\beta}(R_2-\gamma, R_1\,|\,\boldsymbol{P},\boldsymbol{W}) \qquad (59)
$$

$$
C^{\beta}(\epsilon, R_1\,|\,\boldsymbol{W}) \ge \sup_{\boldsymbol{P}} I^{\beta}(\epsilon, R_1\,|\,\boldsymbol{P},\boldsymbol{W}). \qquad (60)
$$

For arbitrary R, using the random coding method, we show that there exists a sequence of codes $\{\Phi_n\}$ such that $\frac{1}{n}\log|\Phi_n| \to R$ and $\varlimsup_{n\to\infty} P_{e,W^n}(\Phi_n) \le I_p(R\,|\,\boldsymbol{P},\boldsymbol{W})$. This method is essentially the same as Verdú and Han's method [14].

First, for $N_n \stackrel{\text{def}}{=} e^{nR - n^{\beta/2}}$, we focus on the random variables $Z_1, \ldots, Z_{N_n}$ subject to the distribution P^n. We assume that the random variable Z_i is independent of the random variable Z_j for $i \neq j$. Then, we define the code $\Phi_{n,\vec{Z}_n,R}$ with the size N_n depending on the random variable $\vec{Z}_n = (Z_1, \dots, Z_{N_n})$. We generate the encoder $\phi(\vec{Z}_n)$, in which $x \in \mathcal{X}_n$ is chosen as $\phi(\vec{Z}_n)(i)$ when $Z_i = x$. The decoder $\{\mathcal{D}_{i,\vec{Z}_n}\}_{i=1}^{N_n}$ is chosen by the following inductive method:

$$\mathcal{D}_{i,\vec{Z}_n,R} \stackrel{\text{def}}{=} \left\{ \frac{1}{n} \log \frac{W^n_{\phi(\vec{Z}_n)(i)}(y)}{W^n_{P^n}(y)} > R \right\} \setminus \left(\bigcup_{j=1}^{i-1} \mathcal{D}_{j,\vec{Z}_n} \right).$$

Thus, the average error probability is evaluated as

$$\begin{aligned}
&E_{\vec{Z}_n} P_{e,W^n}(\Phi_{n,\vec{Z}_n,R}) \\
&\le E_{\vec{Z}_n} \frac{1}{N_n} \sum_{i=1}^{N_n} W^n_{\phi(\vec{Z}_n)(i)} \\
&\quad \left(\left\{ \frac{1}{n} \log \frac{W^n_{\phi(\vec{Z}_n)(i)}(y)}{W^n_{P^n}(y)} > R \right\}^c \right. \\
&\quad\quad \left. \bigcup \left(\bigcup_{j=1}^{i-1} \left\{ \frac{1}{n} \log \frac{W^n_{\phi(\vec{Z}_n)(j)}(y)}{W^n_{P^n}(y)} > R \right\} \right) \right) \\
&\le E_{\vec{Z}_n} \frac{1}{N_n} \sum_{i=1}^{N_n} W^n_{\phi(\vec{Z}_n)(i)} \left\{ \frac{1}{n} \log \frac{W^n_{\phi(\vec{Z}_n)(i)}(y)}{W^n_{P^n}(y)} \le R \right\} \\
&\quad + E_{\vec{Z}_n} \frac{1}{N_n} \sum_{i=1}^{N_n} \sum_{j=1}^{i-1} W^n_{\phi(\vec{Z}_n)(i)} \left\{ \frac{1}{n} \log \frac{W^n_{\phi(\vec{Z}_n)(j)}(y)}{W^n_{P^n}(y)} \ge R \right\} \\
&= \sum_x P^n(x) W^n_x \left\{ \frac{1}{n} \log \frac{W^n_x(y)}{W^n_{P^n}(y)} \le R \right\} \\
&\quad + \frac{1}{N_n} \sum_{i=1}^{N_n} \sum_{j=1}^{i-1} E_{\vec{Z}_n} \left(E_{\vec{Z}_n} W^n_{\phi(\vec{Z}_n)(i)} \right) \\
&\quad \left\{ \frac{1}{n} \log \frac{W^n_{\phi(\vec{Z}_n)(j)}(y)}{W^n_{P^n}(y)} \ge R \right\}.
\end{aligned}$$

The second term is evaluated as

$$\begin{aligned}
&\frac{1}{N_n} \sum_{i=1}^{N_n} \sum_{j=1}^{i-1} E_{\vec{Z}_n} \left(E_{\vec{Z}_n} W^n_{\phi(\vec{Z}_n)(i)} \right) \left\{ \frac{1}{n} \log \frac{W^n_{\phi(\vec{Z}_n)(j)}(y)}{W^n_{P^n}(y)} \ge R \right\} \\
&= \frac{1}{N_n} \frac{N_n(N_n-1)}{2} \sum_x P(x) W^n_P \left\{ \frac{1}{n} \log \frac{W^n_x(y)}{W^n_{P^n}(y)} \ge R \right\} \\
&= \frac{N_n - 1}{2} \sum_x P(x) W^n_P \left\{ W^n_x(y) e^{-nR} \ge W^n_{P^n}(y) \right\} \\
&\le \frac{N_n}{2} e^{-nR} \le \frac{e^{-n^{\beta/2}}}{2} \to 0.
\end{aligned}$$

Any two sequences $\{a_n\}$ and $\{b_n\}$ satisfy the inequality $\overline{\lim}_{n\to\infty} a_n + b_n \le \overline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} b_n$. Hence, $\overline{\lim}_{n\to\infty} E_{\vec{Z}_n} P_{e,W^n}(\Phi_{n,\vec{Z}_n,R}) \le I_p(R \,|\, \boldsymbol{P}, \boldsymbol{W})$ because $\overline{\lim}_{n\to\infty} \sum_x P^n(x) W^n_x \{ \frac{1}{n} \log \frac{W^n_x(y)}{W^n_{P^n}(y)} \le R\} = I_p(R \,|\, \boldsymbol{P}, \boldsymbol{W})$. Thus, the convergence $\frac{1}{n} \log |N_n| \to R$ implies the inequality $C_p(R \,|\, \boldsymbol{W}) \le \inf_{\boldsymbol{P}} I_p(R \,|\, \boldsymbol{P}, \boldsymbol{W})$.

Next, in order to prove (57), for any sequence $\boldsymbol{P}$, we construct a code Φ_n such that $\overline{\lim}_{n\to\infty} P_{e,W^n}(\Phi_n) \le \lim_{\gamma\downarrow 0} I_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{W})$. For any k, we choose the integer N_k such that $E_{\vec{Z}_n} P_{e,W^n}(\Phi_{n,\vec{Z}_n,R-1/k}) \le I_p(R - 1/k \,|\, \boldsymbol{P}, \boldsymbol{W}) + 1/k$ for $\forall n \ge N_k$. Then, for any n, we choose $k(n)$ to be the maximum k satisfying $n \ge N_k$. Then, $k(n) \to \infty$ as $n \to \infty$. Thus, $E_{\vec{Z}_n} \Phi_{n,\vec{Z}_n,R-1/k(n)}$ goes to $\lim_{\gamma\downarrow 0} I_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{W})$, and $\frac{1}{n} \log |\Phi_{n,\vec{Z}_n,R-1/k(n)}|$ goes to R. Hence, we obtain the inequality $C_p(R \,|\, \boldsymbol{W}) \le \inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{W})$, i.e., (57).

