Masahito Hayashi

A Group Theoretic Approach to Quantum Information

Springer

A Group Theoretic Approach to Quantum Information

Masahito Hayashi

A Group Theoretic Approach to Quantum Information

Springer

Masahito Hayashi
Graduate School of Mathematics
Nagoya University
Nagoya
Japan

Copyright for the English version only!

ISBN 978-3-319-45239-5 ISBN 978-3-319-45241-8 (eBook)
DOI 10.1007/978-3-319-45241-8

Library of Congress Control Number: 2016950361

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is the English edition of the Japanese book *A Group Theoretic Approach to Quantum Information*, which was originally published by Kyoritsu Shuppan, Tokyo, Japan in January 2014. Hence, it is the English translation from the original Japanese book.

Group symmetry is a fundamental concept of quantum theory and has been applied to various fields in physics, particle physics, quantum field theory, cosmology, nuclear physics, condensed matter physics, and quantum optics. Since quantum information is an area to utilize the properties of quantum theory, group symmetry plays an important role in quantum information as well. In contrast, most of studies of quantum information have concentrated on application of methods in information science and proposals of new protocols. Hence, group symmetry has not taken a central role of quantum information. Mathematical foundation of quantum information mainly has been based on the viewpoint of operator algebra. While several individual subareas of quantum information employed group theoretic methods, many people consider that group symmetry does not play an important role in quantum information.

However, there exist so many symmetries based on group representation *behind* several quantum information processes although the symmetries have not been noticed sufficiently. Due to the unawareness, no text book has systematically summarized the method of group symmetry in quantum information. That is, due to the lack of the awareness of the group symmetric structure, these topics have been treated in a completely different way; besides they have common structures based on the group symmetric structure. To resolve such problems, this book deals with fundamental topics of quantum information that are essentially based on group representation so that the reader can understand how group symmetry works as a basic infrastructure of quantum information. These topics can be more deeply understood via group theoretical viewpoint, which brings an easier generalization of respective topics.

Chapter 1 introduces fundamental concepts of quantum information, measurement, state, composite system, computation basis, entanglement, etc, and prepares mathematical notations for quantum system. Then, Chap. 2 summarizes fundamental knowledge for quantum channel and information quantities, which are basis of quantum information. These contents can be understood only with linear algebra, calculus, and elementary probability. These two chapters are preliminary and do not discuss group representation. After these preparations, we deal with respective topics of quantum information by assuming the knowledge of group representation theory. However, the reader does not need to worry about the lack of the knowledge of group representation theory because all the required knowledge for group representation theory is summarized from the basics of group representation in the book *Group Representations for Quantum Theory* [44]. If the reader reads the book, the reader can understand the contents of this book only with linear algebra, calculus, and elementary probability without any additional knowledge for group representation theory. For simplicity, this book refers only the book [44] for any topics of group representation while they are obtained in different literature first. To adopt the notations in the above book, we describe the (projective) unitary representation by the font f. Also, the generators of these representations are written with the same font with the capital letter, e.g., Q.

In fact, if we do not employ any knowledge of representation theory, the contents of the remaining chapters cannot be discussed. This property shows that group representation-approach plays an important role in quantum information. This book has a unique characteristic to explicate quantum information based on group representation. On the other hand, quantum information can be discussed more rigorously than other topics in physics so that the contents of this book are more interesting from the mathematical viewpoint. However, quantum information cannot attract sufficient attention from mathematicians. The main reason seems in the point that the existing studies mainly focus not on the common mathematical structure but on the individual topics. To resolve this issue, this book aims to clarify the common mathematical structure via group representation theory.

Now, we proceed to the details of remaining chapters that address individual topics in quantum information. Chapter 3 addresses entanglement of quantum system via group representation. While existing books mainly deal with entanglement of the bipartite system, this chapter deals with entanglement of the multi-partite system as well as entanglement of the bipartite system. This topic seems to have no deep relation with group representation. However, this topic is essentially based on group representation in an unexpected way behind. For example, many entanglement measures are proposed. However, it is quite difficult to calculate them. A larger part of resolved cases are essentially based on the group symmetry because many entanglement measures are based on optimization and employing the group symmetry can reduce the freedom of the optimization.

Chapter 4 discusses covariant measurements with respect to group representation and gives the theory for optimal measurement. This approach can be applied to so many topics in quantum information. Although quantum state estimation is a very fundamental task in quantum information, the problem to optimize the estimation procedure cannot be exactly solved with the finite resource setting, in general, and it can be exactly solved only when the problem has group symmetric property. Hence, we can say that group symmetric property is essential for state estimation. This idea can be applied to the estimation of unknown unitary action. This problem is much more deeply related to group representation because this problem can be investigated by using Fourier transform in the sense of group representation theory. Further, we deal with approximate quantum state cloning in the same way because we can say the same thing for approximate quantum state cloning, which is also an important topic rooted in the foundation of in quantum theory. That is, all of solved examples of approximate quantum state cloning are based on the group symmetry.

Chapter 5 deals with quantum error correction. As quantum error correction, a stabilizer code and a CSS (Calderbank–Shor–Smolin) code are well known. Since these are essentially based on the discrete Heisenberg representation, group representation plays a crucial role in quantum error correction. To address this topic, we first summarize classical algebraic error correction. Then, based on the discrete Heisenberg representation, we address a stabilizer code, which is a typical quantum error correction, and a CSS code, which is a special case of a stabilizer code. Further, we consider a Clifford code, which is a general framework of algebraic quantum error correction. In addition, we investigate the security analysis of quantum cryptography based on our analysis on CSS codes.

Chapter 6 addresses universal information processing based on Schur duality, which is the joint representation of the special unitary group and the permutation group and can be regarded as quantum analogue of the method of types by Csiszár and Körner, which is one of typical approaches to classical information theory. The universality is the independence of the protocol from the channel or the information source. This chapter contains estimation of density matrix, hypothesis testing of quantum state, entanglement concentration, quantum data compression, and classical-quantum channel. When we do not care about the universality, we can discuss these topics without use of representation theory. However, to construct protocols to achieve the universality, we need to employ Schur duality theory because they cannot be constructed without use of Schur duality.

In the above way, the topics of this book cannot be discussed without group representation. Using the group representation this book deals with these fundamental topics in quantum information very efficiently. Unfortunately, the method of group representation is not the typical approach to quantum information. However, many leading researchers consider that the demand of group symmetry is increasing in the area of quantum information. This is because the area of quantum information is well matured and needs a more systematic approach. Further, to lead the reader to understand, the book contains 32 figures and 71 exercises with solutions. Finally, we make a remark for the organization of this book. We put the symbol * on the title of a section or an example that is very complicated. A part with * will be used only in parts with * so that the reader can understand major parts without reading parts with *. So, the reader is recommended to omit such parts if he/she is not familiar to group representation.

The author will be grateful when the readers become interested in quantum information via this book. Finally, the author expresses the acknowledgments to all persons who cooperated to produce this book. Especially, the author would like to thank Prof. Rytaroh Matsumoto of Tokyo Institute of Technology, Prof. Shun

Watanabe of Tokyo University of Agriculture and Technology, and Mr. Yuxiang Yang of The University of Hong Kong for providing helpful comments. The author would also like to thank Mr. Hideo Kotobuki and Ms. Kei Akagi of Kyoritsu Shuppan for supporting the original Japanese version. The author would also like to thank Dr. Claus E. Ascheron of Springer Science+Business Media for his encouragement and patience during the preparation of the manuscript.

Nagoya, Japan Masahito Hayashi
July 2016

Contents

About the Author

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

He worked in Kyoto University as a Research Fellow of the Japan Society of the Promotion of Science (JSPS) from 1998 to 2000, and worked in the Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN from 2000 to 2003, and worked 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. He worked in the Graduate School of Information Sciences, Tohoku University as Associate Professor from 2007 to 2012. In 2012, he joined the Graduate School of Mathematics, Nagoya University as Professor. He also worked in Centre for Quantum Technologies, National University of Singapore as Visiting Research Associate Professor from 2009 to 2012 and as Visiting Research Professor from 2012 to now. In 2011, he received the Information Theory Society Paper Award (2011) for Information-Spectrum Approach to Second-Order Coding Rate in Channel Coding. In 2016, he received the Japan Academy Medal from the Japan Academy and the JSPS Prize from Japan Society for the Promotion of Science.

He is a member of the Editorial Board of the International Journal of Quantum Information and International Journal On Advances in Security. His research interests include classical and quantum information theory, information-theoretic security, and classical and quantum statistical inference.

Chapter 1
Mathematical Foundation of Quantum System

Abstract This book covers several topics in quantum information via group representation. For this purpose, this chapter introduces basic concepts of quantum theory, measurement, state, composite system, many-body system, and entanglement. Since a larger part of this chapter overlaps Chap. 1 of my book, *Group Representations for Quantum Theory* [44], the reader who has already read it can skip this part. However, since Sects. 1.1 and 1.2 contain parts that do not appear in the above book, the reader need to read such a part. Important notations used in this book are introduced in this chapter. Chapter 4 of the other book *Introduction to Quantum Information Science* [43] can be recommended as the detail of this contents.

1.1 System, State, and Measurement

In quantum theory that describes microscopic physics, the target is called a **quantum system** or a **system**, and mathematically denoted by a complex Hilbert space $\mathcal{H}$ with a Hermitian inner product. A complex vector space $\mathcal{H}$ is called a **Hilbert space** when it equips a Hermitian inner product. Even when its dimension is finite.[1] Since $\mathcal{H}$ has a Hermitian inner product, there exists a **completely orthonormal system** (**CONS**) $\{e_i\}$. Each normalized base e_i represents a state in the quantum system that is distinguished from each other. An arbitrary state of the system is given as a normalized vector $x \in \mathcal{H}$, which is called a **state vector**. Once we fix a CONS as a standard basis, any vector x describing a state is written as a superposition (a liner combination) $\sum_i x_i e_i$. Quantum theory has two types of notations for an element x of $\mathcal{H}$. One is a ket vector $|x\rangle$, and the other is a bra vector $\langle x|$. These descriptions are defined so that they have linearity with respect to real coefficients and they satisfy

$$|ax\rangle = a|x\rangle, \quad \langle ax| = \bar{a}\langle x|. \tag{1.1}$$

for a complex number $a \in \mathbb{C}$. In particular, for a standard basis e_i, $|e_i\rangle$ and $\langle e_i|$ are simplified to $|i\rangle$ and $\langle i|$, respectively.

[1] When its dimension is infinite, we assume that the space satisfies the **completeness** under the given Hermitian inner product.

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8_1

On the other hand, when the Hermitian inner product $\langle x|y\rangle$ for $x, y \in \mathcal{H}$ is defined so that $\langle ax|by\rangle = \bar{a}b\langle x|y\rangle$ for $a, b \in \mathbb{C}$, the inner product $\langle x|y\rangle$ of x and y can be regarded as the product $\langle x| \cdot |y\rangle$ of the bra vector and the ket vector. In this case, the product $|y\rangle\langle x|$ with the opposite way can be regarded as a linear map from $\mathcal{H}$ to $\mathcal{H}$. When $x \in \mathcal{H}$ is a normalized vector, $|x\rangle\langle x|$ is a one-dimensional projection. We often denote the state corresponding to the normalized vector $x \in \mathcal{H}$ by the one-dimensional projection $|x\rangle\langle x|$. More generally, a state of quantum system $\mathcal{H}$ is described by a Hermitian matrix ρ with non-negative eigenvalues and trace 1. Such a Hermitian matrix ρ is called a **density matrix** (**density operator**). Also, a Hermitian matrix is called **positive semi definite** when all of its eigenvalues are non-negative. In particular, when a density matrix ρ is given a one-dimensional projection, it is called a **pure state**. A density matrix ρ that is not a pure state is called a **mixed state**. Indeed, a normalized vector $|x\rangle$ expresses a pure state $|x\rangle\langle x|$ in the above sense. When we use a normalized vector $|x\rangle$ to express the pure state, the normalized vector $|x\rangle$ is called a **vector state**. Further, the density matrix $\rho_{\text{mix}} := \sum_{i=1}^{d} \frac{1}{d}|i\rangle\langle i|$ on the d-dimensional system $\mathcal{H}$ is called the **completely mixed state**. In particular, when we need to clarify the quantum system $\mathcal{H}$ of our interest, we denote the completely mixed state by $\rho_{\text{mix},\mathcal{H}}$.

A measurement is given as a decomposition $\{M_\omega\}_{\omega\in\Omega}$ of the unit matrix I by positive semi-definite matrices on $\mathcal{H}$ (i.e., $\sum_{\omega\in\Omega} M_\omega = I$) [72]. Here, Ω is the set of possible outcomes,and is called the probability space. When the state of the system is given as the density matrix ρ as Fig. 1.1, the probability to obtain the outcome $\omega \in \Omega$ is $\operatorname{Tr} \rho M_\omega$. Since ρ and M_ω are positive semi definite, the value $\operatorname{Tr} \rho M_\omega$ is always non-negative. Further, the above conditions guarantees that $\sum_{\omega\in\Omega} \operatorname{Tr} \rho M_\omega = \operatorname{Tr} \rho \sum_{\omega\in\Omega} M_\omega = \operatorname{Tr} \rho I = 1$, we find that $\operatorname{Tr} \rho M_\omega$ satisfies the axioms of the probability. This probability distribution is written as P_M^ρ. Such a decomposition $\{M_\omega\}_{\omega\in\Omega}$ of the unit matrix I by positive semi definite matrices is called a **Positive operator-valued measure** (**POVM**). Especially, when all of M_ω are projections, it is called a **Projection valued measure** (**PVM**).

Next, let us discuss the case when the set of outcomes is a general topological space Ω. In this case, we cannot assign the matrix M_ω corresponding to the outcome ω in the same way. When a measure $\nu(d\omega)$ on Ω, the integral

$$\int_\Omega M_\omega \nu(d\omega) = I \tag{1.2}$$

can be regarded as the decomposition of the unit matrix I. So, $\{M_\omega\}$ form a POVM. However, a POVM on a general topological space Ω does not necessarily have the above form. Hence, we need to treat a function of a subset of Ω for describing an arbitrary POVM. Since Ω is a topological space, we consider the Borel sets $\mathcal{B}(\Omega)$

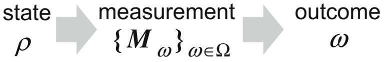


Fig. 1.1 Measuring process

that are generated by open sets of Ω. Then, a map M from $\mathcal{B}(\Omega)$ to $\mathcal{B}_+(\mathcal{H})$ is called a POVM when it satisfies the following conditions. Here, we denote the set of positive semi definite matrices by $\mathcal{B}_+(\mathcal{H})$. When $\mathcal{H}$ is infinite-dimensional, $\mathcal{B}_+(\mathcal{H})$ is the set of positive semi definite bounded operators.

(M1) When the sets $B_j \in \mathcal{B}(\Omega)$ satisfy $B_j \cap B_l = \emptyset$ for distinct elements j, l, we have $\sum_j M(B_j) = M(\cup_j B_j)$.
(M2) $M(\emptyset) = 0$.
(M3) $M(\Omega) = I$.

In particular, M is a PVM when $M(B)$ is a projection for any $B \in \mathcal{B}(\Omega)$. Given a POVM M, we can extend the Hilbert space so that M can be written as a restriction of a PVM E on the enlarged system [99].

Here, we introduce notations for probability. We denote the probability that the occurring event belongs to the set S under the distribution Q by $Q(S)$. When S is given as $\{\omega|$ condition for $\omega\}$, this notation is simplified to $Q\{\omega|$ condition for $\omega\}$. When we focus on the random variable $X(\omega)$, we denote its expectation under the distribution by $\mathrm{E}_Q[X(\omega)]$.

For a Hermitian matrix A, we denote the eigenvalues by $\{a_i\}$, and the projection to the eigenspace with the eigenvalue a_i by E_i. Then, we have a PVM $E = \{E_i\}_i$. Further, we have the relation $A = \sum_i a_i E_i$, which is called the spectral decomposition of the Hermitian matrix A. When the state is given as a density operator ρ, the expectation and the variance are given by $\mathrm{Tr}\, A\rho$ and $\Delta_\rho^2 A := \mathrm{Tr}(A - (\mathrm{Tr}\, A\rho)I)^2\rho$, respectively. That is, we have $\mathrm{E}_{\mathrm{P}_E^\rho}[a_i] = \mathrm{Tr}\, A\rho$ and $\mathrm{E}_{\mathrm{P}_E^\rho}[(a_i - \mathrm{Tr}\, A\rho)^2] = \Delta_\rho^2 A$.

When $\mathcal{H}$ is infinite-dimensional, using a PVM, we can define the spectral decomposition of an operator on $\mathcal{H}$, which can be regarded as the infinite-dimensional extension of diagonzalization. In this case, a self-adjoint operator A on $\mathcal{H}$ does not necessarily have an eigenvector. However, for a self-adjoint operator A on $\mathcal{H}$, there uniquely exists a PVM E with the probability space $\mathbb{R}$ such that

$$A = \int_{\mathbb{R}} x E(dx). \tag{1.3}$$

Here, this fact holds even when the self-adjoint operator A is not bounded [109]. When A is a unitary, the same fact holds by replacing $\mathbb{R}$ by the unit circle $U = \{z \in \mathbb{C} | |z| = 1\}$. In this way, we describe a measurement of a physical quantity given as a self-adjoint operator A on a infinite dimensional space $\mathcal{H}$. When the state is given as a density operator ρ, the expectation and the variance are similarly given by $\mathrm{Tr}\, A\rho$ and $\Delta_\rho^2 A$, respectively.

In fact, a POVM not only gives a probability distribution for the measurement outcome, but also gives a convex decomposition of a density matrix ρ. That is, since $\sum_\omega M_\omega = I$, we have

$$(\mathrm{Tr}\, M_\omega \rho) \sum_\omega \frac{\sqrt{\rho} M_\omega \sqrt{\rho}}{\mathrm{Tr}\, M_\omega \rho} = \sum_\omega \sqrt{\rho} M_\omega \sqrt{\rho} = \sqrt{\rho} I \sqrt{\rho} = \rho. \tag{1.4}$$

For a convex decomposition $\rho = \sum_\omega p_\omega \rho_\omega$ of a density matrix ρ, we can define the POVM $M_\omega := \rho^{-\frac{1}{2}} p_\omega \rho_\omega \rho^{-\frac{1}{2}}$. Then, we can reconstruct the convex decomposition $\rho = \sum_\omega p_\omega \rho_\omega$ by using (1.4).

Exercise 1.1 Consider the case when $M_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{8} \end{pmatrix}$, $M_2 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{8} \end{pmatrix}$, $M_3 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{3}{4} \end{pmatrix}$, and $\rho = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Calculate the distribution P_M^ρ.

Exercise 1.2 Give the spectral decomposition of $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Exercise 1.3 Calculate the average when the measurement is the spectral decomposition of A given in Exercise 1.2 and the state is $\rho = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Exercise 1.4 Calculate the variance in Exercise 1.3.

Exercise 1.5 Focus on the probabilistic decomposition of density matrix $\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Give a POVM based on the decomposition.

1.2 Composite System

Consider two quantum systems $\mathcal{H}_A$ and $\mathcal{H}_B$ that are distinguishable from each other. For example, when the quantum system $\mathcal{H}_A$ expresses the internal system of a particle and the quantum system $\mathcal{H}_B$ expresses the position of the particle, we need the quantum system that describes the freedom of the whole particle. Such a quantum system is called the composite system of the systems $\mathcal{H}_A$ and $\mathcal{H}_B$. The composite system is given by the tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$. The tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$ is defined as the space spanned by the CONS $\{|v_i, u_j\rangle\}_{1 \le i \le k, 1 \le j \le l}$ when $\{|v_i\rangle\}_{i=1}^k$ and $\{|u_j\rangle\}_{j=1}^l$ are CONS of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.

Next, we consider the case when the states of the systems $\mathcal{H}_A$ and $\mathcal{H}_B$ are prepared to be the density matrices ρ and σ, respectively. In this case, when $\rho = \sum_{i,i'=1}^k a_{i,i'} |v_i\rangle\langle v_{i'}|$ and $\sigma = \sum_{j,j'=1}^l b_{j,j'} |u_j\rangle\langle u_{j'}|$, the state of the composite system is the **tensor product state** $\rho \otimes \sigma := \sum_{i,i'=1}^k \sum_{j,j'=1}^l a_{i,i'} b_{j,j'} |v_i, u_j\rangle\langle v_{i'}, u_{j'}|$. A state of the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ cannot be restricted to a tensor product state $\rho \otimes \sigma$ or its convex combination $\sum_j p_j \rho_j \otimes \sigma_j$, which is called a **separable state**. (Here, p_j expresses the probability distribution and ρ_j and σ_j express density matrices on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.) For example, when a state is given by the pure state corresponding to the vector $\sum_j \sqrt{p_j} |v_j, u_j\rangle$, it cannot be written as a convex combination of tensor product state, then, is not separable. Such a state is called

entangled, and has been studied very actively. Especially, such a property of a state is called **entanglement**.

