Quantum universal coding protocols and universal approximation of multi-copy states

Masahito Hayashi
Graduate School of Information Sciences, Tohoku University, Sendai, 980-8579, Japan
E-mail: hayashi@math.is.tohoku.ac.jp

Abstract. We have constructed universal codes for quantum lossless source coding and classical-quantum channel coding. In this construction, we essentially employ group representation theory. In order to treat quantum lossless source coding, universal approximation of multi-copy states is discussed in terms of the quantum relative entropy.

1. Introduction

Main limiting theorems in information theory were proven by Shannon [1], in which the protocols constructed depend on the distributions of the information source/the communication channels. Later, many researchers improved Shannon’s original results so that the protocol to achieve the optimal rate does not depend on the distributions of the information source/the communication channels. Such protocols are called universal, and Csiszár and Körner [2] developed the type method as a unified method to treat them. For the channel coding, they showed that their code has the same exponential rate of the error probability as the Gallager’s exponent [3], which is the optimal one among known codes in the stationary memoryless case. For the source coding, universal coding was discussed for the both cases of fixed-length and variable-length when the information source is given as an independent and identical distribution. In the fixed-length case, the universal code attains the optimal exponential rate of the error probability[2]. Clarke and Barron [4] treated the variable-length case in the following way. Since the average coding length is given by the Shannon entropy when the distribution of information source is known, they regarded the difference between the average coding length and the Shannon entropy as the redundancy of the given code, and studied its asymptotic behavior. Thanks to Kraft inequality, there exists a one-to-one correspondence between a variable-length prefix code and a probability distribution. Hence, the redundancy is given as the relative entropy between the distribution of information source and the distribution corresponding to the code.

Consider the set of commutative densities \( \{ \rho_\theta | \theta \in \Theta \subset \mathbb{R}^m \} \) on \( \mathcal{H} := \mathbb{C}^d \), and a prior distribution \( w(\theta) \) on \( \Theta \). Clarke and Barron [4] showed that the mixture state \( \rho_{w,n} := \int_\Theta \rho_\theta w(\theta) d\theta \) satisfies the relation
\[
D(\rho_{\theta}^\otimes n || \sigma_{w,n}) \approx \frac{m}{2} \log \frac{n}{2\pi e} + \log \sqrt{\det I(\theta)} - \log w(\theta),
\]
where \( I(\theta) \) is the Fisher information matrix (See [5]). In this paper, logarithms are taken to base 2. In particular, Clarke and Barron [6] proved that the mini-max of the constant term is given when the prior \( w \) is chosen as the Jeffreys’ prior \( \frac{\sqrt{\det I(\theta')}}{\int \sqrt{\det I(\theta')} d\theta'} \).
Concerning the quantum system, several universal protocols were given by several papers. Jozsa et al. [7] constructed a universal fixed-length source coding, which depends only on the compression rate and attains the minimum compression rate. Hayashi [8] proved that their code attains the optimal exponential rate of the error probability. Further, Hayashi and Matsumoto [9] constructed a universal variable-length source coding in the quantum system. In their formulation, the quantum state is compressed after the decision of the compression rate. That is, a superposition among states with different coding length is not allowed. Hence, since the measurement is necessary for the decision of the coding length, the quantum state is inevitably demolished. They proved that it is possible to decrease the degree of state demolition to infinitesimal and compress the state with the optimal compression rate even though the entropy rate of the information source is unknown. Therefore, it is impossible to treat the optimization of compression protocol in the same way as Clarke and Barron [4] in this scenario.

In this paper, we consider a different formulation of quantum variable-length source coding, in which a superposition among different coding lengths is allowed. In this formulation, there is a one-to-one correspondence between codes and density matrices. Hence, when the information ensemble is known, the optimal average compression rate is given by its von Neumann entropy.

Hence, the redundancy is given by the quantum relative entropy between the density of information source and the density corresponding to the code. Concerning the quantum case of this problem, in 1996, Krattenthaler and Slater[10] derived its quantum extension for the qubit case. However, the general case has remained an open problem for more than ten years. Their paper did not provide a complete solution to the mini-max problem for the constant term.

As one of main results, we prove the existence of states $\sigma_n$ on $\mathcal{H}^\otimes n$ satisfying

$$D(\rho^{\otimes n}\|\sigma_n) \approx \frac{d^2-1}{2} \log n + O(1)$$  \hspace{1cm} (2)

for all faithful states $\rho$ on $\mathcal{H}$, i.e., states $\rho$ in the set $\mathcal{S} := \{\rho|\text{rank } \rho = d\}$. Since the dimension of the state family $\mathcal{S}$ is $d^2 - 1$, the relation (2) can be regarded as a natural quantum extension of (1). More precisely, we calculate the following mini-max value

$$\min_{\{\sigma_n\}} \sup_{\rho \in \mathcal{S}} \lim_{n \to \infty} \left(D(\rho^{\otimes n}\|\sigma_n) - \frac{d^2-1}{2} \log n\right),$$  \hspace{1cm} (3)

which is one of the main results in this paper. Krattenthaler and Slater[10] treated the same problem for a restricted class of states $\{\sigma_n\}$ in the qubit case.