For proving (59), we choose $N_n = e^{nR_1 + n^\beta R_2 - n^{\beta/2}}$. Substituting $nR_1 + n^\beta R_2$ into nR in the above discussion, we denote the code $\Phi_{n,\vec{Z}_n,R}$ by $\Phi_{n,\vec{Z}_n,R_1,R_2}$. Then

$$\begin{aligned}
&E_{\vec{Z}_n} P_{e,W^n} \left(\Phi_{n,\vec{Z}_n,R_1,R_2} \right) \\
&\le \sum_x P^n(x) W^n_x \left\{ \frac{1}{n^\beta} \left(\log \frac{W^n_{\phi(\vec{Z}_n)(i)}(y)}{W^n_{P^n}(y)} - nR_1 \right) < R_2 \right\} \\
&\quad + \frac{N_n}{2} e^{-(nR_1 + n^\beta R_2)}.
\end{aligned}$$

Since

$$\begin{aligned}
\frac{N_n}{2} e^{-(nR_1 + n^\beta R_2)} &\le \frac{e^{-n^{\beta/2}}}{2} \to 0 \\
\frac{1}{n^\beta} \log \frac{|N_n|}{e^{nR_1}} &\to R_2
\end{aligned}$$

we obtain the inequality $C^\beta_p(R_2, R_1 \,|\, \boldsymbol{W}) \le \inf_{\boldsymbol{P}} I^\beta_p(R_2, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W})$.

For any k, we choose the integer N_k such that $E_{\vec{Z}_n} P_{e,W^n}(\Phi_{n,\vec{Z}_n,R_1,R_2-1/k}) \le I^\beta_p(R_2 - 1/k, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W}) + 1/k$ for $\forall n \ge N_k$. Then, defining $k(n)$ similarly, we obtain $E_{\vec{Z}_n} \Phi_{n,\vec{Z}_n,R_1,R_2-1/k(n)} \to \lim_{\gamma\downarrow 0} I^\beta_p(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W})$, and

$$\frac{1}{n^\beta} \log \frac{|\Phi_{n,\vec{Z}_n,R_1,R_2-1/k(n)}|}{e^{nR_1}} \to R_2.$$

Hence, we obtain the inequality $C^\beta_p(R_2, R_1 \,|\, \boldsymbol{W}) \le \inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I^\beta_p(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W})$, i.e., (59).

For an arbitrary number $R < \sup_{\boldsymbol{P}} I(\epsilon \,|\, \boldsymbol{P}, \boldsymbol{W})$, there exists a sequence of input distributions $\boldsymbol{P}$ such that $I_p(R \,|\, \boldsymbol{P}, \boldsymbol{W}) \le \epsilon$. Therefore, the inequality (58) holds. Similarly, we can show the inequality (60).

B. Converse Part

Next, we prove the converse part, i.e.,

$$C_p(R \,|\, \boldsymbol{W}) \ge \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \tag{61}$$

$$C(\epsilon \,|\, \boldsymbol{W}) \le \sup_{\boldsymbol{P}} \inf_{\boldsymbol{Q}} J(\epsilon \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \tag{62}$$

$$C^\beta_p(R_2, R_1 \,|\, \boldsymbol{W}) \ge \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J^\beta_p(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \tag{63}$$

$$C^\beta(\epsilon, R_1 \,|\, \boldsymbol{W}) \le \sup_{\boldsymbol{P}} \inf_{\boldsymbol{Q}} J^\beta(\epsilon, R_1 \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \tag{64}$$

which completes our proof, because the other inequalities

$$\begin{aligned}
&\inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{W}) \le \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p(R - \gamma \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \\
&\sup_{\boldsymbol{P}} I(\epsilon \,|\, \boldsymbol{P}, \boldsymbol{W}) \ge \sup_{\boldsymbol{P}} \inf_{\boldsymbol{Q}} J(\epsilon \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W})
\end{aligned}$$

$$\inf_{\boldsymbol{P}} \lim_{\gamma\downarrow 0} I_p^\beta(R_2-\gamma, R_1 \,|\, \boldsymbol{P},\boldsymbol{W}) \le \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p^\beta(R_2-\gamma, R_1 \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$$

$$\sup_{\boldsymbol{P}} I^\beta(\epsilon, R_1 \,|\, \boldsymbol{P},\boldsymbol{W}) \ge \sup_{\boldsymbol{P}} \inf_{\boldsymbol{Q}} J^\beta(\epsilon, R_1 \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$$

are trivial based on their definitions. In the converse part, we essentially employ Hayashi and Nagaoka's [16] method. We choose an arbitrary sequence of codes $\{\Phi_n\}_{n=1}^\infty$. Let R be $\underline{\lim}_{n\to\infty}\frac{1}{n}\log|\Phi_n|$. Assume that the code Φ_n consists of the triplet $(N_n, \phi, \{\mathcal{D}_i\}_{i=1}^{N_n})$. Then, for any sequence of output distributions $\boldsymbol{Q} = \{Q^n\}_{n=1}^\infty$ and any real $\gamma > 0$, the inequality

$$P_{e,W^n}(\Phi_n) \ge \sum_{x\in\mathcal{X}^n} P_{\Phi_n}(x) W_x^n \left\{ \frac{1}{n}\log\frac{W_x^n(y)}{Q^n(y)} < R-\gamma \right\} - \frac{e^{n(R-\gamma)}}{N_n} \tag{65}$$

holds, where P_{Φ_n} is the empirical distribution for the $|\Phi_n|$ points $(\phi(1),\ldots,\phi(N_n))$.

Since $\frac{e^{n(R-\gamma)}}{N_n} \to 0$, the relation $\overline{\lim}_{n\to\infty} P_{e,W^n}(\Phi_n) \ge J_p(R-\gamma \,|\, \boldsymbol{P}', \boldsymbol{Q}, \boldsymbol{W})$ holds for any $\boldsymbol{Q}$, where $\boldsymbol{P}' = \{P_{\Phi_n}\}$. Thus, $\overline{\lim}_{n\to\infty} P_{e,W^n}(\Phi_n) \ge \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p(R-\gamma \,|\, \boldsymbol{P}', \boldsymbol{Q}, \boldsymbol{W})$. Therefore, $\overline{\lim}_{n\to\infty} P_{e,W^n}(\Phi_n) \ge \inf_{\boldsymbol{P}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p(R-\gamma \,|\, \boldsymbol{P}', \boldsymbol{Q}, \boldsymbol{W})$, which implies (61).

Now, assume that $\overline{\lim}_{n\to\infty} P_{e,W^n}(\Phi_n) = \epsilon$. Since $\frac{e^{n(R-\gamma)}}{N_n} \to 0$, (65) implies that $R-\gamma \le J(\epsilon \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$. Thus, $R-\gamma \le \sup_{\boldsymbol{P}}\inf_{\boldsymbol{Q}} J(\epsilon \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$. Since γ is an arbitrary positive real number, $R \le \sup_{\boldsymbol{P}}\inf_{\boldsymbol{Q}} J(\epsilon \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$, which implies (62).

Next, consider the case in which $\underline{\lim}_{n\to\infty}\frac{1}{n^\beta}\log\frac{|\Phi_n|}{e^{nR_1}} = R_2$. Replacing $R-\gamma$ by $R_1 + n^{\beta-1}(R_2-\gamma)$ in (65), we obtain $\frac{e^{nR_1+n^\beta(R_2-\gamma)}}{N_n} \to 0$. Thus, $\overline{\lim}_{n\to\infty} P_{e,W^n}(\Phi_n) \ge \inf_{\boldsymbol{P}}\sup_{\boldsymbol{Q}}\lim_{\gamma\downarrow 0} J_p^\beta(R_2-\gamma, R_1 \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})$, which implies (63). Replacing $R_1 + R_2 n^{\beta-1}$ into $R-\gamma$ in (65), similar to (62), we can show (64).