When a state of the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ is given by

$$\rho = \sum_{i,i'=1}^{k} \sum_{j,j'=1}^{l} c_{i,j,i',j'} |v_i, u_j\rangle\langle v_{i'}, u_{j'}|,$$

there exists a density matrix $\mathrm{Tr}_{\mathcal{H}_B} \rho$ on the system $\mathcal{H}_A$ such that the relation

$$\mathrm{Tr}(\mathrm{Tr}_{\mathcal{H}_B} \rho) X = \mathrm{Tr}\, \rho (X \otimes I_{\mathcal{H}_B})$$

holds for a matrix X on the system $\mathcal{H}_A$. In this case, when we focus on only the system $\mathcal{H}_A$, it is suitable to consider that the state of the system $\mathcal{H}_A$ is the density matrix $\mathrm{Tr}_{\mathcal{H}_B} \rho$. The state $\mathrm{Tr}_{\mathcal{H}_B} \rho$ is called the **reduced density matrix** of ρ, and is calculated by

$$\mathrm{Tr}_{\mathcal{H}_B} \rho = \sum_{i,i'=1}^{k} \sum_{j=1}^{l} c_{i,j,i',j} |v_i\rangle\langle v_{i'}|.$$

The operation of taking the reduced density matrix is called the **partial trace**. When the density matrix ρ is diagonal with respect to the basis $\{|v_i, u_j\rangle\}_{i,j}$, the reduced density matrix is the same as the marginal distribution of the probability distribution composed of the diagonal elements. When there is a possibility of confusion, we simplify $\mathrm{Tr}_{\mathcal{H}_B} \rho$ to $\mathrm{Tr}_B \rho$.

Further, if there is a possibility of confusion, for a matrix X on $\mathcal{H}_A$ and a matrix Y on $\mathcal{H}_B$, we simplify the matrices $X \otimes I_{\mathcal{H}_B}$ and $I_{\mathcal{H}_A} \otimes Y$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ to X and Y, respectively. So, the matrix $X \otimes Y$ is simplified to XY. Now, we explain our abbreviations that are applied to the case when standard bases of the systems $\mathcal{H}_A$ and $\mathcal{H}_B$. In this case, we denote the ket vectors of standard basis of both systems by $|j\rangle_A$ and $|j\rangle_B$ so that the system of the ket vector can be distinguished. Hence, the standard basis of the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ is given by $|k\rangle_A \otimes |j\rangle_B$, which is simplified to $|k, j\rangle_{A,B}$.

In the following discussion, for a matrix $X = (x_{k,j})$, we denote the matrix composed of the **complex conjugate** $\overline{x_{k,j}}$ of each entries with respect to the standard basis by $\overline{X}$, and denote the **transposed matrix** with respect to the standard basis by X^T. Then, $X^\dagger$ expresses the **transposed complex conjugate matrix** of X. The matrices $\overline{X}$ and X^T depend on the choice of the standard basis, however, the matrix $X^\dagger$ depends only on the definition of the Hermitian matrix. Then, we denote the vector $\sum_{k,j} x_{k,j} |k, j\rangle_{A,B}$ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ by $|X\rangle\rangle_{A,B}$. So, we have

$$Y \otimes Z |X\rangle\rangle_{A,B} = |Y X Z^T\rangle\rangle_{A,B}. \qquad (1.5)$$

The inner product of two vectors $|X\rangle\rangle_{A,B}$ and $|Y\rangle\rangle_{A,B}$ is calculated to be ${}_{A,B}\langle\langle X|Y\rangle\rangle_{A,B} = \operatorname{Tr} X^\dagger Y$. Hence, a vector $|X\rangle\rangle_{A,B}$ is normalized (the norm is 1) if and only if $\operatorname{Tr} X^\dagger X = 1$. Thus, we have the following formula with respect to the partial trace:

$$\operatorname{Tr}_B |X\rangle\rangle_{A,B}\ {}_{A,B}\langle\langle X| = XX^\dagger, \quad \operatorname{Tr}_A |X\rangle\rangle_{A,B}\ {}_{A,B}\langle\langle X| = X^T\overline{X}. \tag{1.6}$$

Next, let us diagonalize the matrices $XX^\dagger$ and $X^\dagger X$ by using the isometries U and V in the following way:

$$U^\dagger XVV^\dagger X^\dagger U = U^\dagger XX^\dagger U = D, \quad V^\dagger X^\dagger UU^\dagger XV = V^\dagger X^\dagger XV = D,$$

where D is the diagonal matrix whose diagonal elements are in the decreasing order and are non-negative. The rank of D is equal or less than the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$. In this composite system, the rank of the matrix D is called the **Schmidt rank** of the state vector $|X\rangle\rangle$. So, the Schmidt rank is equal or less than the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$. The diagonal elements of $\sqrt{D}$ is called the **Schmidt coefficient** of the state vector $|X\rangle\rangle$. Since the matrix $U^\dagger XV$ and its transposed complex conjugate matrix $V^\dagger X^\dagger U$ are commutative with each other, the squares of the absolute values of the diagonal elements of $U^\dagger XV$ equal the eigenvalues of $\sqrt{D}$. Then, considering a diagonal matrix D' whose diagonal elements have the absolute value 1, we obtain $X = UD'\sqrt{D}V^\dagger$. Rewriting the isometry UD' to U, we obtain the **Schmidt decomposition** of X as

$$X = U\sqrt{D}V^\dagger. \tag{1.7}$$

Let us apply this fact to an arbitrary state vector $|a\rangle$ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$. There exist a CONS $\{|v_i\rangle\}$, $\{|u_j\rangle\}$ and Schmidt coefficients $d_1, \ldots, d_k$ such that

$$|a\rangle = \sum_{j=1}^{k} d_j |v_j\rangle \otimes |u_j\rangle \tag{1.8}$$

when the dimension k of $\mathcal{H}_A$ is not greater than the dimension l of $\mathcal{H}_B$. In some literatures of quantum information, (1.8) is often called the Schmidt decomposition rather than (1.7). The bases $\{|v_i\rangle\}$ and $\{|u_j\rangle\}$ given in (1.8) are called the Schmidt bases.

Given a mixed state ρ on the system $\mathcal{H}_A$, a vector state $|X\rangle\rangle_{A,B}$ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ is called a **purification** of ρ with the **reference system** $\mathcal{H}_B$ when

$$\rho = \operatorname{Tr}_B |X\rangle\rangle_{A,B}\ {}_{A,B}\langle\langle X|. \tag{1.9}$$

Hence, a state vector $|X\rangle\rangle_{A,B}$ is a purification of ρ if and only if $XX^\dagger = \rho$. This condition is equivalent to the following condition: There exists a partial isometry V such that the support of V is the range of $X^\dagger X$ and X is written as

$$X = \sqrt{\rho} V. \tag{1.10}$$

This fact shows that the purification of ρ is uniquely determined up to the freedom with respect to application of the partial isometry on the reference system.

Especially, when the dimensions of both systems $\mathcal{H}_A$ and $\mathcal{H}_B$ are d and all of the diagonal elements of D satisfying (1.7) are $\frac{1}{d}$, the state vector $|X\rangle\rangle_{A,B}$ is called a **maximally entangled state**. This condition is equivalent to the condition of X being a unitary matrix.

Exercise 1.6 Calculate the Schmidt rank and the Schmidt coefficients of $|X\rangle!\rangle$ when $X = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

Exercise 1.7 Show (1.5).

Exercise 1.8 Show $\mathrm{Tr}_B\, |X\rangle\rangle_{A,B\ A,B}\langle\langle Y| = XY^\dagger$.

Exercise 1.9 Give a purification of $\rho = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}$.

Exercise 1.10 Let ρ be a full-rank density matrix on $\mathcal{H}_A$, and $|X\rangle\rangle$ be its purification with the reference $\mathcal{H}_B$, whose dimension equal that of $\mathcal{H}_A$. Show that for any probabilistic decomposition $\rho = \sum_i p_i \rho_i$, there exists a POVM $M = \{M_i\}_i$ on $\mathcal{H}_B$ such that $p_i \rho_i = \mathrm{Tr}_B\, M_i |X\rangle\rangle\langle\langle X|$.

Exercise 1.11 Let ρ be a density matrix on $\mathcal{H}_A$. Show that, for any probabilistic decomposition $\rho = \sum_i p_i |x_i\rangle\langle x_i|$ with pure states, there exist another system $\mathcal{H}_B$, a purification $|X\rangle\rangle$ of ρ on $\mathcal{H}_B$, and a PVM $E = \{E_i\}_i$ on $\mathcal{H}_B$ such that $p_i |x_i\rangle\langle x_i| = \mathrm{Tr}_B\, E_i |X\rangle\rangle\langle\langle X|$.

Exercise 1.12 Calculate the distribution of the outcome when $\mathcal{H}_A$ and $\mathcal{H}_B$ are d-dimensional systems, the state is $\rho_A \otimes \rho_{\mathrm{mix},B}$, and the measurement is written as $\{\frac{1}{d}|U_i\rangle\rangle\langle\langle U_i|\}_{i=1}^{d^2}$ by using d^2 unitaries U_i.

1.3 Many-Body System

Let us consider the case when n particles are given and these particles are characterized by the quantum systems $\mathcal{H}_i$ $(i = 1, \ldots, n)$. When these particles are distinguishable from each other, the composite system corresponding to n particles is given by $(((\mathcal{H}_1 \otimes \mathcal{H}_2) \cdots) \otimes \mathcal{H}_n)$. Since the tensor product does not depend on the order of the tensor product, the above space is the same as $(\mathcal{H}_1(\cdots(\mathcal{H}_{n-1} \otimes \mathcal{H}_n)))$. Hence, we simplify it to $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$. Especially, when each system $\mathcal{H}_i$ is isometric to $\mathcal{H}$, the tensor product space is simplified to $\mathcal{H}^{\otimes n}$. When the state of each system $\mathcal{H}_i$ is independently prepared to be the density matrix ρ_i, the state of the composite system is given as the density matrix $(((\rho_1 \otimes \rho_2) \cdots) \otimes \rho_n)$. Since the tensor product of the matrices does not depend on the order of the tensor product, this density matrix is denoted by $\rho_1 \otimes \rho_2 \cdots \otimes \rho_n$. Especially, when $\rho_i = \rho$, i.e., the n particles are independently prepared in the same ρ, the density matrix of the total system is denoted by $\rho^{\otimes n}$, and is called the n-fold **tensor product state** of the density matrix ρ. The above notation will be applied to the case when ρ and ρ_i are not necessarily density matrices, i.e., are general matrices.

Given a tensor product system $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$ and matrix A on the tensor product system, when we take the partial trace with respect to the specific system $\mathcal{H}_i$, the reduced density matrix of A is denoted by $\mathrm{Tr}_{\mathcal{H}_i} A$. Conversely, when we take the partial trace with respect to all of other systems except for the specific system $\mathcal{H}_i$, the reduced density matrix of A is denoted by $\mathrm{Tr}_{\check{\mathcal{H}}_i} A$. When A is a density matrix ρ, the reduced density matrix $\mathrm{Tr}_{\check{\mathcal{H}}_i} \rho$ is simplified to $\rho_{\mathcal{H}_i}$ or ρ_i.

On the other hand, when a matrix A_i on the system $\mathcal{H}_i$, the matrix $I^{\otimes i-1} \otimes A_i \otimes I^{\otimes n-i}$ on the tensor product system $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$ is simplified to A_i. Then, when $A_i = A$, we simplify $\sum_{i=1}^n A_i$ to $A^{(n)}$.

In fact, this kind of description can be applied to more general case. That is, when we have n systems, we can define the composite system composed of these n systems. when $n \geq 3$, such a system is called a **many-body system**, a **multipartite system**, or a **multi-party system**. The term "many-body system" is mostly used when each system is given as one particle. Other two terms are used for a more general case. The term "multi-party system" is more used in the viewpoint of information science. The term "multipartite system" is used more often in quantum information. So, this book mainly uses the term "multipartite system". To identify the number n, we use the term "n-partite system" to express the multipartite system. In contrast, when $n = 2$, to distinguish the case with $n \geq 3$, this system is called, a **two-body system**, a **bipartite system**, or a **two-party system**.

1.4 Guide for Related Literatures

Here, we introduce literatures that treat quantum information in the viewpoint of the representation theory. Although there are many books for quantum information [12, 42, 43, 45, 97, 101, 129], books with this viewpoint are limited. The first book with this viewpoint is Holevo's book [72], whose Chaps. 3 and 4 discuss the optimization problems with respect to quantum measurement by using the representation theory on a finite-dimensional system. After Holevo's book, several papers employed the representation theory for analysis of quantum information. As the next book of this viewpoint, we can list Christandl [22], which deals with quantum information based on the representation theory on a finite-dimensional system. Chapters 5 and 6 of [72] deals with quantum information in the Bosonic system. Latest progress of this direction is reviewed in Wang [125].

Chapter 2
Quantum Channel, and Information Quantity, and Their Mathematical Structure

Abstract Any quantum operation can be regarded as a quantum channel by considering that the initial and final states are the input and output state, respectively. That is, they are the mathematically same object. In this book, for a unified treatment, we use only the term **quantum channel**. Especially, the main purpose of quantum theory is the precise prediction (description) of the change of the state due to the time evolution in several environments. In the following, we define a quantum channel as a map from an input state to an output state, and this definition brings the mathematically completed framework to satisfy the above purpose. We also discuss information quantities of the quantum system and its relation with quantum channel.

2.1 Channel in Quantum System

When the input and output systems are $\mathcal{H}$ and $\mathcal{K}$, a quantum channel is given as a map Λ from the set $\mathcal{S}(\mathcal{H})$ of density matrices on $\mathcal{H}$ to the set $\mathcal{S}(\mathcal{K})$ of density matrices on $\mathcal{K}$. In this case, the definition of the convex combination $\lambda\rho_1 + (1-\lambda)\rho_2$ $(0 < \lambda < 1)$ implies the equation $\Lambda(\lambda\rho_1 + (1-\lambda)\rho_2) = \lambda\Lambda(\rho_1) + (1-\lambda)\Lambda(\rho_2)$.

Further, when the map Λ is extended to a map from the set $\mathcal{B}(\mathcal{H})$ of Hermitian matrices on $\mathcal{H}$ to the set $\mathcal{B}(\mathcal{K})$ of Hermitian matrices on $\mathcal{K}$, the relation

Linearity: $\Lambda(aA + bB) = a\Lambda(A) + b\Lambda(B)$

holds [43, 45], where a and b are arbitrary real numbers, and A and B are elements of $\mathcal{B}(\mathcal{H})$. However, a liner map Λ does not necessarily give a quantum channel.

In order that a liner map Λ gives a quantum channel, the liner map Λ needs to map an element of $\mathcal{S}(\mathcal{H})$ to an element of $\mathcal{S}(\mathcal{K})$, at least. This condition is equivalent to the following conditions.

Positivity: $A \ge 0 \Rightarrow \Lambda(A) \ge 0$.
Trace-preserving: $\operatorname{Tr} A = \operatorname{Tr} \Lambda(A)$.

However, even though a map Λ satisfies these three conditions, it cannot give a quantum channel, which is a difficult point of quantum theory.

Here, we introduce an ancilla $\mathbb{C}^n$, and consider the quantum channel on the composite system $\mathcal{H} \otimes \mathbb{C}^n$ with the ancilla. Now, we assume that there is no interaction

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8_2

between the ancilla and the system $\mathcal{H}$ and that the ancilla has no state change. When we define $\Lambda \otimes id_n(A \otimes B) := \Lambda(A) \otimes B$, the quantum channel on the composite system $\mathcal{H} \otimes \mathbb{C}^n$ is given as the map $\Lambda \otimes id_n$. Hence, the map $\Lambda \otimes id_n$ needs to satisfy the positivity as well.

n-**positivity**: $X \geq 0 \Rightarrow \Lambda \otimes id_n(X) \geq 0$, where X is a matrix on $\mathcal{H} \otimes \mathbb{C}^n$.

In particular, the map Λ is called completely positive when it is n-positive for any positive integer n.

Completely positivity (CP): Λ is n-positive for any positive integer n.

In particular, a completely positive map and a trace-preserving completely positive map are abbreviated to a CP map and a TP-CP map, respectively. These discussions guarantee that a quantum channel satisfies the linearity, the completely positivity, and the trace-preserving condition, i.e., a quantum channel is given by a TP-CP map. Conversely, when a map Λ satisfies these conditions, there exists a quantum channel corresponding to the map Λ. Then, we denote the set of quantum channels from the system $\mathcal{H}$ to the system $\mathcal{K}$ by $\mathcal{T}(\mathcal{H}, \mathcal{H}_1)$.

Given a quantum channel from the system $\mathcal{H}$ to the system $\mathcal{K}$ and the reference system $\mathcal{R}$ with the same dimension as the system $\mathcal{H}$, when the initial state of the composite system $\mathcal{H} \otimes \mathcal{R}$ is the maximally entangled state, which is a constant times of $|I\rangle\rangle := \sum_l |l, l\rangle_{H,R}$, we focus on the (unnormalized) final state $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ on the composite system $\mathcal{K} \otimes \mathcal{R}$. This state is called the **Choi-Jamiolkowski representation** of Λ because it contains all information with respect to the channel Λ, as explained later [19, 78]. When the input system is finite-dimensional, the trace of the normalized state is the dimension of the input system.

To describe the output state when the input state on $\mathcal{H}$ is $|u\rangle_H = \sum_l u_l |l\rangle_H$, we prepare the following formula by using $|\overline{u}\rangle_R = \sum_l \overline{u}_l |l\rangle_R$. When the relation

$$|u\rangle_H \, {}_H\langle u| = {}_R\langle \overline{u}|(|I\rangle\rangle\langle\langle I|)|\overline{u}\rangle_R = \mathrm{Tr}_{\mathcal{R}}(|I\rangle\rangle\langle\langle I|)I \otimes |\overline{u}\rangle_R \, {}_R\langle \overline{u}|. \tag{2.1}$$

hold, by exchanging the orders between $\mathrm{Tr}_{\mathcal{R}}$ and $(\Lambda \otimes id)$, the output state of the channel Λ can be described as follows

$$\begin{aligned}\Lambda(|u\rangle_H \, {}_H\langle u|) =& \mathrm{Tr}_{\mathcal{R}}(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)I \otimes |\overline{u}\rangle_R \, {}_R\langle \overline{u}| \\ =& {}_R\langle \overline{u}|(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)|\overline{u}\rangle_R.\end{aligned}$$

Since Λ is a CP map, we have

$$(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|) \geq 0. \tag{2.2}$$

Since Λ is trace-preserving, we obtain

$$\mathrm{Tr}_{\mathcal{K}}(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|) = I. \tag{2.3}$$

Hence, we can describe the TP-CP map Λ by using $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ satisfying the relations (2.2) and (2.3).

Next, let $|U\rangle\rangle_{(E,K),R}$ be the purification of Choi-Jamiolkowski representation $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ of Λ whose reference system is $\mathcal{E}$. Then, U can be regarded as a linear map from $\mathcal{R}$ to $\mathcal{E} \otimes \mathcal{K}$. In particular, the reference system of Choi-Jamiolkowski representation is called the **environment system**. Then, the TP-CP map Λ is given by

$$\Lambda(\rho) = \mathrm{Tr}_E\, U\rho U^{\dagger}. \tag{2.4}$$

When Λ is a TP-CP map, U is an isometry because $\mathrm{Tr}\,\rho = \mathrm{Tr}\, U\rho U^{\dagger}$. In the following, the pair of an environment system $\mathcal{E}$ and an isometry U from the system $\mathcal{H}$ to the system $\mathcal{E} \otimes \mathcal{K}$ is called **Stinespring representation** [119] of Λ when it satisfies (2.4).