As the other main result, this paper treats a universal code for quantum channel. That is, we construct a universal coding for a classical-quantum channel, which attains the quantum mutual information and depends only on the coding rate and the ‘type’ of the input system. Note that any universal channel code does not attain the channel capacity and only attain the mutual information even in the classical case because the construction depends on the type of the input system [2]. In the proposed construction, the following three methods play essential roles.

The first method is the information spectrum method, which is essential for construction of the decoder. In the information spectrum method, the decoder is constructed by the square root measurement of the projectors given by the quantum analogue of the likelihood ratio between the signal state and the mixture state[11, 12]. The second method is the irreducible decomposition of the dual representation of the special unitary group and the permutation group. The method of irreducible decomposition provides the universal protocols in quantum setting[7, 9, 13, 14, 15, 16, 17]. However, even in the classical case, the universal channel coding requires the conditional type as well as the type[2]. In the present paper, we introduce a quantum analogue of the conditional type, which is the most essential part of the present paper. The third method is the packing lemma, which yields a suitable combination of the signal states.
independent of the form of the channel in the classical case[2]. This method plays the same role in the present paper.

The remainder of the present paper is organized as follows. In section 2, the notation for group representation theory is presented and a quantum analogue of conditional type is introduced. Group representation theory is essential for our derivation as the previous papers[7, 9, 13, 14, 15, 16, 17]. Using these facts, we show that it is possible to approximate the tensor product state universally. In section 3, we give the first main result, which treats the universal quantum channel coding. In section 4, we give a code that well works universally. In section 5, we treat universal approximation of tensor product states more precisely, and discuss the mini-max problem of this redundancy, which is the second main result of this paper. In section 6, we discuss the relation between the second main result and the quantum data compression.

2. Group representation theory

In this section, we focus on the dual representation on the n-fold tensor product space by the the special unitary group SU(d) and the n-th symmetric group S_n. For this purpose, we focus on the Young diagram and the ‘type’. The former is a key concept in group representation theory and the latter is that in information theory[2]. When the vector of integers \( \vec{n} = (n_1, n_2, \ldots, n_d) \) satisfies the condition \( n_1 \geq n_2 \geq \ldots \geq n_d \geq 0 \) and \( \sum_{i=1}^{d} n_i = n \), the vector \( \vec{n} \) is called the Young diagram (frame) with size \( n \) and depth \( d \), the set of which is denoted as \( \mathcal{Y}_n^d \). When the vector of integers \( \vec{n} \) satisfies the condition \( n_i \geq 0 \) and \( \sum_{i=1}^{d} n_i = n \), the vector \( \vec{p} = \frac{\vec{n}}{n} \) is called the ‘type’ with size \( n \), the set of which is denoted as \( T_n^d \). Further, for \( \vec{p} \in T_n^d \), the subset of \( \mathcal{X}^n \) is defined as:

\[
T_{\vec{p}} := \{ \vec{x} \in \mathcal{X}^n | \text{The empirical distribution of } \vec{x} \text{ is equal to } \vec{p} \}.
\]

The cardinalities of these sets are evaluated as follows:

\[
| \mathcal{Y}_n^d | \leq | T_n^d | \leq (n + 1)^{d-1} \tag{4}
\]

\[
(n + 1)^{-d} e^{-H(\vec{p})} \leq | T_{\vec{p}} | \tag{5}
\]

where \( H(\vec{p}) := -\sum_{i=1}^{d} p_i \log p_i [2] \). Since the sets \( \{ (n_s(i)) | \vec{n} \in \mathcal{Y}_n^d \} \) and \( \{ (n_{s'}(i)) | \vec{n} \in \mathcal{Y}_n^d \} \) are distinct for any \( s \neq s' \in S_d \), the relation \( | \{ \vec{n} | \sum_i n_i = n \} | \cong \frac{n^{d-1}}{(d-1)!} \) implies the following asymptotic behavior of the cardinality \( | \mathcal{Y}_n^d | \):

\[
| \mathcal{Y}_n^d | \cong \frac{n^{d-1}}{d!(d-1)!} \tag{6}
\]