The inequality (65) is shown as follows. We focus on the inequalities

$$\begin{aligned}
&W_{\phi(i)}^n(\mathcal{D}_i) - e^{nR'}Q^n(\mathcal{D}_i) \\
&\le W_{\phi(i)}^n\left(\left\{W_{\phi(i)}^n(y) - e^{nR'}Q^n(y) \ge 0\right\}\right) \\
&\quad - e^{nR'}Q^n\left(\left\{W_{\phi(i)}^n(y) - e^{nR'}Q^n(y) \ge 0\right\}\right) \\
&\le W_{\phi(i)}^n\left(\left\{W_{\phi(i)}^n(y) - e^{nR'}Q^n(y) \ge 0\right\}\right) \\
&= W_{\phi(i)}^n\left\{\frac{1}{n}\log\frac{W_{\phi(i)}^n(y)}{Q^n(y)} \ge R'\right\}
\end{aligned}$$

where the first inequality follows from the fact that any two distributions P and Q and any positive constant a satisfy $\max_{\mathcal{D}}[P(\mathcal{D}) - aQ(\mathcal{D})] = P\{P(\omega) - aQ(\omega) \ge 0\} - aQ\{P(\omega) - aQ(\omega) \ge 0\}$.

Thus

$$\begin{aligned}
1 - P_{e,W^n}(\Phi_n) &= \frac{1}{N_n}\sum_{i=1}^{N_n} W_{\phi(i)}^n(\mathcal{D}_i) \\
&\le \frac{1}{N_n}\sum_{i=1}^{N_n} e^{nR'}Q^n(\mathcal{D}_i) \\
&\quad + W_{\phi(i)}^n\left\{\frac{1}{n}\log\frac{W_{\phi(i)}^n(y)}{Q^n(y)} \ge R'\right\} \\
&= \frac{e^{nR'}}{N_n} + 1 \\
&\quad - \sum_{x\in\mathcal{X}^n} P_{\Phi_n}(x) W_x^n\left\{\frac{1}{n}\log\frac{W_x^n(y)}{Q^n(y)} < R'\right\}
\end{aligned}$$

which implies (65).

X. Proof of the Stationary Memoryless Case

In this section, calculating the second-order information spectrum quantities in the stationary memoryless case with $\beta = \frac{1}{2}$, we prove Theorems 2, 3, and 5.

A. Proof of Theorem 2

In this section, using Theorem 7, we prove Theorem 2 when the cardinality $|\mathcal{X}|$ is finite. For this purpose, we show the following relations in the stationary discrete memoryless case, i.e., the case in which $W_{\vec{x}_n}^n(\vec{y}_n) = W_{\vec{x}_n}^{\times n}(\vec{y}_n) \stackrel{\text{def}}{=} \prod_{i=1}^n W_{x_i}(y_i)$ for $\vec{x}_n = (x_1,\ldots,x_n)$ and $\vec{y}_n = (y_1,\ldots,y_n)$. In this section, abbreviating C_W^{DM} as C, we will prove that

$$\inf_{\boldsymbol{P}}\lim_{\gamma\downarrow 0} I_p(R_2-\gamma, C \,|\, \boldsymbol{P},\boldsymbol{W}) \le \begin{cases} G\left(R_2/\sqrt{V_W^+}\right), & R_2 \ge 0 \\ G\left(R_2/\sqrt{V_W^-}\right), & R_2 < 0. \end{cases} \tag{66}$$

and

$$\inf_{\boldsymbol{P}}\sup_{\boldsymbol{Q}}\lim_{\gamma\downarrow 0} J_p(R_2-\gamma, C \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \ge \begin{cases} G\left(R_2/\sqrt{V_W^+}\right), & R_2 \ge 0 \\ G\left(R_2/\sqrt{V_W^-}\right), & R_2 < 0. \end{cases} \tag{67}$$

The situation of (66) is illustrated by Fig. 5. Showing both inequalities and using Theorem 7, we obtain

$$C_p(R_2, R_1 \,|\, \boldsymbol{W}) = \begin{cases} G\left(R_2/\sqrt{V_W^+}\right), & R_2 \ge 0 \\ G\left(R_2/\sqrt{V_W^-}\right), & R_2 < 0. \end{cases} \tag{68}$$

Since the right-hand side of (68) is continuous with respect to ϵ, (68) implies that

$$C(\epsilon, R_1 \,|\, \boldsymbol{W}) = \begin{cases} \sqrt{V_W^+}G^{-1}(\epsilon), & \epsilon \ge 1/2 \\ \sqrt{V_W^-}G^{-1}(\epsilon), & \epsilon < 1/2. \end{cases}$$

That is, we can show Theorem 2.

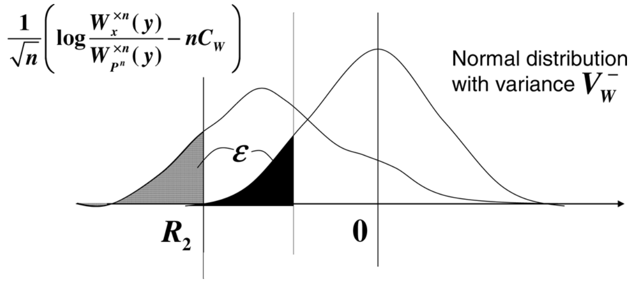


Fig. 5. Limiting behavior of $\frac{1}{\sqrt{n}}\left(\log\frac{W_x^{\times n}(y)}{W_{P^n}^{\times n}(y)}-nC\right)$ and the Gaussian distribution with the variance V_W^-.

In fact, when $\boldsymbol{P}$ is the i.i.d. of P_{M+} or P_{M-}, $I(\epsilon, C\,|\,\boldsymbol{P},\boldsymbol{W})$ is equal to $\sqrt{V_W^+}F^{-1}(\epsilon)$ or $\sqrt{V_W^-}F^{-1}(\epsilon)$. Thus, (66) holds. Therefore, the achievability part (the direct part) of Theorem 2 holds. Therefore, it is sufficient to prove the converse part (67).

We focus on the set T_n of empirical distributions with n outcomes. In this proof, using empirical distributions, we divide the probability space $\mathcal{X}^n$ into two parts, i.e., the main part and the marginal part. Its cardinality $|T_n|$ is evaluated as $|T_n|\le (n+1)^{|\mathcal{X}|}$. In this proof, we use the distribution

$$Q_U^n \stackrel{\text{def}}{=} \sum_{P\in T_n}\frac{1}{|T_n|+1}(W_P)^{\times n}+\frac{1}{|T_n|+1}Q_M^{\times n}$$

and the sets

$$\mathcal{V}_\epsilon \stackrel{\text{def}}{=} \{P\,|\,I(P,W)\ge C-\epsilon\}$$
$$\Omega_n \stackrel{\text{def}}{=} \{\vec{x}_n\in\mathcal{X}^n\,|\,\mathrm{ep}(\vec{x}_n)\in\mathcal{V}_\epsilon\}$$

where Q_M is given in Section II and $\mathrm{ep}(\vec{x}_n)$ is the empirical distribution of $\vec{x}_n\in\mathcal{X}^n$.