Then, the TP-CP map Λ can be written as follows by using a CONS $\{|m\rangle_E\}$ of $\mathcal{E}$ and its Stinespring representation $\mathcal{E}, U$:

$$\Lambda(\rho) = \sum_{m,m'} {}_E\langle m'|U\rho U^{\dagger}|m\rangle_E. \tag{2.5}$$

Defining $F_m := {}_E\langle m|U$, we have

$$\Lambda(\rho) = \sum_m F_m \rho F_m^{\dagger}. \tag{2.6}$$

Due to the trace-preserving condition, $\{F_m\}$ satisfies the condition

$$\sum_m F_m^{\dagger} F_m = I. \tag{2.7}$$

Then, the RHS of (2.7) is called a Kraus representation of Λ [88]. Conversely, when a map Λ is given by (2.6) with use of $\{F_m\}$ satisfying (2.7), Λ is a TP-CP map.

Summarizing these discussions, we obtain the following theorem: (The above discussions give the proofs of **(1)**⇒ **(2)**⇒ **(3)**⇒ **(4)**⇒ **(1)**.)

Theorem 2.1 *The following three conditions are equivalent for a linear map Λ from the set $\mathcal{B}(\mathcal{H})$ of Hermitian matrices on $\mathcal{H}$ to the set $\mathcal{B}(\mathcal{K})$ of Hermitian matrices on $\mathcal{K}$.*

(1) *Λ is a TP-CP map.*
(2) *The Choi-Jamiolkowski representation $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ of Λ satisfies (2.2) and (2.3).*
(3) *There exist an environment system $\mathcal{E}$ and an isometry U from the system $\mathcal{H}$ to $\mathcal{E} \otimes \mathcal{K}$ satisfying (2.4). (Existence of Stinespring representation)*
(4) *Λ is given by (2.6) with use of $\{F_m\}$ satisfying (2.7). (Existence of Kraus representation)*

The above given Stinespring representation has the following physical meaning. Once the state evolution Λ occurs, an interaction with the environment system $\mathcal{E}$ occurs so that the input state ρ on the system $\mathcal{H}$ is transfered to the state on $\mathcal{E} \otimes \mathcal{K}$ via the isometry U. Since the receiver of the channel receives only the output system $\mathcal{K}$, we need to take the partial trace with respect to the environment system $\mathcal{E}$ so that the receiver has the output state $\Lambda(\rho)$.

Hence, the information leaked to the environment system is given by the following TP-CP map Λ_E, which is used for analysis of the information leakage via an eavesdropper.

$$\Lambda_E(\rho) := \mathrm{Tr}_{\mathcal{K}} U \rho U^{\dagger}.$$

Since $U = \sum_m |m\rangle \otimes F_m$, by using Kraus representation $\{F_m\}$, the TP-CP map Λ_E is characterized as

$$\Lambda_E(\rho) = \sum_{m,m'} \mathrm{Tr}\, F_m \rho F_{m'}^{\dagger} |m\rangle\langle m'|. \tag{2.8}$$

Then, the TP-CP map Λ_E does not depend on the choice of the Stinespring representation in the following sense. Hence, without loss of generality, we can assume that Λ_E is given by (2.8). Now, we consider two Stinespring representations, in which, their environment systems are $\mathcal{E}, \mathcal{E}'$ and their output states are $\rho_{R,E,K}$ and $\rho'_{R,E,K}$ when the initial state of the composite system $\mathcal{R}$ and $\mathcal{H}$ is the maximally entangled state. Their output states $\rho_{R,E,K}$ and $\rho'_{R,E,K}$ are the purifications of the Choi-Jamiolkowski representation $\frac{1}{\dim \mathcal{H}}(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ of Λ. Due to the uniqueness of the purification, there exists a partial isometry V from the environment system $\mathcal{E}$ to the environment system $\mathcal{E}'$ such that $V \rho_{R,E,K} V^{\dagger} = \rho'_{R,E,K}$. Letting Λ_E and Λ'_E be TP-CP maps to the environment systems in the respective Stinespring representation, we have

$$V \Lambda_E(\rho) V^{\dagger} = \dim \mathcal{H} V\, \mathrm{Tr}_{K,R}\, \rho_{R,E,K} \overline{\rho} V^{\dagger} = \dim \mathcal{H}\, \mathrm{Tr}_{K,R}\, \rho'_{R,E,K} \overline{\rho} = \Lambda_E(\rho)'.$$

So, the final states in the respective environment systems are mapped to each other via the partial isometry V. When $\Lambda(\rho_{\mathrm{mix},\mathcal{H}}) = \rho_{\mathrm{mix},\mathcal{K}}$, the channel Λ is called **unital**.

When a quantum channel Λ is written as $\Lambda(\rho) = \sum_i (\mathrm{Tr}\, M_i \rho) \rho_i$ by using a POVM $\{M_i\}$ on the input system $\mathcal{H}$ and a set $\{\rho_i\}$ of states on the output system $\mathcal{K}$, the quantum channel Λ is called an **entanglement-breaking channel**. Then, we have the following theorem [45, Theorem 5.2].

Theorem 2.2 *A quantum channel Λ is a entanglement-breaking channel if and only if the Choi-Jamiolkowski representation $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ of the quantum channel Λ is separable.*

2.2 State Reduction Due to Measurement

When we apply the measurement on a system $\mathcal{H}$, the state of the system $\mathcal{H}$ is changed irreversibly. Such a state change is called a state reduction. When the measurement is described by a POVM $\{M_\omega\}_{\omega\in\Omega}$, the state reduction is not uniquely determined from the measurement outcome ω. The typical state reduction is given as follows. When the initial state before the measurement is ρ, the final state with the measurement outcome ω is given by

$$\frac{\sqrt{M_\omega}\rho\sqrt{M_\omega}}{\mathrm{Tr}\,\rho M_\omega}. \tag{2.9}$$

We call the above state reduction the **typical state reduction** due to the POVM $\{M_\omega\}_{\omega\in\Omega}$. Especially, when the POVM is a PVM $\{E_\omega\}_{\omega\in\Omega}$, the state reduction (2.9) is written as

$$\frac{E_\omega\rho E_\omega}{\mathrm{Tr}\,\rho E_\omega}. \tag{2.10}$$

and is called the **projection postulate**. When the measurement is given by a PVM and is performed precisely, it is thought that the state reduction satisfies the projection postulate. A measurement satisfying this condition is called a **non-demolition measurement**. Then, the map $\rho \mapsto \sum_\omega E_\omega\rho E_\omega$ gives the average output state and is called the **pinching**. Especially, when a PVM $\{E_i\}$ is given by the spectral decomposition $\sum_i x_i E_i$ of a Hermitian matrix X, we denote the pinching of the PVM $\{E_i\}$ by Λ_X.

However, the condition (2.9) does necessarily not hold in general. A general state reduction due to a measurement is given as follows by using the set $\{\Lambda_\omega\}_{\omega\in\Omega}$ of CP maps Λ_ω depending on the measurement outcome $\omega \in \Omega$. When the initial state before the measurement is ρ, the final state with the measurement outcome ω is

$$\frac{\Lambda_\omega(\rho)}{\mathrm{Tr}\,\Lambda_\omega(\rho)}, \tag{2.11}$$

and the probability to obtain the measurement outcome ω is $\mathrm{Tr}\,\Lambda_\omega(\rho)$. Since $\sum_{\omega\in\Omega}\mathrm{Tr}\,\Lambda_\omega(\rho) = 1$, the CP-map $\sum_{\omega\in\Omega}\mathrm{Tr}\,\Lambda_\omega$ needs to preserve the trace. Hence, the state reduction due to the measurement is given by (2.11) by using the decomposition $\{\Lambda_\omega\}_{\omega\in\Omega}$ by CP maps of a TP-CP map. Such a decomposition is called an **Instrument** [105–107]. Since a CP map has a Kraus representation, there exist matrices $F_{m,\omega}$ such that $\Lambda_\omega(\rho) = \sum_m F_{m,\omega}\rho F_{m,\omega}^\dagger$ and $\sum_{\omega,m} F_{m,\omega}^\dagger F_{m,\omega} = I$. In fact, an instrument can describe the combination of a state reduction due to the measurement and the application of a quantum channel after the state reduction.

2.3 Convex Set and Inequalities

2.3.1 Convex Set and Convex Function

Given two states ρ_1 and ρ_2 on the system $\mathcal{H}$ and $1 \geq \lambda \geq 0$, the **convex combination** $\lambda\rho_1 + (1-\lambda)\rho_2$ is a state on the system $\mathcal{H}$ as well. A set is called a **convex set** when the set contains any convex combination of any two elements of the set like the set of states on the system $\mathcal{H}$. An element x of a convex set $\mathcal{X}$ is called an **extremal point** of $\mathcal{X}$ when any elements x_1 and x_2 of $\mathcal{X}$ satisfying $x = \lambda x_1 + (1-\lambda)x_2$ are limited to be x. Then, we denote the set of extremal points of $\mathcal{X}$ by $\mathrm{Ext}(\mathcal{X})$. For example, the set $\mathrm{Ext}(\mathcal{S}(\mathcal{H}))$ of extremal points in the set $\mathcal{S}(\mathcal{H})$ of state on the system $\mathcal{H}$ is the set of pure states on the system $\mathcal{H}$. Given a subset $\mathcal{S}$ of $\mathcal{X}$, the subset $\{\sum_j p_j x_j | x_j \in \mathcal{S}\}$ is called the **convex hull** of $\mathcal{S}$. The dimension of the convex set $\mathcal{X}$ is defined to be the minimum dimension m of the real vector space containing $\mathcal{X}$.

A real-valued function f defined on a convex set is called a **convex function** when the inequality $\lambda f(\rho_1) + (1-\lambda) f(\rho_2) \geq f(\lambda\rho_1 + (1-\lambda)\rho_2)$ holds for any real number $\lambda \in (0,1)$. Conversely, f is called a **concave function** when $-f$ is a convex function. For example, the functions $x \mapsto x\log x$, $x \mapsto x^{1+s}$, and $x \mapsto x^{-s}$ defined on $\mathbb{R}_+ := \{x \in \mathbb{R} | x \geq 0\}$ are convex functions, where $s > 0$. The functions $x \mapsto \log x$ and $x \mapsto x^{1-s}$ are concave functions, where $1 > s > 0$.

Given a function f defined on the set $\mathrm{Ext}(\mathcal{X})$ of extremal points of a convex set $\mathcal{X}$, we define the following function $\mathrm{CR}(f)$ on the convex set $\mathcal{X}$ as Fig. 2.1, which is called the **convex roof** of f.

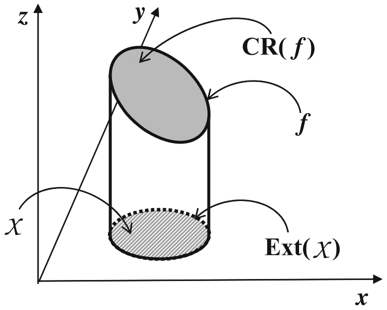
$$\mathrm{CR}(f)(x) := \sup\left\{ \sum_j p_j f(x_j) \middle| \sum_j p_j x_j = x,\ x_j \in \mathrm{Ext}(\mathcal{X}),\ p_j \geq 0 \right\}.$$

Especially, the function f on $\mathcal{X}$ is called **self convex roof** when $f = \mathrm{CR}(f_{\mathrm{Ext}(\mathcal{X})})$. Then, we can show the following lemma from the definition of the convex roof.

Lemma 2.1 *The convex roof* $\mathrm{CR}(f)$ *of* f *is a convex function.*

Further, we have the following lemma.

Fig. 2.1 Convex roof of the function f



Lemma 2.2 *Assume that a function f defined on the set $\mathrm{Ext}(\mathcal{X})$ of extremal points of a convex set $\mathcal{X}$ and a convex function g defined on the convex set $\mathcal{X}$ satisfies $f(x) = g(x)$ for any element in $x \in \mathrm{Ext}(\mathcal{X})$. Then, we have $\mathrm{CR}(f)(x) \geq g(x)$ for any element $x \in \mathcal{X}$.*

Proof We focus on a decomposition $\sum_j p_j x_j = x$ by extremal points satisfying that $\mathrm{CR}(f)(x) = \sum_j p_j f(x_j)$. Since g is a convex function, we have $g(x) \leq \sum_j p_j g(x_j) = \mathrm{CR}(f)(x)$, which implies the desired argument. ∎

Exercise 2.1 Show that the composite function $g \circ f$ is a convex function when f is a linear function and g is a convex function.

Exercise 2.2 Show that the map $\rho \mapsto |\operatorname{Tr} X\rho|^2$ is a convex function for any matrix X.

Exercise 2.3 Let f_a be a convex function defined in a convex set $\mathcal{X}$. Show that the map $x \mapsto \sup_a f_a(x)$ is a convex function.

2.3.2 Majorization and Schur Convex Function

Next, we introduce the concept of majorization to generalize the concept of a convex function. Given a vector $\boldsymbol{a} = (a_j)$ in $\mathbb{R}^d$, we denote the j-th largest entry by $a_j^{\downarrow}$. Then, given two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{R}^d$, we write $\boldsymbol{a} \succ \boldsymbol{b}$ when the relations $\sum_{j=1}^{k} a_j^{\downarrow} \geq \sum_{j=1}^{k} b_j^{\downarrow}$ and $\sum_{j=1}^{d} a_j = \sum_{j=1}^{d} b_j$ hold for $k = 1, \ldots, d$ as Fig. 2.2. Such a relation is called **majorization**. Given two density matrices ρ_1 and ρ_2, we write $\rho_1 \succ \rho_2$ when the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ composed of their eigenvectors satisfies $\boldsymbol{a} \succ \boldsymbol{b}$.

Then, a real-valued function f defined on $\mathbb{R}^d$ is called a **symmetric function** when f is invariant for any permutation $g \in S_d$, i.e., $f(g(\boldsymbol{a})) = f(\boldsymbol{a})$. In particular, a symmetric function f is called **Schur convex function** when the relation $f(\boldsymbol{a}) \geq f(\boldsymbol{b})$ holds for two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{R}^d$ satisfying $\boldsymbol{a} \succ \boldsymbol{b}$. When $-f$ is a Schur convex function, f is called a **Schur concave function**. The following lemma shows that Schur convex function and Schur concave function can be regarded as generalizations of convex function and concave function, respectively.

Lemma 2.3 *When a symmetric function f is a convex function, it is a Schur convex function.*

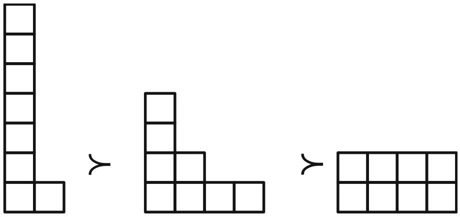


Fig. 2.2 Majorization relation

Proof Assume that $\boldsymbol{a} \succ \boldsymbol{b}$. Then, there exists a distribution p on the set S_d of permutations such that $\sum_{g \in S_d} p(g) g(\boldsymbol{a}) = \boldsymbol{b}$. Hence, we obtain $f(\boldsymbol{b}) \le \sum_{g \in S_d} p(g) f(g(\boldsymbol{a}))$ $= f(\boldsymbol{a})$, which implies the desired argument. ■

The following necessary and sufficient condition is known.

Lemma 2.4 *[91, p.57] Assume that a symmetric function f is partial differential possible. f is a Schur convex function if and only if the relation $(x_i - x_j)(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}) \ge 0$ holds for any i, j.*

Example 2.1 Define Schur function

$$\chi_{\boldsymbol{n}}(a_1, \ldots, a_d) := \frac{\sum_{\sigma \in S_r} \operatorname{sgn}(\sigma) \prod_{l=1}^{r} a_l^{n_{\sigma(l)} + \delta_{\sigma(l)}}}{\prod_{j<l}(a_j - a_l)} \tag{2.12}$$

on the domain $\mathbb{R}_+^d$. When $\boldsymbol{n} = (\underbrace{1, \ldots, 1}_{n}, 0, \ldots, 0)$, Schur function $\chi_{\boldsymbol{n}}(a_1, \ldots, a_d) = \sum_{1 \le i_1 < \cdots < i_n \le d} a_{i_1} \ldots a_{i_n}$ is a Schur concave function because of Lemma 2.4. When $\boldsymbol{n} = (n, 0, \ldots, 0)$, Schur function $\chi_{\boldsymbol{n}}(a_1, \ldots, a_d) = \sum_{1 \le i_1 \le \cdots \le i_n \le d} a_{i_1} \ldots a_{i_n}$ is a Schur convex function because of Lemma 2.4.

Exercise 2.4 Show that the uniform distribution on $\{1, \ldots, d\}$ is the minimum among all of distributions on $\{1, \ldots, d\}$ in the sense of majorization.

2.3.3 Probability Distribution and Inequalities

Next, we introduce useful inequalities when a probability distribution and a random variable X subject to the distribution. When we denote the expectation by E, the definition of convex function yields the following theorem.

Lemma 2.5 (Jensen inequality) *A convex function f satisfies the inequality* $\mathrm{E} f(X) \ge f(\mathrm{E} X)$.

Given a Hermitian matrix A and a real-valued function f defined on $\mathbb{R}$, we define the matrix $f(A) := \sum_i f(a_i) E_i$ by using the spectral decomposition $A = \sum_i a_i E_i$ of A. Now, we apply Jensen inequality to the distribution $\operatorname{Tr} E_i \rho$ defined by a state ρ. Then, a convex function f defined on $\mathbb{R}$ satisfies

$$\operatorname{Tr} \rho f(A) \ge f(\operatorname{Tr} \rho A). \tag{2.13}$$

The following inequality also holds.

Lemma 2.6 (Markov inequality) *When a random variable X takes values in real numbers, the relation* $\mathrm{P}\{X \ge a\} \le \frac{\mathrm{E} X}{a}$ *holds.*

Proof We show the desired inequality only when the random variable X is subject to a discrete distribution. In this case, the desired inequality can be shown as $\mathrm{P}\{X \ge a\} = \sum_{x:x\ge a} \mathrm{P}(x) \le \sum_{x:x\ge a} \mathrm{P}(x)\frac{x}{a} \le \sum_{x} \mathrm{P}(x)\frac{x}{a} = \frac{\mathrm{E}X}{a}$. ∎

2.3.4 Hölder Inequality and Its Related Topics

Given a general topological space Ω and a measure μ on it, we define the p norm $\|f\|_p := (\int_\Omega |f(\omega)|^p \mu(d\omega))^{1/p} = \||f|^p\|_1^{1/p}$ of a function f defined on Ω and a real number p, where p is allowed to take a negative value. When we choose another real number q such that $1/p + 1/q = 1$ and $\infty > p, q > 1$, any non-negative real numbers a and b are known to satisfy **Young inequality**: $ab \le \frac{a^p}{p} + \frac{b^q}{q}$, whose equality holds only when $a = b$. Using Young inequality, we can show $|fg(\omega)| \le \frac{|f(\omega)|^p}{p} + \frac{|g(\omega)|^q}{q}$. Taking the integral with the measure μ, we have

$$\|fg\|_1 \le \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q}. \tag{2.14}$$

Especially, letting $c := \|g\|_q^{q/p}/\|f\|_p$, we obtain $\|cf\|_p^p = \|g\|_q^q = \|cf\|_p\|g\|_q$. Hence, $\|cfg\|_1 \le \frac{\|cf\|_p^p}{p} + \frac{\|g\|_q^q}{q} = \|cf\|_p\|g\|_q$. Dividing by c, we obtain **Hölder inequality**:

$$\|fg\|_1 \le \|f\|_p\|g\|_q. \tag{2.15}$$

Conversely, when $1/p + 1/q = 1$ and $1 > p > 0 > q$, we obtain **reverse Hölder inequality**

$$\|fg\|_1 \ge \|f\|_p\|g\|_q, \tag{2.16}$$

which can be shown by Hölder inequality in the following way. Choose $s := -\frac{1}{q}$. Then, we apply Hölder inequality (2.15) to $p' := \frac{1+s}{s}$, $q' := 1 + s$ and functions $|fg|^{\frac{1}{1+s}}$ and $|f|^{-\frac{1}{1+s}}$. Then, we have

$$\begin{aligned} &\||fg|\|_1^{\frac{1}{1+s}} \||f|^{-\frac{1}{s}}\|_1^{\frac{s}{1+s}} = \|(|fg|^{\frac{1}{1+s}})^{1+s}\|_1^{\frac{1}{1+s}} \|(|f|^{-\frac{1}{1+s}})^{\frac{1+s}{s}}\|_1^{\frac{s}{1+s}} \\ \ge &\||fg|^{\frac{1}{1+s}}|f|^{-\frac{1}{1+s}}\|_1 = \||g|^{\frac{1}{1+s}}\|_1, \end{aligned}$$

which implies that $\||fg|\|_1^{\frac{1}{1+s}} \ge \||g|^{\frac{1}{1+s}}\|_1 \||f|^{-\frac{1}{s}}\|_1^{-\frac{s}{1+s}}$. Taking the $1 + s$-th power, we obtain (2.16).