Using the Young diagram, we can characterize the irreducible decomposition of the natural action of SU(d) and \( S_n \) on the tensor space \( \mathcal{H}^{\otimes n} \):

\[
\mathcal{H}^{\otimes n} = \bigoplus_{\vec{n} \in \mathcal{Y}_n^d} \mathcal{U}_{\vec{n}} \otimes \mathcal{V}_{\vec{n}},
\]

where \( \mathcal{U}_{\vec{n}} \) is the irreducible representation space of SU(d) characterized by \( \vec{n} \), and \( \mathcal{V}_{\vec{n}} \) is the irreducible representation space of \( n \)-th symmetric group \( S_n \) characterized by \( \vec{n} \). Here, the

\[1 \text{ Christandl[18] contains a good survey of representation theory for quantum information.} \]
representation of the $n$-th symmetric group $S_n$ is denoted as $V : s \in S_n \mapsto V_s$. According to Weyl’s dimension formula, the dimension of $U_{\vec{u}}$ can be expressed as

$$\dim U_{\vec{u}} = \prod_{i<j} \frac{n_i - n_j + j - i}{j - i} < n^{\frac{d(d-1)}{2}}. \quad (7)$$

for any $\vec{u} \in Y^d_n$. That is, for a given probability distribution $\vec{p} = (p_1, \ldots, p_d)$ on $\{1, \ldots, d\}$ satisfying the condition $p_1 > p_2 > \cdots > p_d$, when $(p_1, \ldots, p_d) = (\frac{n_1}{n}, \ldots, \frac{n_d}{n})$,

$$\dim U_{\vec{u}} \cong \frac{\prod_{i<j}(p_i - p_j)}{2^d - 1} \frac{d(d-1)}{2} \cdot \cdots \cdot \frac{d(d-1)}{2}. \quad (8)$$

Then, we denote the projection to the subspace $U_{\vec{u}} \otimes V_{\vec{u}}$ by $I_{\vec{u}}$, and define the following.

$$\rho_{\vec{u}} := \frac{1}{\dim U_{\vec{u}} \otimes V_{\vec{u}}} I_{\vec{u}}, \quad \sigma_{U,n} := \sum_{\vec{u} \in Y^d_n} \frac{1}{|Y^d_n|} \rho_{\vec{u}}. \quad (9)$$

For any state $\rho$ and any Young diagram $\vec{u} \in Y^d_n$, the following relation holds.

$$\dim U_{\vec{u}} \rho_{\vec{u}} \geq I_{\vec{u}} \rho^{\otimes n} I_{\vec{u}}.$$ 

Thus, (4), (7), and (9) yield the inequality

$$(n + 1)^{(d+2)(d-1)} \sigma_{U,n} \geq \rho^{\otimes n}. \quad (10)$$

Since $\sigma_{U,n}$ is commutative with $\rho^{\otimes n}$, we have

$$\frac{(d + 2)(d - 1)}{2} \log(n + 1) + \log \sigma_{U,n} \geq \log \rho^{\otimes n}.$$ 

Thus, we obtain

$$D(\rho^{\otimes n} \| \sigma_{U,n}) = \text{Tr} \rho^{\otimes n} (\log \rho^{\otimes n} - \log \sigma_{U,n}) \leq \frac{(d + 2)(d - 1)}{2} \log(n + 1).$$

Therefore, the state $\sigma_{U,n}$ universally approximates the state $\rho^{\otimes n}$ in the sense of the normalized quantum relative entropy:

$$\frac{1}{n} D(\rho^{\otimes n} \| \sigma_{U,n}) \to 0.$$ 

Next, we focus on two systems $X$ and $Y = \{1, \ldots, l\}$. When the distribution of $X$ is given by a probability distribution $\vec{p} = (p_1, \ldots, p_d)$ on $\{1, \ldots, d\}$, and the conditional distribution on $Y$ with the condition on $X$ is given by $\vec{V}$, we denote the joint distribution on $X \times Y$ by $\vec{p} \vec{V}$ and the distribution on $Y$ by $\vec{p} \cdot \vec{V}$. When the empirical distribution of $\vec{x} \in X^n$ is $(\frac{n_1}{n}, \ldots, \frac{n_d}{n})$, the sequence of types $\vec{V} = (\vec{v}_1, \ldots, \vec{v}_d) \in T_n^{\vec{x}_1} \times \cdots \times T_n^{\vec{x}_d}$ is called a conditional type for $\vec{x}$. We denote the set of conditional types for $\vec{x}$ by $V(\vec{x}, Y)$. For any conditional type $V$ for $\vec{x}$, we define the subset of $Y^n$:

$$T_V(\vec{x}) := \left\{ \vec{y} \in Y^n \mid \text{The empirical distribution of } ((x_1, y_1), \ldots, (x_n, y_n)) \text{ is equal to } \vec{p} \vec{V} \right\},$$
where $\tilde{p}$ is the empirical distribution of $x$.