Since $Q_U^n(\vec{y}_n)\ge\frac{1}{|T_n|+1}(W_{\mathrm{ep}(\vec{x}_n)})^{\times n}(\vec{y}_n)$ and $Q_U^n(\vec{y}_n)\ge\frac{1}{|T_n|+1}Q_M^{\times n}(\vec{y}_n)$

$$\begin{aligned}
&\mathrm{P}_{P^n,W^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(\vec{y}_n)}-nC\right)\le R\right\}\\
&=\sum_{x\in\Omega_n}P^n(\vec{x}_n)\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(\vec{y}_n)}-nC\right)\le R\right\}\\
&\quad+\sum_{x\in\Omega_n^c}P^n(\vec{x}_n)\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(\vec{y}_n)}-nC\right)\le R\right\}\\
&\ge\sum_{x\in\Omega_n}P^n(\vec{x}_n)\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(Q_M)^{\times n}(\vec{y}_n)}\right.\right.\\
&\qquad\qquad\left.\left.+\log(|T_n|+1)-nC\right)\le R\right\}\\
&\quad+\sum_{\vec{x}_n\in\Omega_n^c}P^n(\vec{x}_n)\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(W_{\mathrm{ep}(\vec{x}_n)})^{\times n}(\vec{y}_n)}\right.\right.\\
&\qquad\qquad\left.\left.+\log(|T_n|+1)-nC\right)\le R\right\}.
\end{aligned}$$

When $\vec{x}_n\in\mathcal{V}_\epsilon^c$

$$\begin{aligned}
&\mathrm{V}_{W_{\vec{x}_n}^{\times n}}\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{\left(W_{\mathrm{ep}(\vec{x}_n)}\right)^{\times n}(\vec{y}_n)}-nC\right)\\
&=V_{\mathrm{ep}(x),W}<\max_P V_{P,W}\\
&\mathrm{E}_{W_{\vec{x}_n}^{\times n}}\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{\left(W_{\mathrm{ep}(\vec{x}_n)}\right)^{\times n}(\vec{y}_n)}+\log(|T_n|+1)-nC\right)\\
&=\frac{1}{\sqrt{n}}\left(nI(\mathrm{ep}(x),W)+\log(|T_n|+1)-nC\right)\\
&\le\frac{\log(|T_n|+1)}{\sqrt{n}}-\epsilon\sqrt{n}
\end{aligned}$$

where E_P and V_P denote the expectation and the variance under the distribution P. Thus, Chebyshev inequality implies

$$\begin{aligned}
&\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(W_{\mathrm{ep}(\vec{x}_n)})^{\times n}(\vec{y}_n)}+\log(|T_n|+1)-nC\right)\le R\right\}\\
&\ge 1-\frac{\max_P V_{P,W}}{R+\epsilon\sqrt{n}-\frac{\log(|T_n|+1)}{\sqrt{n}}}.
\end{aligned}$$

Define the quantity

$$V'_{P,W}\stackrel{\text{def}}{=}\mathrm{E}_P\mathrm{E}_{W_{\vec{x}_n}}\left(\log\frac{W_x(y)}{Q_M(y)}-D(W_x\,\|\,Q_M)\right)^2.$$

When $\vec{x}_n\in\mathcal{V}_\epsilon$, the inequality

$$\begin{aligned}
&\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(Q_M)^{\times n}(\vec{y}_n)}+\log(|T_n|+1)-nC\right)\le R\right\}\\
&\ge\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(Q_M)^{\times n}(\vec{y}_n)}+\log(|T_n|+1)\right.\right.\\
&\qquad\qquad\left.\left.-nI(\mathrm{ep}(\vec{x}_n),W)\right)\le R\right\}
\end{aligned}\tag{69}$$

holds. Since the random variable

$$\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(Q_M)^{\times n}(\vec{y}_n)}=\sum_{i=1}^n\log\frac{W_{x_i}(y_i)}{(Q_M)(y_i)}$$

has the variance $nV'_{\mathrm{ep}(\vec{x}_n),W}$

$$\begin{aligned}
&\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(Q_M)^{\times n}(\vec{y}_n)}+\log(|T_n|+1)\right.\right.\\
&\qquad\left.\left.-nI(\mathrm{ep}(\vec{x}_n),W)\right)\le R\right\}-G\left(\frac{R}{\sqrt{V'_{\mathrm{ep}(\vec{x}_n),W}}}\right)\\
&\to 0
\end{aligned}\tag{70}$$

as n goes to infinity. Since the random variable $\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(Q_M)^{\times n}(\vec{y}_n)}=\sum_{i=1}^n\log\frac{W_{x_i}(y_i)}{(Q_M)(y_i)}$ is written as a combination of finite number of random variables $\{\log\frac{W_x(y)}{(Q_M)(y)}\}_{x\in\mathcal{X}}$, this convergence is uniform concerning $\vec{x}_n$. Further

$$G\left(\frac{R}{\sqrt{V'_{\mathrm{ep}(\vec{x}_n),W}}}\right)\ge\min_{P\in\mathcal{V}_\epsilon}G\left(\frac{R}{\sqrt{V'_{P,W}}}\right).\tag{71}$$

Hence, from the combination of (69), (70), and (71), for any $\delta > 0$, there exists $N > 0$ such that for $n \geq N$

$$\mathrm{P}_{W^{\times n}_{\vec{x}_n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W^{\times n}_x(y)}{(Q_M)^{\times n}(\vec{y}_n)} + \log(|T_n|+1) - nC \right) \leq R \right\} \geq \min_{P \in \mathcal{V}_\epsilon} G\left(\frac{R}{\sqrt{V'_{P,W}}} \right) - \delta.$$

Therefore

$$\begin{aligned}
&\mathrm{P}_{P^n, W^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W^{\times n}_{\vec{x}_n}(\vec{y}_n)}{Q^n_U(\vec{y}_n)} - nC \right) \leq R \right\} \\
&\geq P^n(\Omega_n) \left(1 - \frac{\max_P V_{P,W}}{R + \epsilon\sqrt{n} - \frac{\log(|T_n|+1)}{\sqrt{n}}} \right) \\
&\quad + P^n(\Omega_n^c) \min_{P \in \mathcal{V}_\epsilon} G\left(\frac{R}{\sqrt{V'_{P,W}}} \right) - \delta \\
&\geq \min_{P \in \mathcal{V}_\epsilon} G\left(\frac{R}{\sqrt{V'_{P,W}}} \right) - \delta
\end{aligned}$$

where Ω_n^c is the complement of Ω_n.

Thus

$$\varlimsup_{n\to\infty} \mathrm{P}_{P^n, W^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W^{\times n}_{\vec{x}_n}(\vec{y}_n)}{Q^n_U(\vec{y}_n)} - nC \right) \leq R \right\} \geq \min_{P \in \mathcal{V}_\epsilon} G\left(\frac{R}{\sqrt{V_{P,W}}} \right) - \delta.$$

Since $\delta > 0$ and $\epsilon > 0$ are arbitrary, when $\boldsymbol{Q} = \{Q^n_U\}$

$$\begin{aligned}
&J_p(R, C \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \\
&= \varlimsup_{n\to\infty} \mathrm{P}_{P^n, W^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W^{\times n}_{\vec{x}_n}(\vec{y}_n)}{Q^n_U(\vec{y}_n)} - nC \right) \leq R \right\} \\
&\geq \min_{P \in \mathcal{V}} G\left(\frac{R}{\sqrt{V_{P,W}}} \right) = \begin{cases} G\left(R/\sqrt{V^+_W} \right), & R \geq 0 \\ G\left(R/\sqrt{V^-_W} \right), & R < 0 \end{cases}
\end{aligned}$$

which implies (67) because of the continuity of the right-hand side.