When the measure μ is a probability measure, we apply the above discussion to two random variables X and Y. When $1/p + 1/q = 1$ and $\infty > p, q > 1$, the inequality (2.14) implies that $\mathrm{E}|XY| \le \frac{\mathrm{E}X^p}{p} + \frac{\mathrm{E}X^q}{q}$. Especially, when $p = q = 2$, we have the following lemma.

Lemma 2.7 *Two random variables X and Y satisfy* $2\mathrm{E}XY \le 2\mathrm{E}|XY| \le \mathrm{E}X^2 + \mathrm{E}Y^2$, *i.e.,* $\mathrm{E}(X+Y)^2 \le 2\mathrm{E}X^2 + 2\mathrm{E}Y^2$.

Next, we apply the above discussion to the case when Ω is the sequence space. Hölder inequality and reverse Hölder inequality guarantee the inequalities

$$\sum_{j=1}^{k} |a_j b_j| \le \Big(\sum_{j=1}^{k} |a_j|^p\Big)^{1/p} \Big(\sum_{j=1}^{k} |b_j|^q\Big)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1, \infty > p, q > 1$$

$$\sum_{j=1}^{k} |a_j b_j| \ge \Big(\sum_{j=1}^{k} |a_j|^p\Big)^{1/p} \Big(\sum_{j=1}^{k} |b_j|^q\Big)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1, 1 > p > 0 > q$$

for two sequences $\{a_j\}_{j=1}^k$ and $\{b_j\}_{j=1}^k$. Then, given a real number s, a probability distribution $\{p_j\}$, and a sequence $\{a_j\}$, the above inequalities imply

$$\sum_{j=1}^{k} |p_j^s a_j| \ge \Big(\sum_{j=1}^{k} |a_j|^p\Big)^{1/p} \Big(\sum_{j=1}^{k} p_j^{sq}\Big)^{1/q} = \Big(\sum_{j=1}^{k} |a_j|^{\frac{1}{1-s}}\Big)^{1-s}, \quad 0 > s \tag{2.17}$$

$$\sum_{j=1}^{k} |p_j^s a_j| \le \Big(\sum_{j=1}^{k} |a_j|^p\Big)^{1/p} \Big(\sum_{j=1}^{k} p_j^{sq}\Big)^{1/q} = \Big(\sum_{j=1}^{k} |a_j|^{\frac{1}{1-s}}\Big)^{1-s}, \quad 1 > s > 0 \tag{2.18}$$

$$\sum_{j=1}^{k} |p_j^s a_j| \ge \Big(\sum_{j=1}^{k} |a_j|^q\Big)^{1/q} \Big(\sum_{j=1}^{k} p_j^{sp}\Big)^{1/p} = \Big(\sum_{j=1}^{k} |a_j|^{\frac{1}{1-s}}\Big)^{1-s}, \quad s > 1, \tag{2.19}$$

where (2.17) and (2.18) follow with $q = \frac{1}{s}$ and (2.19) follows with $p = \frac{1}{s}$.

2.3.5 *Extension to Matrix Version*

It is known that the above discussion can be extended to a matrix version. A real-valued function f defined on a convex subset $\mathcal{S}$ of $\mathbb{R}$ is called a **matrix convex function** defined on the subset $\mathcal{S}$ when Hermitian matrices A and B whose eigenvalues belong to $\mathcal{S}$ satisfy $\lambda f(A) + (1-\lambda) f(B) \ge f(\lambda A + (1-\lambda)B)$ for $1 > \lambda > 0$. A function f is called a **matrix concave function** when $-f$ is matrix convex function. A matrix convex function is closely related to a matrix monotone function defined below. A real-valued function f defined on a convex subset $\mathcal{S}$ of $\mathbb{R}$ is called **matrix monotone** when the following condition: Hermitian matrices A and B satisfy $f(A) \ge f(B)$ when $A \ge B$ and their eigenvalues belong to $\mathcal{S}$. Further the following theorem is known [67, Corollary 2.5.6].

Theorem 2.3 *The following conditions are equivalent for a non-negative real-valued function f defined on* $[0, \infty)$.

(1) *f is matrix monotone.*
(2) *$t/f(t)$ is matrix monotone.*
(3) *f is a matrix concave function.*

For example, when $1 > s > 0$, the functions $x \mapsto -x^{-s}$, $x \mapsto x^s$, $x \mapsto \log x$, $x \mapsto -x^{1+s}$, and $x \mapsto -x \log x$ are matrix concave. The functions $x \mapsto x^s$, $x \mapsto -x^s$, and $x \mapsto \log x$ are matrix monotone. The functions $x \mapsto x^{1+s}$, $x \mapsto x^{-s}$, and $x \mapsto x \log x$ are matrix convex [67, Example 2.5.9].

Further, the following theorem is known for matrix convex functions [67, Theorem 2.5.7].

Theorem 2.4 *The following conditions are equivalent for a real-valued function f defined on* $[0, \infty)$.

(1) *f is matrix convex.*
(2) *When a matrix C satisfies $C^\dagger C \le I$ and the eigenvalues of a Hermitian matrix A belong to the domain of f, we have $f(C^\dagger AC) \le C^\dagger f(A)C$.*
(3) *Assume that a set of matrices $C_1, \ldots, C_k$ satisfies $\sum_j C_j^\dagger C_j = I$. When the eigenvalues of a Hermitian matrix A belong to the domain of f, we have $f(\sum_j C_j^\dagger AC_j) \le \sum_j C_j^\dagger f(A)C_j$.*

The following theorem is also known with respect to trace [9, 67].

Theorem 2.5 (Golden-Thompson) *Two Hermitian matrices A and B satisfy*

$$\operatorname{Tr} e^{A+B} \le \operatorname{Tr} e^A e^B. \tag{2.20}$$

Given a matrix A and a real number p, we define p-**norm** by $\|A\|_p := \||A|^p\|_1^{1/p}$. Then, the following matrix version of Hölder inequality holds [67].

Theorem 2.6 (matrix Hölder inequality) *When $1/p + 1/q = 1$ and $\infty > p, q > 1$, two matrix A and B satisfy*

$$\|AB\|_1 \le \|A\|_p \|B\|_q. \tag{2.21}$$

Using this inequality, we can show the reverse matrix Hölder inequality as follows.

Theorem 2.7 (matrix reverse Hölder inequality) *When $1/p + 1/q = 1$ and $1 > p > 0 > q$, two positive semi definite matrices A and B satisfy*

$$\|AB\|_1 \ge \|A\|_p \|B\|_q. \tag{2.22}$$

Proof We show the desired argument in the case when A and B are invertible. Otherwise, we can show the argument by taking the limit. Choose two matrices $a := \log A$ and $b := \log B$ and a real number $s := -\frac{1}{q}$. We apply matrix Hölder inequality (2.21) to two real numbers $p' := \frac{1+s}{s}$ and $q' := 1+s$ and two matrices $(e^{a+b})^{\frac{1}{1+s}}$ and $(e^{a})^{-\frac{1}{1+s}}$. Then,

$$\begin{aligned}
&\|e^{a+b}\|_1^{\frac{1}{1+s}}\|e^{-\frac{a}{s}}\|_1^{\frac{s}{1+s}} = \|(e^{\frac{a+b}{1+s}})^{1+s}\|_1^{\frac{1}{1+s}}\|(e^{-\frac{a}{1+s}})^{\frac{1+s}{s}}\|_1^{\frac{s}{1+s}} \\
\geq &\|e^{\frac{a+b}{1+s}}e^{-\frac{a}{1+s}}\|_1 \geq \|e^{\frac{b}{1+s}}\|_1,
\end{aligned}$$

where the final inequality follows from (2.20). Hence, $\|e^{a+b}\|_1^{\frac{1}{1+s}} \geq \|e^{\frac{b}{1+s}}\|_1 \|e^{-\frac{a}{s}}\|_1^{-\frac{s}{1+s}}$. Taking the $1+s$-th power, and using the inequality $\|e^{a}e^{b}\|_1 \geq \|e^{a+b}\|_1$ shown by (2.20), we obtain (2.22). ■

Hence, given a real number s, a non-negative matrix A, and a density matrix ρ, similar to (2.17), (2.18), and (2.19), applying matrix Hölder inequality (2.21) and matrix reverse Hölder inequality (2.22), we obtain the following inequalities

$$\|\rho^{s}A\|_1 \geq \|A\|_{\frac{1}{1-s}}, \quad 0 > s \tag{2.23}$$

$$\|\rho^{s}A\|_1 \leq \|A\|_{\frac{1}{1-s}}, \quad 1 > s > 0 \tag{2.24}$$

$$\|\rho^{s}A\|_1 \geq \|A\|_{\frac{1}{1-s}}, \quad s > 1. \tag{2.25}$$

2.4 Information Quantities on Quantum System

2.4.1 Information Quantities of Information Source

In quantum information, to measure the amount of information of the system $\mathcal{H}$ whose state is ρ, we often define information quantities as functions of ρ. Usually, information quantities satisfy convexity or concavity. Now, we additionally focus on the following condition.

Additivity: A function $f(\rho)$ is called additive when $f(\rho_1 \otimes \rho_2) = f(\rho_1) + f(\rho_2)$.

The most important information quantity is **von Neumann entropy**, which is defined as

$$H(\rho) := -\operatorname{Tr}\rho\log\rho. \tag{2.26}$$

Larger von Neumann entropy $H(\rho)$ implies larger randomness in the system $\mathcal{H}$. As discussed in Sect. 6.6, von Neumann entropy $H(\rho)$ expresses the limit of the compression rate in the data compression. Except for von Neumann entropy, the quantity $\psi(s|\rho) := \log\operatorname{Tr}\rho^{1-s}$ expresses the randomness. Using this quantity, we can define **Rényi entropy** of order $1-s$: $H_{1-s}(\rho) := \frac{\psi(s|\rho)}{s}$. Taking the limit $s \to 0$, we obtain von Neumann entropy $H(\rho) = \lim_{s\to 0} H_{1-s}(\rho)$. Larger Rényi entropy $H_{1-s}(\rho)$ also implies larger randomness.

As mentioned in Corollary 2.3 later, von Neumann entropy $H(\rho)$ satisfies concavity, and $e^{\psi(s|\rho)}$ satisfies convexity with $s < 0$ and concavity with $0 < s < 1$. Then, we define min-entropy $H_{\min}(\rho) := -\log \|\rho\|$ andmax-entropy $H_{\max}(\rho) := \log \operatorname{rank} \rho$. These quantities can be characterized as the limit of Rényi entropy in the following way:

$$H_{\min}(\rho) = \lim_{s\to -\infty} H_{1-s}(\rho), \quad H_{\max}(\rho) = \lim_{s\to 1} H_{1-s}(\rho).$$

For $s \neq 0$, the second derivative of $\psi(s|\rho)$ with respect to s can be calculated as

$$\psi''(s|\rho) = \frac{\operatorname{Tr}(\log \rho)^2 \rho^s \operatorname{Tr} \rho^s - (\operatorname{Tr}(\log \rho)\rho^s)^2}{(\operatorname{Tr} \rho^s)^2} > 0. \tag{2.27}$$

Hence, we find that the function $s \mapsto \psi(s|\rho)$ is a convex function. Further, since $\psi(0|\rho) = 0$, $H_{1-s}(\rho) = \frac{\psi(s|\rho)}{s}$ is a monotone increasing with respect to s. Thus, for $s_1 < 0$ and $0 < s_2 < 1$, we have

$$H_{\min}(\rho) \le H_{1-s_1}(\rho) \le H(\rho) \le H_{1-s_2}(\rho) \le H_{\max}(\rho).$$

$H_{\max}(\rho)$ satisfies concavity due to its definition. von Neumann entropy, Rényi entropy, min-entropy, and max-entropy satisfies the additivity.

The second derivative $\psi''(0|\rho)$ is called **varentropy** and is denoted by $V(\rho)$. This quantity is simplified as

$$V(\rho) = \operatorname{Tr} \rho(\log \rho)^2 - H(\rho)^2. \tag{2.28}$$

It expresses the variance when $\log \rho$ is regarded as a random variable. Also, it plays an important role in the second order analysis, as discussed in Chap. 6.

When the state of the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ is given as the pure state $|\Phi\rangle\langle\Phi|$, the set of non-zero eigenvalues of the reduced density matrix $\rho_A = \operatorname{Tr}_B |\Phi\rangle\langle\Phi|$ is the same as that of the other reduced density matrix $\rho_B = \operatorname{Tr}_A |\Phi\rangle\langle\Phi|$. Hence, all of the above information quantities of ρ_A are the same as those of ρ_B. Further, for a probability distribution $\boldsymbol{p} = (p_1, \ldots, p_k)$, we define the density matrix $\rho(\boldsymbol{p}) := \sum_{i=1}^k p_i |i\rangle\langle i|$, **Shannon entropy** $H(\boldsymbol{p}) := H(\rho(\boldsymbol{p}))$, and $\psi(s|\boldsymbol{p}) := \psi(s|\rho(\boldsymbol{p}))$. Then, when $\dim \mathcal{H} = d$, the concavity of von Neumann entropy implies that $H(\rho) \le \log d$. Indeed, the second derivative $\phi''(0|\rho)$ contains the term $\operatorname{Tr} \rho(\log \rho)^2$, and we can maximize $\operatorname{Tr} \rho(\log \rho)^2$ as follows [47]. When $d = 2$,

$$\max_{\operatorname{Tr}\rho=1} \operatorname{Tr}\rho(\log\rho)^2 = \frac{1-\sqrt{1-4/e^2}}{2}(\log\frac{1-\sqrt{1-4/e^2}}{2})^2 + \frac{1+\sqrt{1-4/e^2}}{2}(\log\frac{1+\sqrt{1-4/e^2}}{2})^2.$$

When $d \geq 3$,

$$\max_{\operatorname{Tr}\rho=1} \operatorname{Tr}\rho(\log\rho)^2 = (\log d)^2. \tag{2.29}$$

The case of $d = 2$ can be shown by the simple calculation, and the case of $d \geq 3$ can be shown as follows.

Proof of (2.29). We can show (2.29) in the case of $d = 3$ by calculation. Now, we will show (2.29) in the case of $d \geq 4$ by induction. Let a_d be the LHS of (2.29). Then, we have $a_d \geq (\log d)^2$. Due to the assumption of induction, there exists eigenvalues $\{p_i\}$ of ρ such that $a_d = \sum_{i=1}^{d} p_i(\log p_i)^2$ and $p_i > 0$ $(i = 1, \ldots, d)$.

Now, we apply Lagrange multiplier method. Let λ' be Lagrange multiplier. we obtain $(\log p_i)^2 + 2\log p_i - \lambda' = 0$. Letting $\lambda := \sqrt{1+\lambda'}$, we find that the solution is $\log p_i = -1 \pm \lambda$. Then, there exists $0 \leq r \leq d$ such that the relation

$$\log p_i = \begin{cases} -1+\lambda & \text{if } r \geq i \\ -1-\lambda & \text{if } r < i. \end{cases}$$

holds by reordering p_i. Since $\sum_i p_i = 1$, we have $1 = re^{-1+\lambda} + (d-r)e^{-1-\lambda}$. the real number $x := e^{\lambda}$ satisfies $rx^2 - ex + d - r = 0$. Since the discriminant is greater than 0, we have $e^2 - 4r(d-r) \geq 0$. Solving this inequality, we have $r \leq \frac{d-\sqrt{d^2-e^2}}{2}$, $\frac{d+\sqrt{d^2-e^2}}{2} \leq r$. The function $c(x) := \frac{x-\sqrt{x^2-e^2}}{2}$ is monotone decreasing function on (e, ∞), and satisfies the condition $c(4) < 1$. So, $c(x) < 1$ for $x \geq 4$. Hence, the above inequality for r guarantees that r is 0 or d. Thus, $p_i = 1/d$, which implies (2.29). ■

2.4.2 Information Measures for Two Information Sources

There are several measures for the difference between two density matrices ρ_1 and ρ_2 on the system $\mathcal{H}$. As requirements of such measures, we impose the following properties for a function $f(\rho_1, \rho_2)$ of two density matrices ρ_1 and ρ_2.

Joint convexity: $\lambda f(\rho_1, \rho_2) + (1-\lambda) f(\rho_1', \rho_2') \geq f(\lambda\rho_1 + (1-\lambda)\rho_1', \lambda\rho_2 + (1-\lambda)\rho_2')$.

Additivity: $f(\rho_1 \otimes \rho_1', \rho_2 \otimes \rho_2') = f(\rho_1, \rho_2) + f(\rho_1', \rho_2')$.

Pseudo square distance: We choose one-parametric subset (curve) $\{\rho_t | t \in (0, 1)\}$ of $\mathcal{S}(\mathcal{H})$. Then, for any $tin(0, 1)$ the relation $f(\rho_t, \rho_{t+\epsilon}) = c_t\epsilon^2 + o(\epsilon^2)$ holds with a suitable constant c_t.

Generally, when a function $f(\rho_1, \rho_2)$ is given as the square of a distance, $f(\rho_1, \rho_2)$ is a pseudo square distance.

In the following, we discuss several examples for measures for the difference between two density matrices ρ_1 and ρ_2. The first one is **trace norm distance** $d_1(\rho_1, \rho_2) := \frac{1}{2}\|\rho_1 - \rho_2\|_1$, which is given by the trace norm. The second one is the **fidelity** $F(\rho_1, \rho_2) := \mathrm{Tr}\,|\sqrt{\rho_1}\sqrt{\rho_2}|$. When ρ_1 or ρ_2 is pure, its square is $\mathrm{Tr}\,\rho_1\rho_2$. Using this relation, the **pseuod fidelity** is defined by $\tilde{F}(\rho_1, \rho_2) := \mathrm{Tr}\,\rho_1\rho_2$. Based on these quantities, we define **logarithmic inverse square fidelity** $F_{\log}(\rho_1, \rho_2) := -2\log \mathrm{Tr}\,|\sqrt{\rho_1}\sqrt{\rho_2}|$, and **logarithmic inverse fidelity** $\tilde{F}_{\log}(\rho_1, \rho_2) := -\log \mathrm{Tr}\,\rho_1\rho_2$. Using, we can define **Bures distance** $b(\rho_1, \rho_2)^2 := 1 - \mathrm{Tr}\,|\sqrt{\rho_1}\sqrt{\rho_2}|$. Further, trace norm distance and Bures distance satisfy the axioms of distance.

As another type of measures, we can measure the difference between two density matrices ρ_1 and ρ_2 by modifying the information quantity in the previous subsection. Firstly, based on von Neumann entropy, we define **relative entropy** $D(\rho_1\|\rho_2) := \mathrm{Tr}\,\rho_1(\log\rho_1 - \log\rho_2)$. Also, we define $\psi(s|\rho_1\|\rho_2) := \log \mathrm{Tr}\,\rho_1^{1-s}\rho_2^{s}$. Using this function, we define **relative Renyi entropy** of order $1-s$: $D_{1-s}(\rho_1\|\rho_2) := -\frac{\psi(s|\rho_1\|\rho_2)}{s}$. Taking the limit $s \to 0$, we can recover the relative entropy $D(\rho_1\|\rho_2) = \lim_{s\to 0} D_{1-s}(\rho_1\|\rho_2)$. Further, larger relative Rényi entropy $D_{1-s}(\rho_1\|\rho_2)$ implies larger difference between two density matrices ρ_1 and ρ_2. When $s = 1/2$, $e^{\psi(s|\rho_1\|\rho_2)}$ is analogous to the fidelity $F(\rho_1, \rho_2)$. However, when ρ_1 is not commutative with ρ_2, they are different. In general, we have the relation.

$$F(\rho_1, \rho_2) = \mathrm{Tr}\,|\sqrt{\rho_1}\sqrt{\rho_2}| \geq \mathrm{Tr}\,\sqrt{\rho_1}\sqrt{\rho_2} = e^{\psi(1/2|\rho_1\|\rho_2)}, \tag{2.30}$$

whose equality holds when ρ_1 is commutative with ρ_2.