We define the state $\rho_x$ for $x \in X^n$. For this purpose, we consider a special element $x' = (1, \ldots, 1, 2, \ldots, 2, \ldots, k, \ldots, k)$. The state $\rho_{x'}$ is defined as $\rho_{x'} = \sigma_{U,m_1} \otimes \sigma_{U,m_2} \otimes \cdots \otimes \sigma_{U,m_k}$.

For a general element $x' \in X^n$, we choose a permutation $s \in S_n$ such that $x = sx'$. Then, we define the state $\rho_x$ as $\rho_x = U_s \rho_{x'} U_s^\dagger$, where $U_s$ is the unitary representation of $S_n$. This state plays a similar role as the conditional type in the classical case. Using the inequality (10), we have

$$n^{\frac{k(d-1)}{2}} |Y_n^{d,k}| \rho_x \geq W_n(x).$$

For $\bar{n}_1 \in Y_{m_1}^d, \bar{n}_2 \in Y_{m_2}^d, \ldots, \bar{n}_k \in Y_{m_k}^d$, the density $\rho_{\bar{n}_1} \otimes \rho_{\bar{n}_2} \otimes \cdots \otimes \rho_{\bar{n}_k}$ is commutative with the projector $I_{\bar{n}}$ for $\bar{n} \in Y_n^d[19]$. This fact implies that the density $\rho_x$ is commutative with the density $\rho_{U,n}$. This property is essential for the construction of the proposed decoder.

### 3. Universal classical-quantum channel coding

In the classical-quantum channel, we focus on the set of input alphabets $X := \{1, \ldots, k\}$ and the representation space $H$ of the output system, whose dimension is $d$. Then, a classical-quantum channel is given as the map from $X$ to the set of densities on $H$ with the form $i \mapsto W(i)$. The $n$-th discrete memoryless extension is given as the map from $X^n$ to the set of densities on the $n$-th tensor product system $H^\otimes n$. That is, this extension maps the input sequence $i = (i_1, \ldots, i_n)$ to the state $W_n(i^n) := W(i_1) \otimes \cdots \otimes W(i_n)$. Sending the message $\{1, \ldots, M_n\}$ requires an encoder and a decoder. The encoder is given as a map $\varphi_n$ from the set of messages $\{1, \ldots, M_n\}$ to the set of alphabets $X^n$, and the decoder is given by a POVM $Y^n = \{Y^n_i\}_{i=1}^{M_n}$. Thus, the triplet $\Phi_n := (M_n, \varphi_n, Y^n)$ is called a code. Its performance is evaluated by the size $|\Phi_n| := M_n$ and the average error probability given by

$$\varepsilon[\Phi_n, W] := \frac{1}{M_n} \sum_{i=1}^{M_n} \text{Tr} W_n(\varphi_n(i))(I - Y_i^n).$$

As mentioned in the following main theorem, there exists an asymptotically optimal code that depends only on the coding rate.

**Theorem 1** For any distribution $\tilde{p} = \{p_i\}_{i=1}^k$ on the set of input alphabets $X := \{1, \ldots, k\}$ and any real number $R$, there is a sequence of codes $\{\Phi_n\}_{n=1}^\infty$ such that

$$\lim_{n \to \infty} \frac{-1}{n} \log \varepsilon[\Phi_n, W] \geq \max_{0 \leq t \leq 1} \frac{\phi_{W,\tilde{p}}(t) - tR}{1 + t}$$

$$\lim_{n \to \infty} \frac{1}{n} \log |\Phi_n| = R$$

for any classical-quantum channel $W$, where $\phi_{W,\tilde{p}}(t)$ is given by

$$\phi_{W,\tilde{p}}(t) := -(1 - t) \log \text{Tr}(\sum_{i=1}^{k} p_i W(i)^{1-t})^{\frac{1}{1-t}}.$$

Note that the code $\{\Phi_n\}_{n=1}^\infty$ does not depend on the channel $W$, and depends only on the distribution $\tilde{p}$ and the coding rate $R$. 
The derivative of $\phi_{W,p}(t)$ is given as
\[
\phi'_{W,p}(0) = I(p, W) := \sum_{i=1}^{k} p_i \text{Tr} W(i)(\log W_i - \log W_p), \quad W_p := \sum_{i=1}^{k} p_i W(i).
\]
When the transmission rate $R$ is smaller than the mutual information $I(\vec{p}, W)$,
\[
\max_{0 \leq t \leq 1} \frac{\phi_{W,p}(t) - tR}{1 + t} > 0
\]
because there exists a parameter $t \in (0, 1)$ such that $\phi_{W,p}(t) - tR > 0$. That is, the average error probability $\varepsilon[\Phi_n, W]$ goes to zero.

Bjelakovic and Boche[23] constructed a code universally attaining the mutual information, independently. Indeed, our evaluation for error probability is exponential while they did not give the exponential evaluation. Further, we explicitly give a universal upper bound for average error probability of the constructed code in (12) while they did not provide such an evaluation.