B. Proof of Theorem 3

In this section, using Theorem 9, we prove Theorem 3 when the cardinality $|\mathcal{X}|$ is finite. For this purpose, we show the following relations in the stationary discrete memoryless case, i.e., the case in which $W^n_{\vec{x}_n}(\vec{y}_n) = W^{\times n}_{\vec{x}_n}(\vec{y}_n) \overset{\mathrm{def}}{=} \prod_{i=1}^n W_{x_i}(y_i)$ for $\vec{x}_n = (x_1, \ldots, x_n)$ and $\vec{y}_n = (y_1, \ldots, y_n)$, and $c_n(\vec{x}_n) = \sum_{i=1}^n c(x_i)$. In this section, abbreviating C^{DM}_W as C, we will prove that

$$\begin{aligned}
&\inf_{\boldsymbol{P}: \mathrm{supp}(P_n) \subset \mathcal{X}_{n,c,K}} \lim_{\gamma \downarrow 0} I_p(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{W}) \\
&\leq \begin{cases} G\left(R_2/\sqrt{V^+_{W,c,K}} \right), & R_2 \geq 0 \\ G\left(R_2/\sqrt{V^-_{W,c,K}} \right), & R_2 < 0 \end{cases}
\end{aligned} \tag{72}$$

and

$$\begin{aligned}
&\inf_{\boldsymbol{P}: \mathrm{supp}(P_n) \subset \mathcal{X}_{n,c,K}} \sup_{\boldsymbol{Q}} \lim_{\gamma \downarrow 0} J_p(R_2 - \gamma, R_1 \,|\, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{W}) \\
&\geq \begin{cases} G\left(R_2/\sqrt{V^+_{W,c,K}} \right), & R_2 \geq 0 \\ G\left(R_2/\sqrt{V^-_{W,c,K}} \right), & R_2 < 0. \end{cases}
\end{aligned} \tag{73}$$

Showing both inequalities and using Theorem 9, we obtain

$$C_p(R_2, R_1 \,|\, \boldsymbol{W}, \boldsymbol{c}, K) = \begin{cases} G\left(R_2/\sqrt{V^+_{W,c,K}} \right), & R_2 \geq 0 \\ G\left(R_2/\sqrt{V^-_{W,c,K}} \right), & R_2 < 0. \end{cases} \tag{74}$$

Since the right-hand side of (74) is continuous with respect to ϵ, (74) implies that

$$C(\epsilon, R_1 \,|\, \boldsymbol{W}, \boldsymbol{c}, K) = \begin{cases} \sqrt{V^+_{W,c,K}} G^{-1}(\epsilon), & \epsilon \geq 1/2 \\ \sqrt{V^-_{W,c,K}} G^{-1}(\epsilon), & \epsilon < 1/2. \end{cases}$$

That is, we can show Theorem 3.

The inequality (73) can be proven in the same way as (67) by replacing T_n and Q_M by the set of empirical distributions $T_{n,c,K} \overset{\mathrm{def}}{=} \{P \in T_n \,|\, \mathrm{E}_P c(x) \leq K\}$ and $Q_{M,c,K}$. Therefore, the converse part of Theorem 3 holds. Therefore, it is sufficient to prove the direct part (72).

For any distribution P satisfying $\mathrm{E}_P c(\vec{x}_n) \leq K$, we choose the closet empirical distribution $P_n \in T_{n,c,K}$. Let $\boldsymbol{P} = \{P^n\}$ be the uniform distributions on the set $T_{P_n} \overset{\mathrm{def}}{=} \{\vec{x}_n \in \mathcal{X}^n \,|\, \mathrm{ep}(\vec{x}_n) = P_n\}$. It is sufficient to show that

$$I_p(R, C \,|\, \boldsymbol{P}, \boldsymbol{W}) \leq G\left(R/\sqrt{V_{P,W}} \right). \tag{75}$$

Since

$$P^n(\vec{x}_n) \leq |T_n| (P_n)^{\times n}(\vec{x}_n) \tag{76}$$

we have

$$\begin{aligned}
&I_p(R, C \,|\, \boldsymbol{P}, \boldsymbol{W}) \\
&= \varlimsup_{n\to\infty} \mathrm{P}_{P^n, W^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W^{\times n}_{\vec{x}_n}(\vec{y}_n)}{W^{\times n}_{P^n}(\vec{y}_n)} - nC \right) \leq R \right\} \\
&\leq \varlimsup_{n\to\infty} \mathrm{P}_{P^n, W^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W^{\times n}_{\vec{x}_n}(\vec{y}_n)}{(W_{P_n})^{\times n}(\vec{y}_n)} - \log|T_n| - nC \right) \leq R \right\} \\
&\leq G\left(\frac{R}{\sqrt{V_{P,W}}} \right)
\end{aligned} \tag{77}$$

which implies (75).

In order to prove (75) without condition $|\mathcal{X}| < \infty$, we choose a sequence of input distributions $\{P_{\pm}^{(k)}\}_{k=1}^{\infty}$ with finite supports such that

$$\begin{aligned} P^{(k)} &\in T_{n,c,K} \\ I(P_{\pm}^{(k)}, W) &\to \max_{P:\mathrm{E}_P c(x) \le K} I(P,W) \\ V_{P_{\pm}^{(k)},W} &\to V_{W,c,K}^{\pm}. \end{aligned}$$

Choose the distribution P^n as the uniform distributions on the set $T_{P^{(n^{\frac14})}}$. Then, instead of (76), the relation

$$P^n(\vec{x}_n) \le (n+1)^{n^{\frac14}} \left(P^{(n^{\frac14})}\right)^{\times n} (\vec{x}_n)$$

holds. Since $\frac{1}{\sqrt{n}} \log(n+1)^{n^{\frac14}}$ goes to zero, the same discussion as (77) yields (75).

C. Proof of Theorem 5

As is shown in Section X-B, we obtain the direct part, i.e.,

$$C_p^G\left(a, C_{N,S}^G \,|\, N, S\right) \le G\left(a/\sqrt{V_{P_M,W}}\right).$$

Hence, when $c_n(\vec{x}_n) = \sum_{i=1}^n x_i^2$, it is sufficient to prove

$$\inf_{\boldsymbol{P}:\mathrm{supp}(P_n)\subset\mathcal{X}_{n,c,S}} \sup_{\boldsymbol{Q}} \lim_{\gamma\downarrow 0} J_p(R_2-\gamma, R_1 \,|\, \boldsymbol{P},\boldsymbol{Q},\boldsymbol{W}) \ge G\left(a/\sqrt{V_{P_M,W}}\right). \quad (78)$$