Further, we define **relative min-entropy** $D_{\min}(\rho_1\|\rho_2) := -\log \mathrm{Tr}\,\rho_2 P(\rho_1)$ and **relative max-entropy** $D_{\max}(\rho_1\|\rho_2) := \log \|\rho_2^{-\frac{1}{2}}\rho_1\rho_2^{-\frac{1}{2}}\|$, where $P(\rho)$ is the projection to the range of ρ. Here we should remark that max and min are assigned in the way opposite to the case with min-entropy and max-entropy. Then, the relative min-entropy $D_{\min}(\rho_1\|\rho_2)$ is given as the limit of relative Rényi entropy, i.e., $D_{\min}(\rho_1\|\rho_2) = \lim_{s\to 1} D_{1-s}(\rho_1\|\rho_2)$. On the other hand, the relative max-entropy $D_{\max}(\rho_1\|\rho_2)$ satisfies $\|\rho_2^{-\frac{1}{2}}\rho_1\rho_2^{-\frac{1}{2}}\| = \|\rho_1^{\frac{1}{2}}\rho_2^{-1}\rho_1^{\frac{1}{2}}\|$. Also, this quantity is the minimum real number λ satisfying $\lambda\rho_2 \geq \rho_1$. Since $\mathrm{Tr}\,\rho_1(\rho_1^{\frac{1}{2}}\rho_2^{-1}\rho_1^{\frac{1}{2}}) \leq \|\rho_1^{\frac{1}{2}}\rho_2^{-1}\rho_1^{\frac{1}{2}}\|$, we have

$$D_{-1}(\rho_1\|\rho_2) = \log \mathrm{Tr}\,\rho_1(\rho_1^{\frac{1}{2}}\rho_2^{-1}\rho_1^{\frac{1}{2}}) \leq \log \|\rho_1^{\frac{1}{2}}\rho_2^{-1}\rho_1^{\frac{1}{2}}\| = D_{\max}(\rho_1\|\rho_2).$$

So, the function $s \mapsto \psi(s|\rho_1\|\rho_2)$ is a convex function. Hence, since $\psi(0|\rho_1\|\rho_2) = 0$, $D_{1-s}(\rho_1\|\rho_2) := \frac{-\psi(s|\rho_1\|\rho_2)}{s}$ is a monotone decreasing with respect to s. Thus, for $-1 \leq s_1 < 0, 0 < s_2 < 1$, we obtain

$$D_{\min}(\rho_1\|\rho_2) \leq D_{s_2}(\rho_1\|\rho_2) \leq D(\rho_1\|\rho_2) \leq D_{s_1}(\rho_1\|\rho_2) \leq D_{\max}(\rho_1\|\rho_2). \tag{2.31}$$

As mentioned in Corollary 2.2 later, trace norm distance, square of Bures distance, relative entropy, relative min-entropy, and $e^{\psi(s|\rho_1\|\rho_2)}$ with $-1 < s < 0$ satisfy joint convexity. $-e^{\psi(s|\rho_1\|\rho_2)}$ with $1 > s > 0$ and $-F(\rho_1, \rho_2)$ also satisfy joint convexity. The square of fidelity $F(\rho_1, \rho_2)^2$ satisfies convexity with respect to each entry ρ_1 and ρ_2. Also, relative entropy, relative Rényi entropy, relative min-entropy, relative max-entropy logarithmic inverse square fidelity, and logarithmic inverse pseudo fidelity satisfy the additivity. Then, relative entropy, relative Rényi entropy, relative min-entropy, and relative max-entropy are pseudo square distances. Further, for two distributions $\boldsymbol{p} = (p_1, \ldots, p_k)$ and $\boldsymbol{q} = (p_1, \ldots, p_k)$, we can define **relative entropy** $D(\boldsymbol{p}\|\boldsymbol{q}) := D(\rho(\boldsymbol{p})\|\rho(\boldsymbol{q}))$ by using the diagonal density matrices $\rho(\boldsymbol{p})$ and $\rho(\boldsymbol{q})$.

As a comparison between the states before and after the application of pinching $\Lambda_X(\rho) = \sum_i E_i \rho E_i$ with respect to the Hermitian matrix $X = \sum_i x_i E_i$, we obtain the following relation.

$$D(\rho\|\Lambda_X(\rho)) = H(\Lambda_X(\rho)) - H(\rho). \tag{2.32}$$

In particular, since there exist unitaries U and U' such that $\|(I - E_i)\rho\|_1 = \mathrm{Tr}(I - E_i)\rho U$ and $\|E_i\rho(I - E_i)\|_1 = \mathrm{Tr}\, E_i\rho(I - E_i)U'$, Schwartz inequality for the inner product $\mathrm{Tr}\, X^\dagger \rho Y$ between two matrices X and Y implies that

$$\begin{aligned}
&\|\rho - E_i\rho E_i\|_1 = \|(I - E_i)\rho + E_i\rho(I - E_i)\|_1 \\
=&\|(I - E_i)\rho\|_1 + \|E_i\rho(I - E_i)\|_1 = \mathrm{Tr}(I - E_i)\rho U + \mathrm{Tr}\, E_i\rho(I - E_i)U' \\
\le&\sqrt{\mathrm{Tr}(I - E_i)\rho(I - E_i)\,\mathrm{Tr}\, U^\dagger\rho U} \\
&+ \sqrt{\mathrm{Tr}\, E_i\rho E_i\, \mathrm{Tr}\, U'^\dagger(I - E_i)\rho(I - E_i)U'} \\
=&\sqrt{\mathrm{Tr}(I - E_i)\rho} + \sqrt{\mathrm{Tr}\, E_i\rho\, \mathrm{Tr}(I - E_i)\rho} \le 2\sqrt{\mathrm{Tr}(I - E_i)\rho}.
\end{aligned} \tag{2.33}$$

Hence, trace norm distance between the states before and after the application of pinching Λ_X is evaluated as

$$\begin{aligned}
&\|\rho - \Lambda_X(\rho)\|_1 = \|(\rho - E_i\rho E_i) + \sum_{j\neq i} E_j\rho E_j\|_1 \\
\le&\|\rho - E_i\rho E_i\| + \mathrm{Tr}\sum_{j\neq i} E_j\rho E_j \\
\le&\sqrt{\mathrm{Tr}(I - E_i)\rho} + \sqrt{\mathrm{Tr}\, E_i\rho\, \mathrm{Tr}(I - E_i)\rho} + \mathrm{Tr}(I - E_i)\rho \\
\le&3\sqrt{\mathrm{Tr}(I - E_i)\rho}.
\end{aligned} \tag{2.34}$$

2.4.3 Correlation Between Two Systems

To express the correlation between two systems $\mathcal{H}_A$ and $\mathcal{H}_B$, we define **mutual information** $I_\rho(A:B) := H(\rho_A) + H(\rho_B) - H(\rho)$ for a state ρ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$. As discussed in Exercise 2.5, this quantity is $D(\rho\|\rho_A \otimes \rho_B)$. Hence, $I_\rho(A:B)$ can be regarded as a difference from the tensor product state when the state is ρ. Thus, $I_\rho(A:B)$ is used as a quantity to express the correlation in the state ρ.

Related to mutual information, relative entropy have several important properties as follows. Now, we focus on the relative entropy between a state ρ on the composite system and an arbitrary tensor product state $\sigma_A \otimes \sigma_B$. Then,

$$\begin{aligned}
&D(\rho\|\sigma_A \otimes \sigma_B) = \operatorname{Tr} \rho(\log \rho - \log \sigma_A \otimes \sigma_B)\\
=&\operatorname{Tr} \rho(\log \rho - (\log \sigma_A) \otimes I_B - I_A \otimes (\log \sigma_B))\\
=&\operatorname{Tr} \rho(\log \rho - (\log \rho_A) \otimes I_B - I_A \otimes (\log \rho_B))\\
&+ \operatorname{Tr} \rho((\log \rho_A) \otimes I_B + I_A \otimes (\log \rho_B) - (\log \sigma_A) \otimes I_B - I_A \otimes (\log \sigma_B))\\
=&\operatorname{Tr} \rho(\log \rho - \log(\rho_A \otimes \rho_B))\\
&+ \operatorname{Tr} \rho_A(\log \rho_A - \log \sigma_A) + \operatorname{Tr} \rho_B(\log \rho_B - \log \sigma_B)\\
=&D(\rho\|\rho_A \otimes \rho_B) + D(\rho_A\|\sigma_A) + D(\rho_B\|\sigma_B)\\
=&D(\rho\|\rho_A \otimes \rho_B) + D(\rho_A \otimes \rho_B\|\sigma_A \otimes \sigma_B).
\end{aligned} \tag{2.35}$$

That is, we have the following Pythagorean theorem.

Theorem 2.8 (Pythagorean theorem) *We have the following relation.*

$$D(\rho\|\sigma_A \otimes \sigma_B) = D(\rho\|\rho_A \otimes \rho_B) + D(\rho_A \otimes \rho_B\|\sigma_A \otimes \sigma_B). \tag{2.36}$$

Especially, when $\sigma_A = \rho_A$, we have

$$D(\rho\|\rho_A \otimes \sigma_B) = D(\rho\|\rho_A \otimes \rho_B) + D(\rho_B\|\sigma_B). \tag{2.37}$$

This theorem has the following meaning. Consider the set $\mathcal{S}$ of tensor product density matrices. The relative entropy$D(\rho\|\sigma_A \otimes \sigma_B)$ from a density matrix ρ to a tensor product density matrix $\sigma_A \otimes \sigma_B$ is written as the sum of the relative entropy $D(\rho\|\rho_A \otimes \rho_B)$ from the density matrix ρ to the closed point $\rho_A \otimes \rho_B$ in $\mathcal{S}$ and the relative entropy $D(\rho_A \otimes \rho_B\|\sigma_A \otimes \sigma_B)$ from $\rho_A \otimes \rho_B$ to $\sigma_A \otimes \sigma_B$. That is, as shown in Fig. 2.3, the closet point $\rho_A \otimes \rho_B$ is the foot of perpendicular from the density matrix ρ to the set $\mathcal{S}$. This fact yields that

$$D(\rho\|\sigma_A \otimes \sigma_B) \geq D(\rho\|\rho_A \otimes \rho_B). \tag{2.38}$$

When the input system $\mathcal{H}_A$ of the quantum channel is a classical system $\mathcal{X} := \{1, \ldots, k\}$, the channel is given by the output state $W(x)$ on the output system $\mathcal{H}_B$

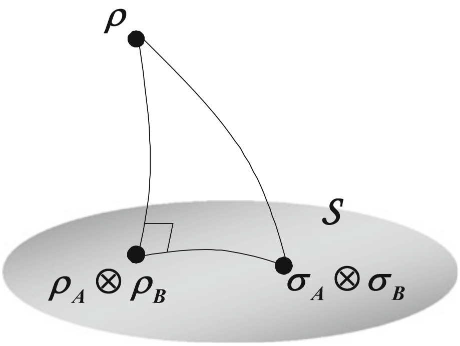


Fig. 2.3 Pythagorean theorem

depending on the input information $x \in \mathcal{X}$. As explained in Sect. 6.7, the map $W : x \mapsto W(x)$ is called a classical-quantum channel. Then, we often focus on the mutual information of the composite system under the state $\sum_{x=1}^{k} p(x)|x\rangle\langle x| \otimes W(x)$.

This value is called **quantum transmission information**, and equals

$$\begin{aligned} I(p, W) :=& H\Big(\sum_{x=1}^{k} p(x)W(x)\Big) - \sum_{x=1}^{k} p(x)H(W(x)) \\ =& \sum_{x=1}^{k} p(x)D\Big(W(x)\Big\| \sum_{x=1}^{k} p(x)W(x)\Big). \end{aligned} \tag{2.39}$$

The concavity of von Neumann entropy guarantees that $I(p, W)$ is a concave function with respect to the probability distribution p. Applying the above inequality (2.37), we have

$$\begin{aligned} &\sum_{x=1}^{k} p(x)D(W(x)\|\sigma) \\ =& \sum_{x=1}^{k} p(x)D\Big(W(x)\Big\| \sum_{x=1}^{k} p(x)W(x)\Big) + D\Big(\sum_{x=1}^{k} p(x)W(x)\Big\|\sigma\Big) \end{aligned} \tag{2.40}$$

for any density matrix σ on the system $\mathcal{H}_B$.

We also have the group invariant version of Pythagorean theorem.

Theorem 2.9 *we consider a unitary representation* f *of a compact group* G *on the system* $\mathcal{H}_B$. *Then, we assume that*σ_B *is invariant with respect to* f, *i.e., the relation* $\mathsf{f}(g)\sigma_B = \sigma_B \mathsf{f}(g)$ *holds for any element* $g \in G$. *Now, using the invariant measure* μ_G, *we define the averaged state* $\overline{\rho} := \int_G \mathsf{f}(g)\rho\mathsf{f}(g)^\dagger \mu_G(dg)$, *Then, we have*

$$D(\rho\|\sigma) = D(\rho\|\overline{\rho}) + D(\overline{\rho}\|\sigma) = H(\overline{\rho}) - H(\rho) + D(\overline{\rho}\|\sigma) \tag{2.41}$$

The following is the reason why this theorem can be regarded as the group invariant version of Pythagorean theorem. Let $\mathcal{S}$ be the subset of density matrices invariant with respect to f. The relative entropy $D(\rho\|\sigma)$ from a density matrix ρ to a invariant density matrix σ is the sum of the relative entropy $D(\rho\|\overline{\rho})$ from the density matrix ρ to the point $\overline{\rho}$ in $\mathcal{S}$ closest to ρ in the sense of relative entropy and the relative entropy $D(\overline{\rho}\|\sigma)$ from $\overline{\rho}$ to σ. That is, in the same way as explained in Fig. 2.3, $\overline{\rho}$ is the foot of perpendicular from the density matrix ρ to the set $\mathcal{S}$.

Proof First, we notice that the averaged state $\overline{\rho}$ is invariant with respect to f. So, since any density matrix ρ satisfies $D(\mathsf{f}(g)\rho\mathsf{f}(g)^\dagger\|\sigma) = D(\rho\|\sigma)$, we have

$$\begin{aligned} D(\rho\|\sigma) =& \int_G D(\mathsf{f}(g)\rho\mathsf{f}(g)^\dagger\|\sigma)\mu_G(dg) \\ =& \int_G D(\mathsf{f}(g)\rho\mathsf{f}(g)^\dagger\|\overline{\rho})\mu_G(dg) + D(\overline{\rho}\|\sigma) \\ =& D(\rho\|\overline{\rho}) + D(\overline{\rho}\|\sigma) = H(\overline{\rho}) - H(\rho) + D(\overline{\rho}\|\sigma). \end{aligned} \tag{2.42}$$

■

Exercise 2.5 Show that $I_\rho(A:B) = D(\rho\|\rho_A \otimes \rho_B)$.

Exercise 2.6 Show that any eigenvalue λ of a density matrix ρ on a system $\mathcal{H}$ satisfies that

$$H(\rho) \le (1-\lambda)\log\dim\mathcal{H} + h(\lambda), \tag{2.43}$$

where $h(x) := -x\log x - (1-x)\log(1-x)$.

Exercise 2.7 Show the following type of Pythagorean theorem.

$$D(\rho\|\sigma_A \otimes \sigma_B) = D(\rho\|\sigma_A \otimes \rho_B) + D(\rho_B\|\sigma_B). \tag{2.44}$$

Exercise 2.8 We consider the pinching Λ_E with respect to the spectral decomposition E of σ. Show the following relations by using Theorem 2.9.

$$\begin{aligned} D(\rho\|\sigma) =& D(\rho\|\Lambda_E(\rho)) + D(\Lambda_E(\rho)\|\sigma) \\ =& H(\Lambda_E(\rho)) - H(\rho) + D(\Lambda_E(\rho)\|\sigma). \end{aligned} \tag{2.45}$$

2.4.4 Quantum Channel and Information Quantities

Next, we discuss information quantities of a quantum channel based on the above given information quantities. First, when the input and output systems of Λ are $\mathcal{H}$, we introduce the entanglement fidelity as a quantity to express how precisely the quantum channel Λ transmits the input state. We prepare the reference system $\mathcal{H}_R$ whose dimension is the same as that of $\mathcal{H}$. Given an input state ρ on the system $\mathcal{H}$, we choose a purification $|\Phi\rangle$ of ρ on the composite system $\mathcal{H} \otimes \mathcal{H}_R$. Then, we define **entanglement fidelity** to be [111]

$$F_e^2(\rho, \Lambda) := \langle \Phi | (\Lambda \otimes id)(|\Phi\rangle\langle\Phi|) |\Phi\rangle. \tag{2.46}$$

Remark that this quantity does not depend on the choice of the purification $|\Phi\rangle$. By using a Kraus representation $\{F_m\}$ of Λ, the entanglement fidelity is written as

$$F_e^2(\rho, \Lambda) = \sum_m |\operatorname{Tr} F_m \rho|^2. \tag{2.47}$$

Since the function $\rho \mapsto |\operatorname{Tr} F_m \rho|^2$ is a convex function, $F_e^2(\rho, \Lambda)$ is convex with respect to ρ. Especially, when ρ is a pure state, $F_e^2(\rho, \Lambda)$ is the square $F(\rho, \Lambda(\rho))^2$ of the fidelity. Hence, when $\rho = \sum_j p_j |x_j\rangle\langle x_j|$, we have

$$F_e^2(\rho, \Lambda) \le \sum_j p_j \langle x_j | \Lambda(|x_j\rangle\langle x_j|) |x_j\rangle. \tag{2.48}$$

When ρ on the system $\mathcal{H}$ is the completely mixed state ρ_{mix}, we simplify $F_e^2(\rho, \Lambda)$ to $F_e^2(\Lambda)$. Let Θ be the set of pure states on $\mathcal{H} = \mathbb{C}^r$, which is a homogeneous space with respect to the Lie group SU(r). Then, the invariant measure μ_Θ on Θ satisfies [45, (8.27)]

$$\frac{r}{r+1}(1 - F_e^2(\Lambda)) = 1 - \int_\Theta \langle \theta | \Lambda(|\theta\rangle\langle\theta|) |\theta\rangle \mu_\Theta(d\theta).$$

This equation shows that the entanglement fidelity $F_e^2(\Lambda)$ of the quantum channel Λ gives the average performance when a pure state is transmitted via the the quantum channel Λ. Hence, we can evaluate the performance of a channel by evaluating its entanglement fidelity.

Next, we introduce **coherent information** $I_c(\rho, \Lambda) := H(\Lambda(\rho)) - H(\Lambda \otimes id_R (|\Phi\rangle\langle\Phi|))$ to measure how much the quantum channel Λ keeps the coherence [112], where $|\Phi\rangle$ is a purification of the state ρ. This quantity can be defined even when the output system $\mathcal{H}_B$ is different from the input system $\mathcal{H}_A$, and it does not depend on the choice of the purification $|\Phi\rangle$. Now, we choose a Stinespring representation U of Λ. So, $U \otimes I(|\Phi\rangle\langle\Phi|)U^\dagger \otimes I$ is a purification of $\Lambda \otimes id_R(|\Phi\rangle\langle\Phi|)$. Then, we have $\operatorname{Tr}_{B,R} U \otimes I(|\Phi\rangle\langle\Phi|)U^\dagger \otimes I = \operatorname{Tr}_B U\rho U^\dagger = \Lambda_E(\rho)$. That is, the set of non-zero eigenvalues of the density matrix $\Lambda \otimes id_R(|\Phi\rangle\langle\Phi|)$ is the same as that of the density matrix $\Lambda_E(\rho)$. Hence, we have $H(\Lambda \otimes id_R(|\Phi\rangle\langle\Phi|)) = H(\Lambda_E(\rho))$, which implies $I_c(\rho, \Lambda) = H(\Lambda(\rho)) - H(\Lambda_E(\rho))$.

Exercise 2.9 Show (2.47).

Exercise 2.10 Show (2.47).

Exercise 2.11 Show (2.48) by using Exercise 1.11.