4. Construction of universal code

According to Csiszár and Körner[2], the proposed code is constructed as follows. For a type $\vec{p} \in T_n^d$ and a real positive number $R < H(\vec{p})$, there exist $M_n := e^{nR - \sqrt{n}}$ distinct elements $M_n := \{\vec{x}_1, \ldots, \vec{x}_{M_n}\} \subset T_{\vec{p}}$ such that their empirical distributions are $\vec{p}$ and
\[
|T_{\vec{p}}(\vec{x}) \cap (M_n \setminus \{\vec{x}\})| \leq |T_{\vec{p}}(\vec{x})| e^{-n(H(\vec{p}) - R)}
\]
for $\vec{x} \in M_n \subset T_{\vec{p}}$ and $\vec{V} \in V(\vec{x}, X)$. As is explained in [19], this argument can be shown by substituting the identical map into $\vec{V}$ in Lemma 5.1 in Csiszár and Körner[2]. Our encoder is constructed from the above argument.

Next, for any $\vec{x} \in X^n$ and any real number $C_n$, we define the projection
\[
P(\vec{x}) := \{\rho_\vec{x} - C_n\sigma_{U,n} \geq 0\},
\]
where $\{X \geq 0\}$ presents the projection $\sum_{i} x_i E_i$ for a Hermitian matrix $X$ with the diagonalization $X = \sum_i x_i E_i$. Remember that the density $\rho_\vec{x}$ is commutative with the other density $\sigma_{U,n}$. Using the projection $P(\vec{x})$, we define the decoder:
\[
Y_{\vec{x}} := \sqrt{\sum_{\vec{x} \in M_n} P(\vec{x})^{-1}} P(\vec{x}) \sqrt{\sum_{\vec{x} \in M_n} P(\vec{x})^{-1}}.
\]
Then, the above-constructed code $(e^{nR - \sqrt{n}}, M_n, \{Y_{\vec{x}}\}_{\vec{x} \in M_n})$ is denoted by $\Phi_{U,n}(\vec{p}, R)$.

As is shown in [19], the properties (5) and (11) yields the evaluation
\[
\varepsilon(\Phi_{U,n}(\vec{p}, R), W) 
\leq 2(n + 1)^{d+\frac{k(k-1)}{2}} |Y_n^d|^{-1} e^{-n(\phi_{W,p}(t) - t(R + r(t)))} + 4n \frac{d(d-1)}{2} |Y_n^d| e^{-nr(t)}, \quad (12)
\]
which implies that
\[
\lim_{n \to \infty} -\frac{1}{n} \log \varepsilon(\Phi_{U,n}(\vec{p}, R), W) \geq \max_{t \in (0, 1)} \frac{\phi_{W,p}(t) - tR}{1 + t}
\]
for any channel $W$. Therefore, we obtain Theorem 1.
5. Approximation of multi-copy states

Next, we discuss the information-theoretic approximation more precisely. That is, we calculate the value $D(\rho^{\otimes n} \| \sigma_{U,n})$ more deeply. In the following calculation, we focus on the set $Y^d := \{|\vec{p} = (p_1, p_2, \ldots, p_{d-1}, 1 - p_1 - \ldots - p_{d-1})| p_1 > p_2 > \ldots > p_{d-1} > 1 - p_1 - \ldots - p_{d-1} > 0\}$ of the probability distributions on $\{1, \ldots, d\}$, and the density

$$P(i) = \sum_{i=1}^{d} p_i |i\rangle \langle i|,$$

where $\{|i\rangle\}_{i=1}^{d}$ is the standard orthonormal basis of $\mathcal{H}$. Thus, for any state $\rho$, there exist $\Omega \in SU(d)/U(1)^{d-1}$ and $\vec{p} \in Y^d$ such that $\rho = \rho_{\Omega} \vec{p} := U_\Omega \rho U_\Omega^\dagger$ where $U_\Omega$ is a representative of $\Omega$. In this calculation, it is essential to calculate the average of the random variable $\log |Y^d_n| + \log \dim \mathcal{U}_n + \log \dim \mathcal{V}_n$ under the distribution $Q_{\vec{p}}(\vec{n}) := Tr(\rho(\vec{p})^{\otimes n} I_n)$. In order to treat $\sum_{\vec{n} \in Y^d_n} Q_{\vec{p}}(\vec{n}) \log \dim \mathcal{V}_n$ asymptotically, Matsumoto and Hayashi [17] introduced the quantity $n_1^n := \frac{n_1!}{n_1^n n_2^n \cdots n_d^n}$.