For the following discussion, we define the empirical distribution $\mathrm{ep}(\vec{x}_n)$ in the continuous case

$$\mathrm{ep}(\vec{x}_n) \stackrel{\mathrm{def}}{=} \sum_{i=1}^n \frac{1}{n}\delta_{x_i}$$

where δ_x is the delta measure at x. In the following discussion, we use the distribution

$$\begin{aligned} Q_U^n &\stackrel{\mathrm{def}}{=} \frac12 \left(W_{P_M}\right)^{\times n} + \frac12 \left(W_{P_{M,\epsilon}}\right)^{\times n} \\ P_{M,\epsilon} &\stackrel{\mathrm{def}}{=} \frac{1}{\sqrt{2\pi(S-\epsilon)}} e^{-\frac{x^2}{2(S-\epsilon)}} \end{aligned}$$

and the sets

$$\begin{aligned} \mathcal{V}_\epsilon &\stackrel{\mathrm{def}}{=} \{P \,|\, \mathrm{E}_P x^2 \le S-\epsilon\} \\ \Omega_n &\stackrel{\mathrm{def}}{=} \{\vec{x}_n \in \mathcal{X}^n | \mathrm{ep}(\vec{x}_n) \in \mathcal{V}_\epsilon\} \end{aligned}$$

where we obtain

$$\begin{aligned} &\mathrm{P}_{P^n,W^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(y)} - nC \right) \le R \right\} \\ &= \sum_{\vec{x}_n\in\Omega_n} P^n(\vec{x}_n) \mathrm{P}_{W_{\vec{x}_n}^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(\vec{y}_n)} - nC \right) \le R \right\} \\ &\quad + \sum_{\vec{x}_n\in\Omega_n^c} P^n(\vec{x}_n) \mathrm{P}_{W_{\vec{x}_n}^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(\vec{y}_n)} - nC \right) \le R \right\} \\ &\ge \sum_{x\in\Omega_n} P^n(\vec{x}_n) \mathrm{P}_{W_{\vec{x}_n}^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{\left(W_{P_M}\right)^{\times n}(\vec{y}_n)} + \log 2 - nC \right) \le R \right\} \\ &\quad + \sum_{x\in\Omega_n^c} P^n(\vec{x}_n) \mathrm{P}_{W_{\vec{x}_n}^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{\left(W_{P_{M,\epsilon}}\right)^{\times n}(y)} + \log 2 - nC \right) \le R \right\}. \end{aligned}$$

When $\vec{x}_n \in \mathcal{V}_\epsilon^c$, the random variable

$$\frac{1}{\sqrt{n}} \left(\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(W_{P_{M,\epsilon}})^{\times n}(\vec{y}_n)} + \log 2 - nC \right)$$

has the expectation

$$\begin{aligned} &\frac{1}{\sqrt{n}} \left(\frac{n}{2}\log\left(1+\frac{S-\epsilon}{N}\right) + \frac{\frac{\|\vec{x}_n\|^2}{nN} - \frac{S-\epsilon}{N}}{2\left(1+\frac{S-\epsilon}{N}\right)} - \frac{n}{2}\log\left(1+\frac{S}{N}\right) + \log 2 \right) \\ &\le \frac{\log 2}{\sqrt{n}} - \frac{\sqrt{n}}{2} \log \frac{1+\frac{S}{N}}{1+\frac{S-\epsilon}{N}} \end{aligned}$$

and the variance

$$\frac{\frac{(S-\epsilon)^2}{N^2} + 2\frac{\|\vec{x}_n\|^2}{nN}}{2\left(1+\frac{S-\epsilon}{N}\right)^2} \le \frac{\frac{(S-\epsilon)^2}{N^2} + 2\frac{S-\epsilon}{N}}{2\left(1+\frac{S-\epsilon}{N}\right)^2}.$$

Thus, Chebyshev inequality implies

$$\begin{aligned} &\mathrm{P}_{W_{\vec{x}_n}^{\times n}} \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{\left(W_{P_{M,\epsilon}}\right)^{\times n}(\vec{y}_n)} + \log 2 - nC \right) \le R \right\} \\ &\ge 1 - \frac{\frac{\frac{(S-\epsilon)^2}{N^2} + 2\frac{S-\epsilon}{nN}}{2(1+\frac{S-\epsilon}{N})^2}}{R + \frac{\sqrt{n}}{2}\log\frac{1+\frac{S}{N}}{1+\frac{S-\epsilon}{N}} - \frac{\log 2}{\sqrt{n}}} \to 1. \end{aligned}$$

When $x \in \mathcal{V}_\epsilon$, under the n-variable Gaussian distribution $W_{\vec{x}_n}^{\times n}$, the random variable $\log \frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n+\vec{x}_n)}{(W_{P_M})^{\times n}(\vec{y}_n+\vec{x}_n)}$ is calculated to be

$$\frac{1}{2\left(1+\frac{S}{N}\right)} \left(-\frac{S\|\vec{y}_n\|^2}{N^2} + \frac{2\vec{x}_n\cdot\vec{y}_n}{N} + \frac{\|\vec{x}_n\|^2}{N} \right) \frac{n}{2}\log\left(1+\frac{S}{N}\right).$$

The expectation is

$$\frac{\frac{\|\vec{x}_n\|^2}{N} - n\frac{S}{N}}{2\left(1+\frac{S}{N}\right)} + \frac{n}{2}\log\left(1+\frac{S}{N}\right)$$

and the variance is

$$\frac{2n\frac{S^2}{N^2}+4\frac{\|\vec{x}_n\|^2}{N}}{4\left(1+\frac{S}{N}\right)^2}.$$

The random variable

$$\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n+\vec{x}_n)}{(W_{P_M})^{\times n}(\vec{y}_n+\vec{x}_n)} - \frac{\frac{\|x\|^2}{N}-n\frac{S}{N}}{2\left(1+\frac{S}{N}\right)} - \frac{n}{2}\log\left(1+\frac{S}{N}\right)\right)$$

converges to the normal distribution when n goes to infinity. Due to the property of Gaussian distribution, this convergence is uniform when $\|x\|$ is bounded. Hence

$$\begin{aligned}&\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(W_{P_M})^{\times n}(\vec{y}_n)}+\log 2-nC\right)\le R\right\}\\&\ge \mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(W_{P_M})^{\times n}(\vec{y}_n)}+\log 2 - \frac{\frac{\|\vec{x}_n\|^2}{N}-n\frac{S}{N}}{2\left(1+\frac{S}{N}\right)} - \frac{n}{2}\log\left(1+\frac{S}{N}\right)\right)\le R\right\}\end{aligned} \tag{79}$$

and

$$G\left(\frac{R}{\sqrt{\frac{2\frac{S^2}{N^2}+4\frac{\|\vec{x}_n\|^2}{nN}}{4\left(1+\frac{S}{N}\right)^2}}}\right)\ge\begin{cases}G\left(\frac{R}{\sqrt{\frac{2\frac{S^2}{N^2}+4\frac{S-\epsilon}{N}}{4\left(1+\frac{S}{N}\right)^2}}}\right), & R\le 0\\ G\left(\frac{R}{\sqrt{\frac{2\frac{S^2}{N^2}+4\frac{S}{N}}{4\left(1+\frac{S}{N}\right)^2}}}\right), & R>0.\end{cases} \tag{80}$$

Similarly to (70), we obtain the uniform convergence

$$\mathrm{P}_{W_{\vec{x}_n}^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{(W_{P_M})^{\times n}(\vec{y}_n)}+\log 2 - \frac{\frac{\|\vec{x}_n\|^2}{N}-n\frac{S}{N}}{2\left(1+\frac{S}{N}\right)} - \frac{n}{2}\log\left(1+\frac{S}{N}\right)\right)\le R\right\} - G\left(\frac{R}{\sqrt{\frac{2\frac{S^2}{N^2}+4\frac{\|\vec{x}_n\|^2}{nN}}{4\left(1+\frac{S}{N}\right)^2}}}\right)\to 0. \tag{81}$$