2.5 Qubit System

Next, we consider a typical example of a quantum system, the two-dimensional system $\mathbb{C}^2$, which can be regarded as a quantum analogue of "bit" and is called qubit. Using a three-dimensional vector $\boldsymbol{x} \in \mathbb{R}^3$ and three matrices

$$E^1 := \frac{1}{2}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad E^2 := \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad E^3 := \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{2.49}$$

we can characterize any density matrix on the qubit system as $\rho_{\boldsymbol{x}} := \frac{1}{2}I + \sum_{j=1}^{3} x_j E^j$. Then, the eigenvectors of the Hermitian matrix $\rho_{\boldsymbol{x}}$ are $\frac{1+\|\boldsymbol{x}\|}{2}$ and $\frac{1-\|\boldsymbol{x}\|}{2}$. The Hermitian matrix $\rho_{\boldsymbol{x}}$ is a density matrix if and only if $\|\boldsymbol{x}\| \le 1$. Then, it is a pure state if and only if $\|\boldsymbol{x}\| = 1$. Hence, we have $H(\rho_{\boldsymbol{x}}) = h(\frac{1+\|\boldsymbol{x}\|}{2})$, where the function $h(x)$ is **binary entropy** and is defined as $h(x) := -x\log(x) - (1-x)\log(1-x)$. The information quantities for two states defined in Sect. 2.4 are calculated as

$$\begin{aligned}
F(\rho_{\boldsymbol{x}}, \rho_{\boldsymbol{x}'})^2 =& \frac{1 + \boldsymbol{x}\cdot\boldsymbol{x}' + \sqrt{1-\|\boldsymbol{x}\|^2}\sqrt{1-\|\boldsymbol{x}'\|^2}}{2} \qquad (2.50)\\
D(\rho_{\boldsymbol{x}}\|\rho_{\boldsymbol{x}'}) =& h(\frac{1+\|\boldsymbol{x}\|}{2}) - \frac{1+\boldsymbol{x}\cdot\boldsymbol{x}'/\|\boldsymbol{x}'\|}{2}\log(\frac{1+\|\boldsymbol{x}'\|}{2}) \\
& - \frac{1-\boldsymbol{x}\cdot\boldsymbol{x}'/\|\boldsymbol{x}'\|}{2}\log(\frac{1-\|\boldsymbol{x}'\|}{2}) \\
d_1(\rho_{\boldsymbol{x}}, \rho_{\boldsymbol{x}'}) =& \frac{\|\boldsymbol{x}-\boldsymbol{x}'\|}{2} \\
e^{\psi(s|\rho_{\boldsymbol{x}}\|\rho_{\boldsymbol{x}'})} =& \frac{1+\boldsymbol{x}\cdot\boldsymbol{x}'/(\|\boldsymbol{x}\|\cdot\|\boldsymbol{x}'\|)}{2}(\frac{1+\|\boldsymbol{x}\|}{2})^{1-s}(\frac{1+\|\boldsymbol{x}'\|}{2})^{s} \\
&+ \frac{1-\boldsymbol{x}\cdot\boldsymbol{x}'/(\|\boldsymbol{x}\|\cdot\|\boldsymbol{x}'\|)}{2}(\frac{1-\|\boldsymbol{x}\|}{2})^{1-s}(\frac{1+\|\boldsymbol{x}'\|}{2})^{s} \\
&+ \frac{1-\boldsymbol{x}\cdot\boldsymbol{x}'/(\|\boldsymbol{x}\|\cdot\|\boldsymbol{x}'\|)}{2}(\frac{1+\|\boldsymbol{x}\|}{2})^{1-s}(\frac{1-\|\boldsymbol{x}'\|}{2})^{s} \\
&+ \frac{1+\boldsymbol{x}\cdot\boldsymbol{x}'/(\|\boldsymbol{x}\|\cdot\|\boldsymbol{x}'\|)}{2}(\frac{1-\|\boldsymbol{x}\|}{2})^{1-s}(\frac{1-\|\boldsymbol{x}'\|}{2})^{s} \\
D_{\max}(\rho_{\boldsymbol{x}}\|\rho_{\boldsymbol{x}'}) =& \log\frac{(1-\boldsymbol{x}\cdot\boldsymbol{x}')^2 + \sqrt{\|\boldsymbol{x}-\boldsymbol{x}'\|^2 + \|\boldsymbol{x}\times\boldsymbol{x}'\|^2}}{1-\|\boldsymbol{x}'\|} \\
D_{\min}(\rho_{\boldsymbol{x}}\|\rho_{\boldsymbol{x}'}) =& \begin{cases} -\log\frac{1+\boldsymbol{x}\cdot\boldsymbol{x}'}{2} & \text{when } \|\boldsymbol{x}\| = 1 \\ 0 & \text{when } \|\boldsymbol{x}\| \neq 1, \end{cases}
\end{aligned}$$

where $\boldsymbol{x} \times \boldsymbol{x}'$ is the cross product between $\boldsymbol{x}$ and $\boldsymbol{x}'$.

Further, when the two states are two vector states $|y\rangle$ and $|z\rangle$, we can choose a suitable two-dimensional subspace so that the supports of these two states are included in the subspace. That is, we can choose two qubit states $\rho_{\boldsymbol{x}}$ and $\rho_{\boldsymbol{x}'}$ such that $\|\boldsymbol{x}\| = \|\boldsymbol{x}'\| = 1$ and $|\langle y|z\rangle|^2 = \frac{1+\boldsymbol{x}\cdot\boldsymbol{x}'}{2}$. Hence, for $s \in (0,1)$, we have $e^{\psi(s\||y\rangle\langle y|\||z\rangle\langle z|)} = |\langle y|z\rangle|^2$, $d_1(|y\rangle\langle y|, |z\rangle\langle z|) = \sqrt{1-|\langle y|z\rangle|^2}$, and $D_{\min}(|y\rangle\langle y|, |z\rangle\langle z|) = -\log|\langle z|y\rangle|^2$. On the other hand, when $|y\rangle\langle y| \neq |z\rangle\langle z|$, the relative Renyi entropy, relative entropy with $s < 0$ and the relative max entropy are infinity.

Exercise 2.12 Calculate $H_{1+s}(\rho_{\boldsymbol{x}})$.

2.6 Information Processing Inequalities

This section explains the relation between quantum channel and information quantities given in Sect. 2.4. Section 2.4 introduces several kinds of functions $f(\rho_1, \rho_2)$ to measure the difference between two density matrices ρ_1 and ρ_2 on the system $\mathcal{H}$. Now, we introduce another condition for information quantities as follows.

Information processing inequality: Any quantum channel Λ satisfies as Fig. 2.4

$$f(\rho_1, \rho_2) \geq f(\Lambda(\rho_1), \Lambda(\rho_2)). \tag{2.51}$$

Theorem 2.10 *Trace norm distance $d_1(\rho_1, \rho_2)$ [42, 45], Bures distance $b(\rho_1, \rho_2)$ [42, 45], relative entropy $D(\rho_1\|\rho_2)$ [42, 45, 102], relative Rényi entropy $D_{1-s}(\rho_1\|\rho_2)$ with $1 > s > -1$ [102], relative min-entropy $D_{\min}(\rho_1\|\rho_2)$ [110], relative max-entropy $D_{\max}(\rho_1\|\rho_2)$ [110], and logarithmic inverse square fidelity $F_{\log}(\rho_1, \rho_2)$ [42, 45] satisfy the information processing inequality.*

On the other hand, the pseudo fidelity $\tilde{F}(\rho_1, \rho_2)$ does not necessarily satisfy the information processing inequality. Since any quantum channel has Stinespring representation, the information processing inequality for the partial trace guarantees the information processing inequality (2.51) for any quantum channel.

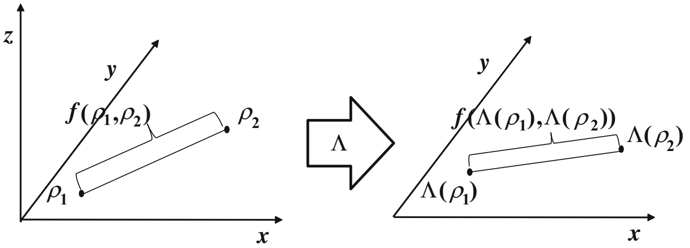


Fig. 2.4 Information processing inequality: This picture shows the case when the projection from 3-dimensional space to 2 dimensional space

When ρ_2 is the completely mixed state ρ_{mix}, the relative entropy, relative Rényi entropy, relative min-entropy, and relative max-entropy are written by using von Neumann entropy, Rényi entropy, max-entropy, and min-entropy as follows

$$D(\rho\|\rho_{\mathrm{mix}}) = \log d - H(\rho), \qquad D_{1-s}(\rho\|\rho_{\mathrm{mix}}) = \log d - H_{1-s}(\rho)$$
$$D_{\min}(\rho\|\rho_{\mathrm{mix}}) = \log d - H_{\max}(\rho), \qquad D_{\max}(\rho\|\rho_{\mathrm{mix}}) = \log d - H_{\min}(\rho),$$

where d is the dimension of the quantum system. The above fact and Theorem 2.10 yield the following theorem.

Corollary 2.1 *When a quantum channel Λ from the system $\mathcal{H}$ to the system $\mathcal{H}$ is unital, the relations*

$$H(\rho) \ge H(\Lambda(\rho)), \qquad H_{1-s}(\rho) \ge H_{1-s}(\Lambda(\rho))$$
$$H_{\max}(\rho) \ge H_{\max}(\Lambda(\rho)), \qquad H_{\min}(\rho) \ge H_{\min}(\Lambda(\rho))$$

hold for $1 > s \ge -1$.

Given a real number $1 > \lambda > 1$ and states ρ_1, ρ_2, ρ_1', and ρ_2' on the system $\mathcal{H}$, we consider the states $\tilde{\rho_1} := \begin{pmatrix} \lambda\rho_1 & 0 \\ 0 & (1-\lambda)\rho_1' \end{pmatrix}$ and $\tilde{\rho_2} := \begin{pmatrix} \lambda\rho_2 & 0 \\ 0 & (1-\lambda)\rho_2' \end{pmatrix}$ on the composite system $\mathcal{H} \otimes \mathbb{C}^2$. We apply Theorem 2.10 to the partial trace with respect to the system $\mathbb{C}^2$. Calculating the respective information quantities with the states $\tilde{\rho_1}$ and $\tilde{\rho_2}$, we obtain the following corollary.

Corollary 2.2 *Trace norm distance, square of Bures distance, relative entropy, relative min-entropy, and $e^{\psi(s|\rho_1\|\rho_2)}$ with $-1 < s < 0$ satisfy joint convexity. Also, $-e^{\psi(s|\rho_1\|\rho_2)}$ with $1 > s > 0$ and $-F(\rho_1, \rho_2)$ satisfies joint convexity.*

Further, applying Corollary 2.2 to the case when ρ_2 and ρ_2' are the completely mixed state ρ_{mix}, we obtain the following corollary.

Corollary 2.3 *von Neumann entropy $H(\rho)$ satisfies concavity. Also $e^{\psi(s|\rho)}$ satisfies convexity for ρ with $s < 0$ and concavity for ρ with $0 < s < 1$.*

2.7 Relative Entropy and Rényi Entropy

In quantum information, given a density matrix ρ and a real number R, the quantities $\min_{H(\rho')\le R} D(\rho'\|\rho)$ and $\min_{H(\rho')\ge R} D(\rho'\|\rho)$ often have operational significance. These quantities can be described by using $\psi(s) := \psi(s|\rho)$. In this treatment, choosing a suitable standard basis, we can assume that $\rho = \rho(\boldsymbol{p})$ without loss of generality. In this case, we often employ the notation $\psi(s|\boldsymbol{p}) := \psi(s|\rho(\boldsymbol{p}))$. Now, we will show that the derivative of $\psi(s)$ is written by relative entropy. Using this fact, we will give formulas to calculate the above quantities based on $\psi(s)$. In the following, without loss of generality, we assume that $\boldsymbol{p} = (p_1, \dots, p_d)$ satisfies $p_j \ge p_{j+1}$ and $p_d > 0$.

First, we focus on the derivative of $\psi(s)$. The relation (2.27) guarantees that the derivative $\psi'(s)$ is monotone increasing. Since

$$\lim_{s\to-\infty}\psi'(s)=\log p_1=H_{\min}(\rho),\quad \psi'(1)=-\sum_{j=1}^{r}\log p_j,$$

we can define $s(S)$ for $S\in(H_{\min}(\rho),\psi'(1)]$ as the inverse function of $\psi'(s)$. That is, we choose $s(S)\in(-\infty,1]$ satisfying

$$S=\psi'(s(S)). \tag{2.52}$$

Further, we choose the integer k such that $p_1=p_k>p_{k+1}$. Then, the relation

$$\begin{aligned}\frac{d}{dS}(1-s(S))S+\psi(s(S))&=1-s(S)-s'(S)S+s'(S)\psi'(s(S))\\&=1-s(S)>0\end{aligned}$$

holds for $S\in(H_{\min}(\rho),\psi'(1)]$. So, $(1-s(S))S+\psi(s(S))$ is monotone increasing for S. Then, we have $(1-1)\psi'(1)+\psi(1)=\psi(1)=\log d$. As will be shown later, the relation

$$\lim_{s\to-\infty}(1-s)\psi'(s)+\psi(s)=\log k \tag{2.53}$$

holds. Hence, for $R\in(\log k,\log d]$, we ca define the inverse function S_R as follows.

$$R=(1-s(S_R))S_R+\psi(s(S_R)). \tag{2.54}$$

Conversely, when a real number R satisfies $R\le\log k$, the relation

$$R\ <-s\psi'(s)+\psi(s). \tag{2.55}$$

holds for any s.
Proof of (2.53). We have

$$\begin{aligned}\psi'(s)(1-s)+\psi(s)&=-(1-s)\sum_{j=1}^{d}\frac{p_j^{1-s}\log p_j}{\sum_{j'=1}^{d}a_{j'}^{1-s}}+\log\sum_{j'=1}^{d}p_{j'}^{1-s}\\
=-(1-s)&\sum_{j=k+1}^{d}\frac{p_j^{1-s}\log p_j}{\sum_{j'=1}^{d}p_{j'}^{1-s}}+\log\sum_{j'=1}^{d}p_{j'}^{1-s}-\log kp_1^{1-s}\\
&+\left(-k(1-s)\frac{p_1^{1-s}\log p_1}{\sum_{j'=1}^{d}p_{j'}^{1-s}}+(1-s)\log p_1\right)+\log k\end{aligned}$$

$$
\begin{aligned}
= & -(1-s)\sum_{j=k+1}^{d}\frac{p_j^{1-s}\log p_j}{\sum_{j'=1}^{d} p_{j'}^{1-s}}+\log\frac{\sum_{j'=1}^{d} p_{j'}^{1-s}}{kp_1^{1-s}} \\
& +(1-s)\frac{\sum_{j=k+1}^{d} p_j^{1-s}}{\sum_{j'=1}^{d} p_{j'}^{1-s}}\log p_1+\log k.
\end{aligned}
$$

For $j > k$, the quantity $\frac{p_j^{1-s}}{\sum_{j'=1}^{d} p_{j'}^{1-s}}$ exponentially approaches to zero as $s \to -\infty$. On the other hand, the quantity $\frac{\sum_{j=1}^{d} p_j^{1-s}}{kp_1^{1-s}}$ converges to 1. Hence, we obtain (2.53). ■

Using the above defined function $s(S)$ and S_R, we can characterize $\min_{H(\rho')=R} D(\rho'\|\rho)$ as follows.

Lemma 2.8 *The relations*

$$
\begin{aligned}
S_R-R=S_R s(S_R)-\psi(s(S_R))= & \frac{s(S_R)R-\psi(s(S_R))}{1-s(S_R)} && (2.56)\\
= & \min_{H(\boldsymbol{q})=R} D(\boldsymbol{p}\|\boldsymbol{p})=\min_{H(\rho')=R} D(\rho'\|\rho) && (2.57)
\end{aligned}
$$

hold for $R \in (\log k, \log d]$. *Also, the relations*

$$
\min_{H(\boldsymbol{q})=R} D(\boldsymbol{q}\|\boldsymbol{p})=\min_{H(\rho')=R} D(\rho'\|\rho)=-\log p_1-R \qquad (2.58)
$$

hold for $R \in [0, \log k]$.

Proof The first equation in (2.56) follows from (2.54). The relation (2.54) also guarantees that $S_R=\frac{R-\psi(s(S_R))}{1-s(S_R)}$. Substituting this equation into S_R-R, we obtain the second equation in (2.56).

In the following, we show (2.57). Choosing $\rho_s := \frac{\rho^{1-s}}{\operatorname{Tr}\rho^{1-s}}$, we have

$$
\begin{aligned}
& D(\rho'\|\rho)-D(\rho_s\|\rho) \\
= & \operatorname{Tr}\rho'(\log\rho'-\log\rho)-\operatorname{Tr}\frac{\rho^{1-s}}{\operatorname{Tr}\rho^{1-s}}\left(\log\frac{\rho^{1-s}}{\operatorname{Tr}\rho^{1-s}}-\log\rho\right) \\
= & \operatorname{Tr}\rho'\left(\log\rho'-\log\frac{\rho^{1-s}}{\operatorname{Tr}\rho^{1-s}}\right)+\operatorname{Tr}\left(\rho'-\frac{\rho^{1-s}}{\operatorname{Tr}\rho^{1-s}}\right)\left(\log\frac{\rho^{1-s}}{\operatorname{Tr}\rho^{1-s}}-\log\rho\right) \\
= & D(\rho'\|\rho_s)-s\operatorname{Tr}\left(\rho'-\frac{\rho^{1-s}}{\operatorname{Tr}\rho^{1-s}}\right)\log\rho, \qquad (2.59)
\end{aligned}
$$

and

$$\begin{aligned}
& - H(\rho') + H(\rho_s) \\
= & \operatorname{Tr} \rho' \left(\log \rho' - \log \frac{\rho^{1-s}}{\operatorname{Tr} \rho^{1-s}} \right) + \operatorname{Tr} \left(\rho' - \frac{\rho^{1-s}}{\operatorname{Tr} \rho^{1-s}} \right) \log \frac{\rho^{1-s}}{\operatorname{Tr} \rho^{1-s}} \\
= & D(\rho' \| \rho_s) + (1-s) \operatorname{Tr} \left(\rho' - \frac{\rho^{1-s}}{\operatorname{Tr} \rho^{1-s}} \right) \log \rho .
\end{aligned} \tag{2.60}$$

Further, the relation (2.54) implies $H(\rho_{s(S_R)}) = R$. When $H(\rho') = R$, the relations (2.59) and (2.60) imply

$$\begin{aligned}
& \frac{D(\rho' \| \rho_{s(S_R)})}{1 - s(S_R)} = - \operatorname{Tr} \left(\rho' - \rho_{s(S_R)} \right) \log \rho \\
= & \frac{D(\rho' \| \rho) - D(\rho_{s(S_R)} \| \rho) - D(\rho' \| \rho_{s(S_R)})}{s(S_R)} .
\end{aligned}$$

That is, we have

$$D(\rho' \| \rho) - D(\rho_{s(S_R)} \| \rho) = \frac{1}{1 - s(S_R)} D(\rho' \| \rho_{s(S_R)}) \ge 0.$$

Hence, we obtain $D(\rho_{s(S_R)} \| \rho) = \min_{H(\rho')=R} D(\rho' \| \rho)$.

Now, we choose $\boldsymbol{q}$ so that $\rho_s = \rho(\boldsymbol{q})$. Hence, identifying $\boldsymbol{q}$ and $\rho(\boldsymbol{q})$, we find that the above quantity equals $\min_{H(\boldsymbol{q})=R} D(\boldsymbol{q} \| \boldsymbol{p})$. Further, the relation (2.54) implies

$$D(\rho_{s(S_R)} \| \rho) = \psi'(s(S_R)) s(S_R) - \psi(s(S_R)) = S_R s(S_R) - \psi(s(S_R)).$$

So, we obtain (2.57).

Finally, we will show (2.58). For $R \in [0, \log k]$, we choose ρ' such that $H(\rho') = R$. Then, we have

$$\begin{aligned}
& D(\rho' \| \rho) = \operatorname{Tr} \rho' \log \rho' + \operatorname{Tr} \rho' (- \log \rho) \\
\ge & - H(\rho') + \operatorname{Tr} \rho' (- \log p_1) = - \log p_1 - R.
\end{aligned}$$

Since the support of ρ' is the subspace generated by $|1\rangle, \ldots, |k\rangle$, we can choose the state ρ' such that $H(\rho') = R$. Such a state ρ' attain the equality in the above inequality. Thus, we obtain (2.58).

Using Lemma 2.8, we can show the following lemma with respect to $\min_{H(\rho') \le R} D(\rho' \| \rho)$ and $\min_{H(\rho') \ge R} D(\rho' \| \rho)$.