In their Appendix D, they showed that

$$\sum_{\vec{n} \in Y^d_n} Q_{\vec{p}}(\vec{n}) \log \dim \mathcal{V}_n \log \frac{n_1^n}{n_1!} \approx \sum_{s \in S_d} \log (s) \prod_{i} \delta_{i(s)} \log \left( \prod_{i<j} (p_i - p_j) \right) = \log \prod_{i<j} (p_i - p_j) - \sum_{s \in S_d} \log (s) \prod_{i<j} (p_i - p_j) \log \prod_{i} \delta_{i(s)},$$

(13)

where $\delta_i := d - i$ and we have applied the formula $\sum_{s \in S_d} \log (s) \prod_{i} \delta_{i(s)} = \prod_{i<j} (p_i - p_j)$. In their Appendix C, they calculated $\sum_{\vec{n} \in Y^d_n} Q_{\vec{p}}(\vec{n})$ as

$$\sum_{\vec{n} \in Y^d_n} Q_{\vec{p}}(\vec{n}) \log \frac{n_1^n}{n_1!} \approx H(\vec{p})n - \frac{d-1}{2} \log n - \frac{d-1}{2} \log 2\pi e - \frac{1}{2} \sum p_i.$$  

(14)

As is shown in [20], the combination of (6), (8), (13), and (14) yields

$$\sum_{\vec{n} \in Y^d_n} P(n) \log |Y^d_n| + \log \dim \mathcal{U}_n + \log \dim \mathcal{V}_n \approx H(\vec{p})n + \frac{d^2-1}{2} \log n + C_d - \log d!(d-1)! + C(\vec{p}),$$

where

$$C_d := - \frac{d-1}{2} \log 2\pi e - \log 2d-1 \cdot 3d-2 \cdot \cdots \cdot (d-1)!$$

$$C(\vec{p}) := - \sum_{s \in S_d} \log (s) \prod_{i<j} (p_i - p_j) \log \prod_{i} \delta_{i(s)} + 2 \log \prod_{i<j} (p_i - p_j) - \frac{1}{2} \sum p_i.$$

Hence,

$$D(\rho^{\otimes n} \| \sigma_{U,n}) = - Tr \rho^{\otimes n} \log \sigma_{U,n} - n H(\rho) \approx \frac{d^2-1}{2} \log n + C_d - \log d!(d-1)! + C(\vec{p}).$$

2 Appendix D of [17] contains three systematic typos. In order to recover the correct meaning, replace $S_n$, $i > j$, and $\mathbf{n}'$ by $S_d$, $i < j$, and $\mathbf{n}'$, respectively.
From the above discussion, it is possible to reduce the asymptotic approximation error $D(\rho^\otimes n\|\sigma_{U,n})$ to $d^2/2 - 1/2 \log n$ universally. However, the state $\sigma_{U,n}$ is not necessarily optimal for this approximation. The optimal state is given as the solution of the following mini-max problem.

**Theorem 2** We obtain the following mini-max value:

$$\min_{\{\sigma_n\}} \sup_{\rho \in \mathcal{B}} \lim_{n \to \infty} \left( D(\rho^\otimes n\|\sigma_n) - \frac{d^2}{2} - \frac{1}{2} \log n \right) = C_d + \log \int_{Y^d} e^{C(\tilde{\rho})} d\tilde{\rho},$$

(15)

where $d\tilde{\rho} := dp_1 dp_2 \ldots dp_{d-1}$. The above mini-max value is realized when we choose the mixture state $\sigma_{J,n} := \sum_{i \in Y^d} J_n(i) \rho_{ii}$ with the distribution $J_n(i) := \frac{e^{C(i)}}{\sum_{i' \in Y^d} e^{C(\tilde{i')}}}$. This mini-max value is also attained by the mixture state $\tilde{\sigma}_{J,n} := \int_{Y^d} \rho(\tilde{p})^\otimes n J(\tilde{p}) d\tilde{\rho}$, where $\rho(\tilde{p})^\otimes n$ is the mixture with respect to the invariant measure $\mu$ on $SU(d)$, i.e., $\int_{SU(d)} (U \rho(\tilde{p}) U^\dagger)^\otimes n \mu(dU) = \int \rho_{\tilde{p}}^{\otimes n} d\Omega$. In addition, when the state $\sigma_n$ is $\sigma_{J,n}$ or $\tilde{\sigma}_{J,n}$, the limit of (15) exists.

For a proof, see Hayashi [20]. Since the Jeffreys’ prior gives the mini-max solution in the classical case, the distribution $J(\tilde{p}) d\tilde{\rho} d\Omega$ can be regarded as a quantum extension of Jeffreys’ prior.

6. Universal quantum source coding

In this section, we apply the result obtained to quantum variable-length lossless coding. First, we present a formulation of variable-length lossless coding. For this purpose, we focus on the Fock space:

$$\mathcal{H}_\oplus := \bigoplus_{k=0}^\infty (\mathbb{C}^2)^\otimes k.$$

(16)

When we use this space for storing quantum information, we cannot determine the length of the stored state without state demolition.