Therefore, the combination of (79), (80), and (81) yields that

$$\varliminf_{n\to\infty}\mathrm{P}_{P^n,W^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(\vec{y}_n)}-nC\right)\le R\right\} \ge\begin{cases}G\left(\frac{R}{\sqrt{\frac{2n\frac{S^2}{N^2}+4n\frac{S-\epsilon}{N}}{4\left(1+\frac{S}{N}\right)^2}}}\right), & R\le 0\\ G\left(\frac{R}{\sqrt{\frac{2n\frac{S^2}{N^2}+4n\frac{S}{N}}{4\left(1+\frac{S}{N}\right)^2}}}\right), & R>0.\end{cases}$$

Since $\epsilon>0$ is arbitrary, when $\boldsymbol{Q}=\{Q_U^n\}$

$$\begin{aligned}&J_p(R,C\,|\,\boldsymbol{P},\boldsymbol{Q},\boldsymbol{W})\\&=\varliminf_{n\to\infty}\mathrm{P}_{P^n,W^{\times n}}\left\{\frac{1}{\sqrt{n}}\left(\log\frac{W_{\vec{x}_n}^{\times n}(\vec{y}_n)}{Q_U^n(\vec{y}_n)}-nC\right)\le R\right\}\\&\ge G\left(\frac{R}{\sqrt{\frac{2\frac{S^2}{N^2}+4\frac{S}{N}}{4\left(1+\frac{S}{N}\right)^2}}}\right)\end{aligned}$$

which implies (78).

XI. Concluding Remarks and Future Study

We have obtained a general asymptotic formula for channel coding in the sense of the second-order coding rate. That is, it has been shown that the optimum second-order transmission rate with the error probability ϵ is characterized by the second-order asymptotic behavior of the logarithmic likelihood ratio between the conditional output distribution and the nonconditional output distribution. Using this result, we have derived this type of optimal transmission rate for the discrete memoryless case, the discrete memoryless case with a cost constraint, the additive Markovian case, and the Gaussian channel case with an energy constraint. The performance in the second-order coding rate is characterized by the average of the variance of the logarithmic likelihood ratio with the single letterized expression. When the input distribution producing the capacity is not unique, it is characterized by its minimum and its maximum. We give a typical example such that the minimum is different from the maximum. Furthermore, both quantities have been verified to satisfy the additivity.

The main results of this study are as follows. While the application of the information spectrum method to the second-order coding rate was initiated by Hayashi [6], his research indicated that there is no difficulty in extending general formulas to the second-order coding rate. Therefore, in the i.i.d. case, the second-order coding rate of the source coding and intrinsic randomness are solved by the central limit theorem. However, channel coding cannot been treated using the method of Hayashi [6] except for the additive noise case with no cost constraint because the present problem contains the optimization concerning the input distribution in the nonadditive noise case. In the converse part, we have to treat the general sequence of input distributions. In order to resolve this difficulty, we have treated the logarithmic likelihood ratio between the conditional output distribution and the distribution Q_U^n, which is introduced in Section X-A.

Furthermore, we can consider the quantum extension of our results. There is considerable difficulty concerning noncommutativity in this direction. In addition, the third-order coding rate is expected but appears difficult. The second order is the order $\sqrt{n}$, and it is not clear whether the third order is a constant order or the order $\log n$. This is an interesting problem for future study.

Appendix I
Proof of the Second Equation of (1)

Recall the following minimax theorem.

Lemma 4: Consider two vector spaces V_1 and V_2 and consider a real-valued function $f(v_1, v_2)$ with the domain $V_1 \times V_2$. If f is convex with respect to v_2 and concave with respect to v_1, then (see [17, Ch. VI, Prop. 2.3])[2]

$$\max_{v_1 \in S_1} \min_{v_2 \in S_2} f(v_1, v_2) = \min_{v_2 \in S_2} \max_{v_1 \in S_1} f(v_1, v_2)$$

where S_1 and S_2 are convex subsets of V_1 and V_2.

Since $J(P,Q,W) - I(P,W) = D(W_p \,\|\, Q) \geq 0$, $I(P,W) = \min_{Q \in \mathcal{P}(\mathcal{Y})} J(P,Q,W)$. The joint convexity of the divergence implies that $Q \mapsto D(W_x \| Q)$ is convex. Thus, Lemma 4 can be applied. Therefore, we obtain

$$\begin{aligned}\max_{P \in \mathcal{P}(\mathcal{X})} I(P,W) &= \max_{P \in \mathcal{P}(\mathcal{X})} \min_{Q \in \mathcal{P}(\mathcal{Y})} J(P,Q,W) \\ &= \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{P \in \mathcal{P}(\mathcal{X})} J(P,Q,W) \\ &= \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D(W_x \| Q).\end{aligned}$$

Now, the second equation of (1) is proved.

Appendix II
Proof of (16)

For a given $R < 0$, we prove (16). Since $\frac{d^2\psi_P}{ds^2}(s) > 0$, the function ψ_P is convex. Choosing s_n such that

$$C_W^{\mathrm{DM}} + \frac{R_2}{\sqrt{n}} = -\frac{d\psi_P}{ds}(s_n) = -\frac{d\psi_P}{ds}(0) - \int_0^{s_n} \frac{d^2\psi_P}{ds^2}(t)dt$$

we have the relation

$$\frac{R_2}{\sqrt{n}} = -\int_0^{s_n} \frac{d^2\psi_P}{ds^2}(t)dt. \tag{82}$$

Then, the minimum of $C_W^{\mathrm{DM}} s + \frac{R_2}{\sqrt{n}} s + \psi_P(s)$ is attained when $s = s_n$. Since $\frac{d^2\psi_P}{ds^2}(s)$ is continuous and bounded, s_n approaches zero as n goes to infinity. More precisely, (82) implies

$$R_2 = -\lim_{n\to\infty} \sqrt{n} \int_0^{s_n} \frac{d^2\psi_P}{ds^2}(t)dt = -\lim_{n\to\infty} (\sqrt{n} s_n) \frac{d^2\psi_P}{ds^2}(0).$$

That is

$$\lim_{n\to\infty} (\sqrt{n} s_n) = \frac{-R_2}{\frac{d^2\psi_P}{ds^2}(0)}.$$

When the function $\epsilon(u)$ is chosen to be $\frac{d^2\psi_P}{ds^2}(u) - \frac{d^2\psi_P}{ds^2}(0)$, $\epsilon(u)$ approaches zero as u goes to zero.

Thus, we have

$$\begin{aligned}& n \min_{0 \leq s \leq 1} \left(C_W^{\mathrm{DM}} s + \frac{R_2}{\sqrt{n}} s + \psi_P(s) \right) \\ &= n \left(C_W^{\mathrm{DM}} s_n + \frac{R_2}{\sqrt{n}} s_n + \psi_P(s_n) \right) \\ &= n \left(\frac{R_2}{\sqrt{n}} s_n + \int_0^{s_n} \int_0^t \frac{d^2\psi_P}{ds^2}(u) du dt \right) \\ &= \sqrt{n} R_2 s_n + n \frac{s_n^2}{2} \frac{d^2\psi_P}{ds^2}(0) + n \int_0^{s_n} \int_0^t \epsilon(u) du dt \\ &\to \frac{-R_2^2}{2 \frac{d^2\psi_P}{ds^2}(0)}\end{aligned}$$

which implies (16).