Lemma 2.9 *When $H(\rho) < R < \log d$, we have*

$$S_R - R = \min_{H(\boldsymbol{q})\geq R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho')\geq R} D(\rho'\|\rho) \tag{2.61}$$

$$0 = \min_{H(\boldsymbol{q})\leq R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho')\leq R} D(\rho'\|\rho). \tag{2.62}$$

When $\log k < R < H(\rho)$, we have

$$0 = \min_{H(\boldsymbol{q})\geq R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho')\geq R} D(\rho'\|\rho) \tag{2.63}$$

$$S_R - R = \min_{H(\boldsymbol{q})\leq R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho')\leq R} D(\rho'\|\rho). \tag{2.64}$$

When $0 \leq R \leq \log k$,

$$0 = \min_{H(\boldsymbol{q})\geq R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho')\geq R} D(\rho'\|\rho) \tag{2.65}$$

$$\log p_1 - R = \min_{H(\boldsymbol{q})\leq R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho')\leq R} D(\rho'\|\rho). \tag{2.66}$$

Proof Taking the derivative with respect to R in (2.52), we have

$$\frac{dS_R}{dR} = \frac{ds(S_R)}{dR}\psi''(s(S_R)). \tag{2.67}$$

Next, taking the derivative with respect to R in (2.54), we have $1 = -\frac{ds(S_R)}{dR}S_R + (1-s(S_R))\frac{dS_R}{dR}s'(S_R)\psi'(s(S_R))$. So, substituting (2.52) into this equation, we have

$$1 = (1-s(S_R))\frac{dS_R}{dR}. \tag{2.68}$$

Combining (2.67) and (2.68), we obtain

$$\frac{ds(S_R)}{dR} = \frac{1}{(1-s(S_R))\psi''(s(S_R))} < 0. \tag{2.69}$$

Then, the relations (2.68) and (2.69) yield

$$\frac{d}{dR}(S_R - R) = \frac{1}{1-s(S_R)} - 1 = \frac{s(S_R)}{1-s(S_R)}$$

$$\frac{d^2}{dR^2}(S_R - R) = \frac{1}{(1-s(S_R))^3}\psi''(s(S_R)) > 0.$$

Hence, the function $R \mapsto S_R - R$ is a convex function. The equality $\frac{d}{dR}(S_R - R) = 0$ holds only when $s(S_R) = 0$, i.e., $R = H(\rho)$. Since $S_{H(\rho)} - H(\rho) = 0$, this function realizes the minimum value 0 when $R = H(\rho)$. This, using (2.57), we obtain the relations (2.61), (2.62), (2.63), (2.64), and (2.65). Also, applying the same discussion to $\log p_1 - R$, we obtain (2.66). ■

We also have the following lemma with respect to the function ψ.

Lemma 2.10 *When $H(\rho) < R < \log d$, the relations*

$$S_R - R = \max_{0 \le s < 1} \frac{sR - \psi(s)}{1-s}, \quad 0 = \max_{s \le 0} \frac{sR - \psi(s)}{1-s} \tag{2.70}$$

hold. When $\log k < R < H(\rho)$, the relations

$$0 = \max_{0 \le s < 1} \frac{sR - \psi(s)}{1-s}, \quad S_R - R = \max_{s \le 0} \frac{sR - \psi(s)}{1-s} \tag{2.71}$$

hold. When $0 \le R \le \log k$, the relations

$$0 = \max_{0 \le s < 1} \frac{sR - \psi(s)}{1-s}, \quad \log p_1 - R = \sup_{s \le 1} \frac{sR - \psi(s)}{1-s} \tag{2.72}$$

hold.

Proof Taking the derivative with respect to s, we have

$$\frac{d}{ds} \frac{sR - \psi(s)}{1-s} = \frac{R - (1-s)\psi'(s) - \psi(s)}{(1-s)^2} \tag{2.73}$$

$$\frac{d}{ds}(R - (1-s)\psi'(s) - \psi(s)) = -(1-s)\psi''(s) \le 0, \tag{2.74}$$

where the equality in (2.74) holds only when $s = 1$.

When $R \in (\log k, \log d)$, the relations (2.52) and (2.54) imply

$$R - (1 - s(S_R))\psi'(s(S_R)) - \psi(s(S_R)) = 0. \tag{2.75}$$

Hence, we have $\max_{s<1} \frac{sR-\psi(s)}{1-s} = \frac{s(S_R)R+\psi(s(S_R))}{1-s(S_R)}$. Especially, the relation (2.69) in the proof of Lemma 2.9 shows that the function $R \mapsto s(S_R)$ is strictly monotone increasing. Hence, $s(S_R) \le 0$ if and only if $R \le H(\rho)$. Thus, using the relation $\frac{0R-\psi(0)}{1} = 0$, we have

$$\max_{0 \le s < 1} \frac{sR - \psi(s)}{1-s} = \begin{cases} \frac{s(S_R)R+\psi(s(S_R))}{1-s(S_R)} & \text{if } H(\rho) < R < \log d \\ 0 & \text{if } \log k < R \le H(\rho) \end{cases}$$

$$\max_{s \le 0} \frac{sR - \psi(s)}{1-s} = \begin{cases} 0 & \text{if } H(\rho) < R < \log d \\ \frac{s(S_R)R+\psi(s(S_R))}{1-s(S_R)} & \text{if } \log k < R \le H(\rho). \end{cases}$$

Therefore, we obtain the second relation of (2.70) and the second relation of (2.71).

When $R \in [0, \log k]$, the relation (2.55) guarantees that the RHS of (2.73) is a negative number for $s < 1$. Thus,

$$\sup_{s<1} \frac{sR - \psi(s)}{1-s} = \lim_{s \to -\infty} \frac{sR - \psi(s)}{1-s} = -\log p_1 - R.$$

Hence,

$$\max_{0 \le s < 1} \frac{sR - \psi(s)}{1-s} = 0, \quad \sup_{s \le 0} \frac{sR - \psi(s)}{1-s} = -\log p_1 - R,$$

which implies the second relation in (2.72). ■

Summarizing the above lemmas, we obtain the following theorem.

Theorem 2.11

$$\begin{aligned} &\min_{\boldsymbol{q}:H(\boldsymbol{q}) \le R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho') \le R} D(\rho'\|\rho) \\ =&\sup_{s \le 0} \frac{sR - \psi(s)}{1-s} = \sup_{s \le 0} \frac{s}{1-s}(R - H_{1-s}(\rho)), \end{aligned} \tag{2.76}$$

$$\begin{aligned} &\min_{\boldsymbol{q}:H(\boldsymbol{q}) \ge R} D(\boldsymbol{q}\|\boldsymbol{p}) = \min_{H(\rho') \ge R} D(\rho'\|\rho) \\ =&\max_{0 \le s < 1} \frac{sR - \psi(s)}{1-s} = \max_{0 \le s < 1} \frac{s}{1-s}(R - H_{1-s}(\rho)). \end{aligned} \tag{2.77}$$

2.8 Rényi Mutual Information

In Sect. 2.4.3, we discussed the mutual information $I_\rho(A:B)$ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ when ρ is a density matrix on $\mathcal{H}_A \otimes \mathcal{H}_B$. Recall tho two expressions for the mutual information $I_\rho(A:B)$ as

$$I_\rho(A:B) = D(\rho\|\rho_A \otimes \rho_B) = \min_{\sigma_B} D(\rho\|\rho_A \otimes \sigma_B). \tag{2.78}$$

When the reduced density matrices ρ_A and ρ_B are invertible, we have two kinds of Rényi extension of mutual information as

$$I^{\uparrow}_{1+s|\rho}(A:B) := D_{1+s}(\rho\|\rho_A \otimes \rho_B) = \frac{1}{s} \log \operatorname{Tr} \rho^{1+s}(\rho_A^{-s} \otimes \rho_B^{-s}) \tag{2.79}$$

$$I^{\downarrow}_{1+s|\rho}(A:B) := \min_{\sigma_B} D_{1+s}(\rho\|\rho_A \otimes \sigma_B) \tag{2.80}$$

for $s \in (-1, 0) \cup (0, \infty)$. Evidently, we have $I^{\uparrow}_{1+s|\rho}(A:B) \ge I^{\downarrow}_{1+s|\rho}(A:B)$. Also, we find that $\lim_{s \to 0} I^{\uparrow}_{1+s|\rho}(A:B) = I_\rho(A:B)$.

To calculate $I^{\downarrow}_{1+s|\rho}(A:B)$, we consider Rényi version of Pythagorean theorem as a generalization of (2.37).

Lemma 2.11 ([115, Lemma 3 in Suppl. Mat.]) *We have*

$$D_{1+s}(\rho\|\rho_A\otimes\sigma_B)=D_{1+s}(\rho\|\rho_A\otimes\sigma_B^*(1+s))+D_{1+s}(\sigma_B^*(1+s)\|\sigma_B), \quad (2.81)$$

where $\sigma_B^*(1+s):=\frac{(\mathrm{Tr}_A\,\rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}}}{\mathrm{Tr}_B(\mathrm{Tr}_A\,\rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}}}$.

Since $D_{1+s}(\sigma_B^*(1+s)\|\sigma_B)\ge 0$, $\rho_A\otimes\sigma_B^*(1+s)$ is the point closet to ρ in the sense of Rényi relative entropy. So, we have

$$I^{\downarrow}_{1+s|\rho}(A:B)=D_{1+s}(\rho\|\rho_A\otimes\sigma_B^*(1+s))=\frac{1+s}{s}\log\mathrm{Tr}_B(\mathrm{Tr}_A\,\rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}}. \quad (2.82)$$

This expression shows that $\lim_{s\to 0}I^{\downarrow}_{1+s|\rho}(A:B)=I_\rho(A:B)$. Therefore, we can define $I^{\downarrow}_{1+s|\rho}(A:B)$ for $s\in(-1,\infty)$.

Since $s\mapsto sD_{1+s}(\rho\|\rho_A\otimes\sigma_B)$ is convex, Exercise 2.3 guarantees that the map $s\mapsto\max_{\sigma_B}sD_{1+s}(\rho\|\rho_A\otimes\sigma_B)=sI^{\downarrow}_{1+s|\rho}(A:B)$ is convex for $s\in(-1,0)$.

Given a classical-quantum channel $W:x\mapsto W(x)$ on $\mathcal{H}_B$ and a distribution p on $\mathcal{X}:=\{1,\dots,k\}$, we consider the state $\rho:=\sum_{x=1}^k p(x)|x\rangle\langle x|\otimes W(x)$. Then, we have

$$tI^{\downarrow}_{1-t|\rho}(A:B)=\phi_{W,\boldsymbol{p}}(t):=-(1-t)\log\mathrm{Tr}\left(\sum_{x=1}^k p(x)W(x)^{1-t}\right)^{\frac{1}{1-t}} \quad (2.83)$$

for $t\in(0,1)$. Then, we find that $\phi_{W,\boldsymbol{p}}(t)$ is a concave function. This function will be used for a universal code for classical-quantum channel in Sect. 6.7.

Proof of Lemma 2.11 Since $(\mathrm{Tr}_B(\mathrm{Tr}_A\,\rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}})^{1+s}=e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B^*(1+s))}$, we have

$$\begin{aligned}e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B)}&=\mathrm{Tr}\,\rho^{1+s}\rho_A^{-s}\sigma_B^{-s}=\mathrm{Tr}_B(\mathrm{Tr}_A\,\rho^{1+s}\rho_A^{-s})^{\frac{1+s}{1+s}}\sigma_B^{-s}\\&=e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B^*(1+s))}\mathrm{Tr}_B(\frac{(\mathrm{Tr}_A\,\rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}}}{\mathrm{Tr}_B(\mathrm{Tr}_A\,\rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}}})^{1+s}\sigma_B^{-s}\\&=e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B^*(1+s))}e^{sD_{1+s}(\sigma_B^*(1+s)\|\sigma_B)}.\end{aligned}$$

■

Exercise 2.13 Give another proof of (2.82) by using matrix Hölder inequality (2.21) and matrix reverse Hölder inequality (2.22).

Chapter 3
Entanglement and Its Quantification

Abstract Quantum mechanics is completely far from everyday intuition not only because the measured outcome can be predicted only probabilistically but also because of a quantum-specific correlation called entanglement. This type of correlation never appears in macroscopic objects. As a typical protocol with entanglement, this chapter explains quantum teleportation. Then, it explains several measures to quantify entanglement. While existing books mainly deal with entanglement of the bipartite system, this chapter deals with entanglement of the multi-partite system as well as entanglement of the bipartite system. This topic seems to have no deep relation with group representation. However, this topic is essentially based on group representation in an unexpected way. Although many entanglement measures are proposed, it is quite difficult to calculate them. A larger part of resolved cases are essentially based on the group symmetry because many entanglement measures are based on a optimization and employing the group symmetry can reduce the freedom of the optimization.

3.1 Locality Conditions

When the composite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ is composed of several distant systems, it is usually difficult to perform a quantum operation across both systems. Similarly, it is often difficult to generate a quantum state across the whole of the composite system. On the other hand, when an operation is composed of quantum operations on the respective systems and classical communications, it does not have such difficulty because it does not need such a quantum operation across distant systems. Hence, such an operation does not increase the amount of entanglement. In the following, such an operation is called Local Operations and Classical Communications, and is abbreviated to an **LOCC**.

When the composite system is composed of two distant systems, a maximally entangled state is the most fundamental state among states across two distant systems. For example, as mentioned in Sect. 3.2, an operation to transmit a quantum system between two distant systems can be realized by a combination of a maximally entangled state and an LOCC. As mentioned in Sect. 3.6, any entangled state between

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8_3

two systems can be realized by a combination of a maximally entangled state and an LOCC. A maximally entangled state and an LOCC are fundamental tools to understand entangled states in this sense.

An LOCC is called a **one-way LOCC** when the classical communication is restricted only to one-direction communications among these plural distant systems. A one-way LOCC is concretely described as follows. Firstly, we apply a measurement described by an instrument $\{\Lambda^1_{\omega_1}\}$ on the first system $\mathcal{H}_1$. Then, we send the first measurement outcome ω_1 to the other systems. In the second system $\mathcal{H}_2$, we choose the instrument $\{\Lambda^{2,\omega_1}_{\omega_2}\}$ dependently of the first measurement outcome ω_1, and apply it to obtain the second measurement outcome ω_2. Then, we send it to the other systems. In the j-the system $\mathcal{H}_j$, we choose the instrument $\{\Lambda^{j,\omega_1,\ldots,\omega_{j-1}}_{\omega_j}\}$ dependently of the measurement outcomes $\omega_1, \ldots, \omega_{j-1}$ and apply it to obtain the j-th measurement outcome ω_j. Then, we send it to the other systems. This LOCC is characterized by the following TP-CP map:

$$\sum_{\omega_1,\omega_2,\ldots,\omega_n} \Lambda^1_{\omega_1} \otimes \Lambda^{2,\omega_1}_{\omega_2} \otimes \cdots \otimes \Lambda^{n,\omega_1,\ldots,\omega_{n-1}}_{\omega_n}. \tag{3.1}$$

On the other hand, an LOCC is called a **two-way LOCC** when bidirectional classical communications are allowed among these plural distant systems. For example, when the composite system is composed of two distant systems and we apply measurement on each system alternately with interactive communication for measurement outcomes, the whole operation is a two-way LOCC. In general, an LOCC means a two-way LOCC.

As a larger class of quantum operations, we have the class of separable quantum channels. A quantum channel Λ from the composite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ to the composite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ is called **separable** when Λ has the following Kraus representation

$$\Lambda(\rho) = \sum_m F^1_m \otimes F^2_m \otimes \cdots \otimes F^n_m \rho (F^1_m \otimes F^2_m \otimes \cdots \otimes F^n_m)^\dagger. \tag{3.2}$$

A one-way LOCC is a two-way LOCC, and a two-way LOCC is a separable operation (quantum channel). The relation among these relations are summarized as Fig. 3.1.

Such a restriction for quantum channels based on the structure of tensor product on the quantum system is called a **locality condition**. When a POVM can be realized by one-way LOCC, two-way LOCC, or separable, it is called a **one-way LOCC POVM**, a **two-way LOCC POVM**, or a **separable POVM**, respectively. Especially, a POVM $M = \{M_j\}_j$ is a separable POVM if and only if its entries M_j can be written as

$$M_j = \sum_l M^{1,l}_j \otimes \cdots \otimes M^{r,l}_k, \tag{3.3}$$

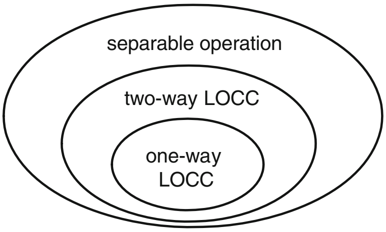


Fig. 3.1 Relation among local operations

where $M_j^{m,l}$ is a positive semi definite matrix on the system $\mathcal{H}_m$. Now, we recall the majorization relation $\succ$ between two density matrices ρ_1 and ρ_2 defined in Sect. 2.3.2.

Theorem 3.1 (Nielsen [100]) *Given two vector states $|\Phi\rangle$ and $|\Psi\rangle$ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$, we define the two states $\rho_1 := \mathrm{Tr}_B\, |\Phi\rangle\langle\Phi|$ and $\rho_2 := \mathrm{Tr}_B\, |\Psi\rangle\langle\Psi|$. Then, there exists a two-way LOCC that converts $|\Phi\rangle$ to $|\Psi\rangle$ if and only if $\rho_2 \succ \rho_1$.*

Exercise 3.1 Show that any vector state $|\Psi\rangle$ on a given bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$ from a maximally entangled state $|\Phi\rangle$ on the bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$ by using Theorem 3.1.

3.2 Quantum Teleportation

As a typical protocol that efficiently uses an entangle state, this section explains quantum teleportation. Given an irreducible projective unitary representation f of a group G on a finite-dimensional system $\mathcal{H}_A$ and a maximally entangled state $\frac{1}{\sqrt{d}}|I\rangle\rangle_{B,C}$ across sender's system $\mathcal{H}_B$ and receiver's system $\mathcal{H}_C$ that have the same dimension d as $\mathcal{H}_A$, there is a protocol that can transmit a quantum state by sending a classical information and applying local operations. Such a protocol is called **quantum teleportation** [6], and is given as follows.

Firstly, the sender prepares the state $|x\rangle = \sum_j x_j |j\rangle$ to be sent on the system $\mathcal{H}_A$. Then, the sender applies the measurement for the POVM $d \int_G |\mathsf{f}(g)\rangle\rangle_{A,B\ A,B}\langle\langle \mathsf{f}(g)| \mu_G(dg)$ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$. Then, the sender sends the measurement outcome g to the receiver. Finally, the receiver apply the unitary $\mathsf{f}(g)$ dependently of the received measurement outcome g, as Fig. 3.2.

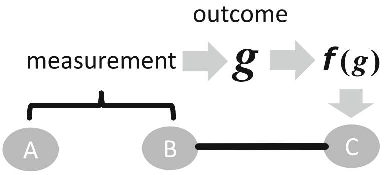


Fig. 3.2 Quantum teleportation

Since

$$
{}_{A,B}\langle\!\langle \mathsf{f}(g) | | x\rangle_A \frac{1}{\sqrt{d}} |I\rangle\!\rangle_{B,C} = \frac{1}{\sqrt{d}} \sum_{j,k,l} \overline{\mathsf{f}(g)_{j,k}} x_j \delta_{k,l} |l\rangle_C
$$

$$
= \frac{1}{\sqrt{d}} \sum_l (\overline{\mathsf{f}(g)}^T x)_l |l\rangle_C ,
$$

the final state on the receiver's system $\mathcal{H}_C$ is $\mathsf{f}(g) \sum_l (\overline{\mathsf{f}(g)}^T x)_l |l\rangle_C = \sum_l x_l |l\rangle_C$. This discussion shows that this protocol enables us to transmit a quantum state $|x\rangle$ only with a LOCC.

Usual teleportation protocol is this protocol when the irreducible projective unitary representation f is discrete Heisenberg representation. In this case, the required size of classical information to be sent is the square of the dimension d of the system $\mathcal{H}_A$ to be transmitted. The square of the dimension of the representation space of the irreducible projective unitary representation f is equal or smaller than the order $|G|$ of the group G. (For its proof, see [44, (2.56)].) The case of the discrete Heisenberg representation realizes the minimum amount of classical information to be sent because it attains the equality condition in the above evaluation for the amount of classical information to be sent.

Exercise 3.2 We consider three distant players Alice, Bob, and Charlie. Then, we assume that Alice and Bob have the systems $\mathcal{H}_A$ and $\mathcal{H}_B$, whose composite system is in the maximally entangled state. Also, we assume that Bob and Charlie have the systems $\mathcal{H}_{B'}$ and $\mathcal{H}_C$, whose composite system is in the maximally entangled state. In this case, Alice and Charlie do not share any entangled stated. Give an LOCC protocol to generate a maximally entangled state between Alice and Charlie by using quantum teleportation. This protocol is called **entanglement swapping**.