Hence variable-length lossless coding is formulated as follows. When we compress the quantum state on the $n$-fold tensor product system of $\mathcal{H}(= \mathbb{C}^2)$, i.e., $\mathcal{H}^\otimes n$, the encoder is given by the isometry $U_n$ from $\mathcal{H}^\otimes n$ to $\mathcal{H}_\oplus$, and the decoder is given as the inverse map from the image $U_n$ to $\mathcal{H}^\otimes n$.

Note that our definition is more general than that of Boström and Felbinger [21] because they assume that the compressed state of the basis state belongs to the space $(\mathbb{C}^2)^\otimes n$, i.e., that it is not a superposition of states of different length. Furthermore, when we store a quantum state in terms of a Fock state, we do not have to know the length of the stored state. From another viewpoint, the average energy is an important quantity for physical realization because a higher energy damages the communication channel and storage. Note that it is not necessary to measure the energy. In this case, it is natural to treat the Hamiltonian $H := \sum_{k=0}^\infty k P_k$, where $P_k$ is the projection to the space $(\mathbb{C}^2)^\otimes k$. In this case, when the initial state is given by $\rho$, the average energy is given as $\text{Tr} H U \rho U^\dagger$. Hence, it is appropriate to consider the minimization of the average energy for a given ensemble $\{(\rho_i, p_i)\}$ on $\mathcal{H}^\otimes n$.

Next, we consider the small class of quantum lossless codes. When we concatenate two general quantum lossless codes, the concatenated lossless code is not necessarily determined. In order to avoid this problem, we consider the prefix quantum lossless code. A code $U_n$ on $\mathcal{H}^\otimes n$ is called a prefix when there exists a basis $\{ | e_i \rangle \}_{i=1}^{d^n}$ on $\mathcal{H}^\otimes n$ such that the state $U_n | e_i \rangle$ has the form $| \phi_1(i) \cdots \phi_k(i) \rangle \in (\mathbb{C}^2)^\otimes k$, and the classical code $\phi(i) := (\phi_1(i) \cdots \phi_k(i))$ satisfies the prefix condition. Note that the base $| e_i \rangle$ does not necessarily have the tensor product form. Since the classical concatenated code of two classical prefix codes can be defined, the concatenated
quantum lossless code of two prefix quantum lossless codes can be defined. Hence, it is natural to restrict our quantum lossless codes to prefix lossless codes.

As is explained in [20], there is one-to-one correspondence between quantum prefix lossless code and density matrix. When we use a prefix quantum lossless code corresponding to a state $\sigma$ and the true mixture state is $\rho_p$, the average energy is

$$- \text{Tr} \, \rho_p \log \sigma = H(\rho_p) + D(\rho_p\|\sigma).$$

Therefore, the relative entropy $D(\rho_p\|\sigma)$ can be regarded as the redundancy of the quantum lossless code $\sigma$ with respect to $\rho_p$. As is mentioned in [20], when the true mixture is $\rho^{\otimes n}$ and the prefix quantum lossless code $\sigma_{P,n} := \sum_{\bar{n} \in Y_n} P_n(\bar{n}) \rho_{\bar{n}}$ is applied, the asymptotic redundancy is given by

$$D(\rho^{\otimes n}||\sigma_{P,n}) \leq \frac{d^2-1}{2} \log n + C_d + \log \int_{Y^d} e^{C(\bar{p})} d\bar{p} + \sum_{\bar{n}} Q_{\bar{p}}(\bar{n}) (\log J(\bar{p}) - \log P_n(\bar{n}) - (d-1) \log n).$$

Hence, the mini-max asymptotic redundancy is $\frac{d^2-1}{2} \log n + C_d + \log \int_{Y^d} e^{C(\bar{p})} d\bar{p}$, which is attained when we choose the mini-max code $\sigma_{J,n}$ or $\bar{\sigma}_{J,n}$. Therefore, the prefix quantum lossless code $\sigma_{J,n}$ and $\bar{\sigma}_{J,n}$ can be used for universal quantum lossless data compression.

7. Discussion

As the first main result, we have constructed a universal code attaining the quantum mutual information based on the combination of information spectrum method, group representation theory, and the packing lemma. The presented code well works because any tensor product state $\rho^{\otimes n}$ is close to the state $\rho_{\bar{p},n}$. Further, Hayashi [22] derived an exponential decreasing rate of error probability in classical-quantum channel, which is close to the capacity. Hence, if a more sophisticated evaluation is applied, a better exponential decreasing rate can be expected. Such an evaluation is left as a future problem.