Appendix III
Proof of Lemma 1

For this proof, we define the maps $\mathcal{E}_A$ and $\mathcal{E}_B$ as

$$\begin{aligned}(\mathcal{E}_A Q)(a,b) &\stackrel{\mathrm{def}}{=} P^A(a) Q^{B|A}(b|a) \\ (\mathcal{E}_B Q)(a,b) &\stackrel{\mathrm{def}}{=} P^B(b) Q^{A|B}(a|b)\end{aligned}$$

where $P^A(0) = 1/2$ and $P^B(0) = q_1$. When the distribution Q' satisfies that ${Q'}^A = P^A$, the following Pythagorean-type inequality holds:

$$D(Q' \| Q) = D(Q' \| \mathcal{E}_A(Q)) + D(\mathcal{E}_A(Q) \| Q). \tag{83}$$

Similarly, when the distribution Q' satisfies that ${Q'}^B = P^B$, the following Pythagorean-type inequality holds:

$$D(Q' \| Q) = D(Q' \| \mathcal{E}_B(Q)) + D(\mathcal{E}_B Q \| Q). \tag{84}$$

Define

$$\begin{aligned}Q_{2k} &\stackrel{\mathrm{def}}{=} \underbrace{\mathcal{E}_B \circ \mathcal{E}_A \circ \cdots \circ \mathcal{E}_B \circ \mathcal{E}_A}_{2k} Q \\ Q_{2k+1} &\stackrel{\mathrm{def}}{=} \underbrace{\mathcal{E}_A \circ \mathcal{E}_B \circ \mathcal{E}_A \circ \cdots \circ \mathcal{E}_B \circ \mathcal{E}_A}_{2k+1} Q.\end{aligned}$$

[2] This relation holds even if V_1 is infinite dimensional, as long as S_2 is a closed and bounded set.

Then, $D(Q_{2k+1}\|Q_{2k}) = D(\mathcal{E}_A Q_{2k}\|\mathcal{E}_A Q_{2k-1}) \leq D(Q_{2k}\|Q_{2k-1})$, and $D(Q_{2k}\|Q_{2k-1}) \leq D(Q_{2k-1}\|Q_{2k-2})$. For any $Q' \in \mathcal{Z}_1$, we have

$$D(Q'\|Q) = D(Q'\|Q_n) + \sum_{k=1}^{n} D(Q_k\|Q_{k-1})$$

i.e.,

$$\sum_{k=1}^{n} D(Q_k\|Q_{k-1}) \leq D(Q'\|Q)$$

which implies

$$\sum_{k=1}^{\infty} D(Q_k\|Q_{k-1}) \leq D(Q'\|Q).$$

Thus, $D(Q_k\|Q_{k-1})$ converges to zero. Therefore, there exists a distribution Q_∞ such that $Q_k \to Q_\infty$. Hence

$$D(Q'\|Q) = D(Q'\|Q_\infty) + \sum_{k=1}^{\infty} D(Q_k\|Q_{k-1})$$

which implies (22).

Further, for any $P_2 \in \mathcal{Z}_2$, we assume that Q satisfies $Q^A = P^A$. Since the concavity of log implies the inequality $\log \sum_a P^A(a) Q^{B|A}(b|a) \geq \sum_a P^A(a) \log Q^{B|A}(b|a)$, the following Pythagorean-type inequality holds:

$$\begin{aligned} D(P_2\|Q) &= H(P_2) - \sum_a \sum_b P_2^A(a) P_2^B(b) \log Q(a,b) \\ &= H(P_2) - \sum_a P_2^A(a) \log Q^A(a) \\ &\quad - \sum_a \sum_b P_2^A(a) P_2^B(b) \log Q^{B|A}(b|a) \\ &= H(P_2) - \sum_a P_2^A(a) \log Q^A(a) \\ &\quad - \sum_b P_2^B(b) \log Q^B(b) + \sum_b P_2^B(b) \log Q^B(b) \\ &\quad - \sum_a \sum_b P_2^A(a) P_2^B(b) \log Q^{B|A}(b|a) \\ &= D(P_2\|P_2^A \times P_2^B) + \sum_b P_2^B(b) \log Q^B(b) \\ &\quad - \sum_b P_2^B(b) \sum_a P_2^A(a) \log Q^{B|A}(b|a) \\ &= D\left(P_2\|P_2^A \times P_2^B\right) \\ &\quad + \sum_b P_2^B(b) \left(\log \sum_a P^A(a) Q^{B|A}(b|a) \right. \\ &\qquad \left. - \sum_a P^A(a) \log Q^{B|A}(b|a) \right) \\ &\geq D(P_2\|P_2^A \times P_2^B). \end{aligned} \tag{85}$$

Combination of (84) and (85) yields (23).

Acknowledgment

The author would like to thank Prof. T. Uyematsu for informing him about [3]. He would also like to thank the referees and the editor for helpful comments concerning this manuscript.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[2] M. Hayashi, "Practical evaluation of security for quantum key distribution," *Phys. Rev. A, Gen. Phys.*, vol. 74, 2006, 022307.

[3] V. Strassen, "Asymptotische abschätzugen in Shannon's informationstheorie," in *Proc. 3rd Conf. Inf. Theory*, Prague, Czech Republic, 1962, pp. 689–723.

[4] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[5] T.-S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003.

[6] M. Hayashi, "Second order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4619–4637, Oct. 2008.

[7] S. Watanabe, R. Matsumoto, and T. Uyematsu, "Noise tolerance of the BB84 protocol with random privacy amplification," *Int. J. Quant. Inf.*, vol. 4, pp. 935–946, 2006.

[8] J. Hasegawa, M. Hayashi, T. Hiroshima, A. Tanaka, and A. Tomita, "Experimental decoy state quantum key distribution with unconditional security incorporating finite statistics," [Online]. Available: arXiv:0705.3081

[9] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379– 423 and 623–656, 1948.

[10] I. Csiszár and J. Koürner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[11] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.

[12] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*. Berlin, Germany: Springer-Verlag, 1997.

[13] Y. Kabashima and D. Saad, "Statistical mechanics of low-density parity-check codes," *J. Phys. A, Math. Gen.*, vol. 37, pp. R1–R43, 2004.

[14] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[15] M. Ohya, D. Petz, and N. Watanabe, "On capacities of quantum channels," *Prob. Math. Statist.*, vol. 17, pp. 179–196, 1997.

[16] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1753–1768, Jul. 2003.

[17] I. Ekeland and R. Téman, *Convex Analysis and Variational Problems*. Amsterdam, The Netherlands: North-Holland, 1976.

Masahito Hayashi was born in Japan in 1971. He received the B.S. degree from the Faculty of Sciences, Kyoto University, Kyoto, Japan, in 1994 and the M.S. and Ph.D. degrees in mathematics from Kyoto University, Japan, in 1996 and 1999, respectively.

He was a Research Fellow of the Japan Society of the Promotion of Science (JSPS), Kyoto University, from 1998 to 2000, and worked in the Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN, from 2000 to 2003, and in ERATO Quantum Computation and Information Project, Japan Science and Technology Agency (JST) as the Research Head from 2000 to 2006. He also worked in the Superrobust Computation Project Information Science and Technology Strategic Core (21st Century COE by MEXT) Graduate School of Information Science and Technology, The University of Tokyo, as Adjunct Associate Professor from 2004 to 2007. Since 2009, he also has been a Visiting Research Associate Professor at the Centre for Quantum Technologies, National University of Singapore, Singapore. He published the book *Quantum Information: An Introduction* (New York: Springer-Verlag, 2006). In 2007, he joined the Graduate School of Information Sciences, Tohoku University, Sendai, Japan, as Associate Professor. His research interests include quantum information theory and quantum statistical inference.

Dr. Hayashi is on the Editorial Board of the *International Journal of Quantum Information* and the *International Journal on Advances in Security*.