As the second main result, we have found a sequence of states $\{\sigma_n\}$ such that the relative entropy $D(\rho^{\otimes n}||\sigma_n)$ behaves universally as $\frac{d^2-1}{2} \log n$. While this result was known for the case of a qubit, the general case has remained open for more than ten years. In this derivation, the calculation of Matsumoto and Hayashi [17] plays an essential role. Furthermore, we have solved the asymptotic mini-max problem concerning $D(\rho^{\otimes n}||\sigma_n) - \frac{d^2-1}{2} \log n$. It has been checked that
our optimal value is better than the result of Krattenthaler and Slater [10] for the qubit case. Our discussion is different from the original discussion of Clarke and Barron[4, 6] because they considered optimization with respect to mixtures of \( n \) i.i.d. distributions while we minimize it among all densities. Our method can be translated to the classical case when the family is taken as the full multinomial distribution family. In such a case, the derivation of the solution to the mini-max problem can be expected to be shorter than the original derivation[4, 6].

As is mentioned in Section 6, there is a one-to-one correspondence between the prefix quantum lossless code and a density matrix. Using this relation, we have applied our result (2) and (15) to quantum variable-length lossless data compression. It has been established that it is possible to compress multi-copy states universally by means of the lossless code \( \sigma_{J,n} \) or \( \tilde{\sigma}_{J,n} \).

Acknowledgments
This research was partially supported by a Grant-in-Aid for Scientific Research on Priority Area ‘Deepening and Expansion of Statistical Mechanical Informatics (DEX-SMI)’, No. 18079014 and a MEXT Grant-in-Aid for Young Scientists (A) No. 20686026.

References
[1] Shannon C E 1948 A Mathematical Theory of Communication Bell System Tech. J. 27 379 (Part I), 623 (Part II)
[2] Csiszár I and Körner J 1981 Information Theory: Coding Theorems for Discrete Memoryless Systems (Academic Press)
[3] Gallager R G 1968 Information Theory and Reliable Communication (John Wiley & Sons)
[4] Clarke B S and Barron A R 1990 Information-theoretic asymptotics of Bayes methods IEEE Trans. Inform. Theory 36 453
[5] Amari S and Nagaoka H 2000 Methods of Information Geometry (AMS & Oxford University Press)
[6] Clarke B S and Barron A R 1994 Jeffreys’ prior is asymptotically least favorable under entropy risk J. Statist. Planning and Inference 41 37
[7] Jozsa R, Horodecki M, Horodecki P and Horodecki R 1998 Universal Quantum Information Compression Phys. Rev. Lett. 81 1714
[8] Hayashi M 2002 Exponents of quantum fixed-length pure state source coding Phys. Rev. A 66 032321
[9] Hayashi M and Matsumoto K 2002 Quantum universal variable-length source coding Phys. Rev. A 66 022311
[10] Krattenthaler C and Slater P 2000 Asymptotic Redundancies for Universal Quantum Coding IEEE Trans. Inform. Theory 46 801 (Preprint quant-ph/9612043)
[11] Verdú S and Han T S 1994 A general formula for channel capacity IEEE Trans. Inform. Theory 40 1147
[12] Hayashi M and Nagaoka H 2003 General formulas for capacity of classical-quantum channels IEEE Trans. Inform. Theory 49 1753
[13] Hayashi M 2001 Asymptotics of quantum relative entropy from a representation theoretical viewpoint J. Phys. A: Math. and Gen. 34 3413
[14] Keyl M and Werner R F 2001 Estimating the spectrum of a density operator Phys. Rev. A 64 052311
[15] Hayashi M 2002 Optimal sequence of POVMs in the sense of Stein’s lemma in quantum hypothesis J. Phys. A 35 10759
[16] Bjelaković I, Deuschel J D, Kruger T, Seller R, Siegmund-Schultze R and Szkola A 2005 A Quantum Version of Sanov’s Theorem Comm. Math. Phys. 260 659
[17] Matsumoto K and Hayashi M 2007 Universal distortion-free entanglement concentration Phys. Rev. A 76 062338
[18] Christandl M 2006 The Structure of Bipartite Quantum States - Insights from Group Theory and Cryptography PhD thesis, February 2006, University of Cambridge (Preprint quant-ph/0604183)
[19] Hayashi M 2008 Universal coding for classical-quantum channel Preprint arXiv:0805.4092
[20] Hayashi M 2008 Universal approximation of multi-copy states and universal quantum lossless data compression Preprint arXiv:0806.1091
[21] Boström K and Felbinger T 2002 Lossless quantum data compression and variable-length coding Phys. Rev. A 65 032313
[22] Hayashi M 2007 Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding Phys. Rev. A 76 062301
[23] Bjelakovic I and Boche H 2007 Classical Capacities of Averaged and Compound Quantum Channels, Preprint arXiv:0710.3027