3.3 Examples of Entangled States

3.3.1 Discrete Heisenberg Representation Revisited

Since we give examples of entangled states by using discrete Heisenberg representation [44, Chap. 8], we firstly revisit discrete Heisenberg representation. In this section, we address stabilizer code as a typical class of quantum error correction. For this purpose, we simply review discrete Heisenberg representation. For its detail, see [44, Chap. 8].

First of all, we review discrete Heisenberg representation of the algebra $\mathbb{Z}_d$, which is a d-dimensional irreducible projective unitary representation as follows. We define two matrices $\mathsf{X}_{\mathbb{Z}}$ and $\mathsf{Z}_{\mathbb{Z}}$ on the system with the computational basis $\{|0\rangle, \ldots, |d-1\rangle\}$ as follows.

$$\mathsf{X}_{\mathbb{Z}}|x\rangle = |x+1\rangle, \quad \mathsf{Z}_{\mathbb{Z}}|x\rangle = \omega_{\mathbb{Z}}^{x}|x\rangle, \tag{3.4}$$

where $\omega_{\mathbb{Z}} := e^{i2\pi/d}$. This definition is equivalent to the following definition.

$$\mathsf{X}_{\mathbb{Z}} = \sum_x |x+1\rangle\langle x|, \quad \mathsf{Z}_{\mathbb{Z}} = \sum_x \omega_{\mathbb{Z}}^{x}|x\rangle\langle x|. \tag{3.5}$$

Then, we define W as

$$\tilde{\mathsf{W}}_{\mathbb{Z}}(s,t) := \mathsf{X}^s\mathsf{Z}^t = \sum_s \omega_{\mathbb{Z}}^{xt}|x+s\rangle\langle x|, \quad \forall(s,t) \in \mathbb{Z}_d^2. \tag{3.6}$$

Since $\omega_{\mathbb{Z}}\mathsf{X}\mathsf{Z} = \mathsf{Z}\mathsf{X}$, the relation

$$\tilde{\mathsf{W}}_{\mathbb{Z}}(s,t)\tilde{\mathsf{W}}_{\mathbb{Z}}(s',t') = \omega_{\mathbb{Z}}^{s't}\tilde{\mathsf{W}}_{\mathbb{Z}}(s+s',t+t') \tag{3.7}$$

holds. So, $\tilde{\mathsf{W}}_{\mathbb{Z}}$ is a projective unitary representation of $\mathbb{Z}_d^2$.

When d is odd, we set $\tau_{\mathbb{Z}} := \omega_{\mathbb{Z}}^{(d+1)/2}$, and when d is even, we set $\tau_{\mathbb{Z}} := e^{i\pi/d}$. Then, we define the projective unitary representation $\mathsf{W}_{\mathbb{Z}}(s,t)$ as

$$\mathsf{W}_{\mathbb{Z}}(s,t) := \tau_{\mathbb{Z}}^{st}\tilde{\mathsf{W}}_{\mathbb{Z}}(s,t) = \tau_{\mathbb{Z}}^{st}\sum_s \omega_{\mathbb{Z}}^{xt}|x+s\rangle\langle x|, \tag{3.8}$$

where st takes values in integers between 0 and $d-1$. Then, we have

$$\mathsf{W}_{\mathbb{Z}}(s,t)\mathsf{W}_{\mathbb{Z}}(s',t') = \begin{cases} \tau_{\mathbb{Z}}^{s't-t's}\mathsf{W}_{\mathbb{Z}}(s+s',t+t') & \text{when } d \text{ is odd} \\ (-1)^{\epsilon}\tau_{\mathbb{Z}}^{s't-t's}\mathsf{W}_{\mathbb{Z}}(s+s',t+t') & \text{when } d \text{ is even,} \end{cases} \tag{3.9}$$

where ϵ is 0 or 1 dependently of s, t, s', t'. $\mathsf{W}_{\mathbb{Z}}(s,t)$ is called **discrete Heisenberg representation** of $\mathbb{Z}_d^2$.

Here, given the computational basis $\{|x\rangle\}_{x\in\mathbb{Z}_d}$, the **dual computational basis** is defined as

$$|\hat{e}_{\mathbb{Z}}(l)\rangle := \frac{1}{\sqrt{d}}\sum_{k\in\mathbb{Z}_d}\omega_{\mathbb{Z}}^{-kl}|k\rangle, \quad l \in \mathbb{Z}_d. \tag{3.10}$$

Then, the dual computational basis are the eigenvectors of $\mathsf{X}_{\mathbb{Z}}$, and $\mathsf{Z}_{\mathbb{Z}}$ permutes the dual computational basis.

Now, we define the discrete Heisenberg group $\mathrm{H}(2, \mathbb{Z}_d)$ by using $\omega_d := e^{\frac{2\pi i}{d}}$ as follows. When d is odd, $\mathrm{H}(2, \mathbb{Z}_d)$ is the set $\{(s, t, \omega_d^n)|s, t \in \mathbb{Z}_d, n \in \mathbb{Z}\}$ with multiplication $(s, t, \omega_d^n)(s', t', \omega_d^{n'}) = (s + s', t + t', \omega_d^{n+n'+st'-s't})$. When d is even, $\mathrm{H}(2, \mathbb{Z}_d)$ is the set $\{(s, t, \omega_{2d}^n)|s, t \in \mathbb{Z}_d, n \in \mathbb{Z}\}$ with multiplication $(s, t, \omega_{2d}^n)(s', t', \omega_{2d}^{n'}) = (s + s', t + t', \omega_{2d}^{n+n'+st'-s't})$. Then, the representation $\mathsf{W}_{\mathbb{Z},\mathrm{H}}$ of the discrete Heisenberg group $\mathrm{H}(2, \mathbb{Z}_d)$ is defined as

$$\mathsf{W}_{\mathbb{Z},\mathrm{H}}(s, t, \omega_d^n) := \tau_{\mathbb{Z}}^n \mathsf{W}_{\mathbb{Z}}(s, t) \quad \text{when } d \text{ is odd}$$
$$\mathsf{W}_{\mathbb{Z},\mathrm{H}}(s, t, \omega_{2d}^n) := \tau_{\mathbb{Z}}^n \mathsf{W}_{\mathbb{Z}}(s, t) \quad \text{when } d \text{ is even.}$$

The fact that $\mathsf{W}_{\mathbb{Z},\mathrm{H}}$ satisfies the condition of representation can be checked by the isometric relation between the central extension of $\mathbb{Z}_d^2$ based on the factor system given in [44, (8.7)] and the discrete Heisenberg group $\mathrm{H}(2, \mathbb{Z}_d)$ defined above.

Next, we consider the case with Galois extension $\mathbb{F}_q$ of the finite field $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$. Identifying the vector space $\mathbb{F}_p^m$ with $\mathbb{F}_q$, we can describe the linear map $x \mapsto zx$ by an $m \times m$ matrix M_z on $\mathbb{F}_p$. Then, we denote $\mathrm{Tr}\, M_z$ by trz. Using $\omega_{\mathbb{F}} := e^{i2\pi/p}$, we define the matrices $\mathsf{X}_{\mathbb{F}}$ and $\mathsf{Z}_{\mathbb{F}}$ on the space $\mathcal{H}$ whose CONS is given by the computational basis $\{|x\rangle\}_{x \in \mathbb{F}_q}$ as

$$\mathsf{X}_{\mathbb{F}}(s)|x\rangle = |x + s\rangle, \quad \mathsf{Z}_{\mathbb{F}}(t)|x\rangle = \omega_{\mathbb{F}}^{trtx}|x\rangle. \tag{3.11}$$

This definition is equivalent to the following definition.

$$\mathsf{X}_{\mathbb{F}}(s) = \sum_{x \in \mathbb{F}_q} |x + s\rangle\langle x|, \quad \mathsf{Z}_{\mathbb{F}}(t) = \sum_{x \in \mathbb{F}_q} \omega_{\mathbb{Z}}^{trtx}|x\rangle\langle x|.$$

Then, we define the representation $\tilde{\mathsf{W}}_{\mathbb{F}}(s, t) := \mathsf{X}_{\mathbb{F}}(s)\mathsf{Z}_{\mathbb{F}}(t) = \sum_{x \in \mathbb{F}_q} \omega_{\mathbb{Z}}^{trtx}|x + s\rangle\langle x|$ for $\forall (s, t) \in \mathbb{F}_q^2$. Since we have the commutation relation

$$\mathsf{X}_{\mathbb{F}}(s)\mathsf{Z}_{\mathbb{F}}(t)\mathsf{X}_{\mathbb{F}}(-s) = \omega_{\mathbb{F}}^{-trst}\mathsf{Z}_{\mathbb{F}}(t), \tag{3.12}$$

$\tilde{\mathsf{W}}_{\mathbb{F}}$ satisfies $\tilde{\mathsf{W}}_{\mathbb{F}}(s, t)\tilde{\mathsf{W}}_{\mathbb{F}}(s', t') = \omega_{\mathbb{F}}^{trs't}\tilde{\mathsf{W}}_{\mathbb{F}}(s + s', t + t')$. Hence, $\tilde{\mathsf{W}}_{\mathbb{F}}$ is a projective unitary representation of $\mathbb{F}_q^2$.

We define $\tau_{\mathbb{F}} := \omega_{\mathbb{Z}}^{(p+1)/2}$ when the character p of $\mathbb{F}_q$ is not 2, and we define $\tau_{\mathbb{F}} := i$ otherwise. Then, we define the projective representation $\mathsf{W}_{\mathbb{F}}(s, t)$ as

$$\mathsf{W}_{\mathbb{F}}(s, t) := \tau_{\mathbb{F}}^{trts}\tilde{\mathsf{W}}_{\mathbb{F}}(s, t) = \tau_{\mathbb{F}}^{trts} \sum_{x \in \mathbb{F}_q} \omega_{\mathbb{F}}^{trtx}|x + s\rangle\langle x|,$$

where $trts$ takes values in an integer between 0 and $p-1$. Then, we have

$$\mathsf{W}_{\mathbb{F}}(s,t)\mathsf{W}_{\mathbb{F}}(s',t') = \begin{cases} \tau_{\mathbb{F}}^{tr(s't-t's)}\mathsf{W}_{\mathbb{F}}(s+s',t+t') & \text{when } p \neq 2 \\ (-1)^{\epsilon}\tau_{\mathbb{F}}^{tr(s't-t's)}\mathsf{W}_{\mathbb{F}}(s+s',t+t') & \text{when } p = 2, \end{cases} \tag{3.13}$$

where ϵ is 0 or 1 dependently of s, t, s', t'. The projective unitary representation $\mathsf{W}_{\mathbb{F}}(s,t)$ of $\mathbb{F}_q^2$ is called the **discrete Heisenberg representation** of $\mathbb{F}_q^2$.

When q is a prime p, by identifying $\mathbb{F}_p^2$ with $\mathbb{Z}_p^2$, the discrete Heisenberg representation of $\mathbb{F}_p^2$ equals the discrete Heisenberg representation $\mathbb{Z}_p^2$. However, when q is a power p^m of a prime p, the discrete Heisenberg representation of $\mathbb{F}_q^2$ does not equal the discrete Heisenberg representation $\mathbb{Z}_q^2$. Now, based on the computational basis $\{|x\rangle\}_{x\in\mathbb{F}_q}$, we define the **dual computational basis** as

$$|\hat{e}_{\mathbb{F}}(l)\rangle := \frac{1}{\sqrt{q}}\sum_{k\in\mathbb{F}_q}\omega_{\mathbb{F}}^{-trkl}|k\rangle, \quad l \in \mathbb{F}_q. \tag{3.14}$$

Then, dual computational basis are eigenvectors of $\mathsf{X}_{\mathbb{F}}$, and are permutated by $\mathsf{Z}_{\mathbb{F}}$.

Next, we define the following projective unitary representation of the group $\mathbb{Z}_d^{2r}$ or the vector space $\mathbb{F}_q^{2r}$ on $\mathcal{H}^{\otimes r} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_r$ from the discrete Heisenberg representation of $\mathbb{Z}_d^2$ or $\mathbb{F}_q^2$, where $\mathcal{H}_i$ is a space isomorphic to $\mathcal{H}$. The computational basis of $\mathcal{H}^{\otimes r}$ is written as $|s\rangle$ by using elements s of $\mathbb{Z}_d^r$ or $\mathbb{F}_q^r$. In the following, we employ an unified notation, i.e., denote $\mathbb{F}_q$ or $\mathbb{Z}_d$ by $\mathbb{X}$. Then, we let $\mathsf{W}_{\mathbb{X},i}(s,t)$ be the operation $\mathsf{W}_{\mathbb{X}}(s,t)$ on the i-th space $\mathcal{H}_i$. Then, we define

$$\mathsf{W}_{\mathbb{X}}^r(\vec{s}) := \mathsf{W}_{\mathbb{X},1}(s_1,t_1)\otimes\cdots\otimes\mathsf{W}_{\mathbb{X},r}(s_r,t_r), \quad \forall\vec{s} = (s_1,\ldots s_r,t_1,\ldots,t_r) \in \mathbb{X}^{2r}.$$

For $\boldsymbol{s} = (s_1,\ldots s_r)$, $\boldsymbol{t} = (t_1,\ldots t_r) \in \mathbb{X}^r$, we define

$$\begin{aligned} (\boldsymbol{s},\boldsymbol{t})_{\mathbb{Z}} &:= s_1t_1 + \cdots + s_rt_r, \\ (\boldsymbol{s},\boldsymbol{t})_{\mathbb{F}} &:= tr(s_1t_1 + \cdots + s_rt_r), \end{aligned} \tag{3.15}$$

and for $\vec{s} = (\boldsymbol{s},\boldsymbol{t})$, $\vec{s}' = (\boldsymbol{s}',\boldsymbol{t}') \in \mathbb{X}^{2r}$, we define $\langle\vec{s},\vec{s}'\rangle_{\mathbb{X}} := (\boldsymbol{s}',\boldsymbol{t})_{\mathbb{X}} - (\boldsymbol{s},\boldsymbol{t}')_{\mathbb{X}}$. So, we have

$$\mathsf{W}_{\mathbb{X}}^r(\vec{s})\mathsf{W}_{\mathbb{X}}^r(\vec{s}') = \begin{cases} \tau_{\mathbb{X}}^{\langle\vec{s},\vec{s}'\rangle_{\mathbb{X}}}\mathsf{W}_{\mathbb{X}}^r(\vec{s}+\vec{s}') & \text{when } d \text{ or } p \text{ is odd} \\ (-1)^{\epsilon}\tau_{\mathbb{X}}^{\langle\vec{s},\vec{s}'\rangle_{\mathbb{X}}}\mathsf{W}_{\mathbb{X}}^r(\vec{s}+\vec{s}') & \text{when } d \text{ or } p \text{ is even}, \end{cases} \tag{3.16}$$

where ϵ is 0 or 1 dependently of s, t, s', t'. Hence, $\mathsf{W}_{\mathbb{Z}}^r$ and $\mathsf{W}_{\mathbb{F}}^r$ are projective unitary representations of $\mathbb{Z}_d^{2r}$ and $\mathbb{F}_q^{2r}$, respectively. The projective unitary representations $\mathsf{W}_{\mathbb{Z}}^r$ and $\mathsf{W}_{\mathbb{F}}^r$ are called **discrete Heisenberg representation** of $\mathbb{Z}_d^{2r}$ and $\mathbb{F}_q^{2r}$, respectively.

3.3.2 Stabilizer State

As explained in Sect. 1.2, a vector state on the composite system $\mathcal{H}_1 \otimes \mathcal{H}_2$ (**bipartite system**) of two systems is described as $|X\rangle\rangle$ by using a matrix X. In general, it is thought that the amount of entanglement is not changed by the applications of unitaries on individual systems, which is called a **local unitary**. Hence, in entanglement theory, when two entangled states are converted to each other by local unitary, the amounts of their entanglement are considered to be equal to each other. So, we say that these two states are called equivalent to each other as entangled states. Thus, the amount of the entanglement can be characterized by the Schmidt coefficient of the matrix X given in (1.8). Especially, when the Schmidt rank is d and all of the Schmidt coefficients are one fixed value, the vector state $|X\rangle\rangle$ is a maximally entangled state with rank d. When both dimensions of $\mathcal{H}_1$ and $\mathcal{H}_2$ are d, the vector state $|X\rangle\rangle$ is a maximally entangled of rank d if and only if $\sqrt{d}X$ is a unitary matrix.

On the other hand, it is not so easy to characterize the amount of entanglement when the state is mixed. Given discrete Heisenberg representation $\mathsf{W}^r_{\mathbb{X}}$ of $\mathbb{X}^{2r}$ on the system $\mathcal{H}_1 = \mathcal{H}_2$, the density matrix $\sum_{\vec{s}\in\mathbb{X}^{2r}} \mathrm{P}(\vec{s})|\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rangle\rangle\langle\langle\mathsf{W}^r_{\mathbb{X}}(\vec{s})|$ with a probability distribution P on $\mathbb{X}^{2r}$ is called a **Bell diagonal state**. Such a state can be easily treated in comparison with a general mixed state.

However, when the whole system is composed of more than three systems, the multipartite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_r$ is not so easy to treat the amount of entanglement in the whole system. The most famous vector state among entangled states on the multipartite system is a stabilizer state. To define a stabilizer state, we consider the case when the respective system $\mathcal{H}_j$ has dimension $q = p^n$ and the discrete Heisenberg representation $\mathsf{W}^r_{\mathbb{F}_q}$ of $\mathbb{F}_q^{2r}$ is defined on the composite system $(\mathbb{C}^q)^{\otimes r} := \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_r$. Then, the action of the subgroup $\mathrm{Sp}(2, \mathbb{F}_q)^r$ of discrete symplectic group $\mathrm{Sp}(2r, \mathbb{F}_q)^r$ under Metaplectic representation $\mathsf{S}^r_{\mathbb{F}_q}$ given in [44, Sect. 8.3.2] and the action of $\mathbb{F}_q^{2r}$ under the discrete Heisenberg representation $\mathsf{W}^r_{\mathbb{F}_q}$ are local unitaries. We call a subgroup N of $\mathbb{F}_q^{2r}$ a strictly self-orthogonal subgroup when $N = \{\vec{s} \in \mathbb{F}_q^{2r} | \langle \vec{s}, \vec{s}'\rangle_{\mathbb{F}} = 0, \ \forall \vec{s}' \in N\}$. Now, we assume that a subgroup N of $\mathbb{F}_q^{2r}$ is strictly self-orthogonal. For any elements $\vec{s}, \vec{s}' \in N$, $\mathsf{W}^r_{\mathbb{F}_q}(\vec{s})$ and $\mathsf{W}^r_{\mathbb{F}_q}(\vec{s}')$ are commutative with each other. Hence, their simultaneous eigenvectors exist. That is, for $x \in N^*$, there exists a vector $|v_x\rangle \in (\mathbb{C}^q)^{\otimes r}$ such that

$$\mathsf{W}^r_{\mathbb{F}_q}(\vec{s})|v^N_x\rangle = \omega_{\mathbb{F}}^{x(\vec{s})}|v^N_x\rangle, \quad \forall \vec{s} \in N. \tag{3.17}$$

Such a vector state is called the **stabilizer state** of N. Since for any two elements $x, x' \in N^*$, the vectors $|v_x\rangle$ and $|v_{x'}\rangle$ are mapped to each other via the action of $\mathbb{F}_q^{2r}$, they are equivalent to each other as entangled state. That is, the amount of entanglement of the stabilizer state does not depend on the choice of an element of $x \in N^*$. When for two strictly self-orthogonal subgroups N and N', there exists an element g of the subgroup $\mathrm{Sp}(2, \mathbb{F}_q)^r$ of the symplectic group $\mathrm{Sp}(2r, \mathbb{F}_q)^r$ such that $N = gN'$, the stabilizer states of N and N' are equivalent to each other as entangled states.

In fact, a strictly self-orthogonal subgroup N can be written as a generating matrix, which is a $2r \times r$ matrix (see [44, Sect. 8.2.3]). Further, when the upper half of a generating matrix A of N is an invertible matrix $g \in \mathrm{GL}(r, \mathbb{F}_q)$, Ag^{-1} is also a generating matrix of N. Hence, when a strictly self-orthogonal subgroup N has a generating matrix satisfying the above condition, N has a generating matrix $\begin{pmatrix} I \\ \zeta \end{pmatrix}$ with use of a symmetric matrix ζ. Given a symmetric matrix ζ and a vector $\boldsymbol{s} \in \mathbb{F}_q^r$, there exists a stabilizer state $|u_{\boldsymbol{s}}^{\zeta}\rangle$ of the strictly self-orthogonal subgroup with generating matrix $\begin{pmatrix} I \\ \zeta \end{pmatrix}$ such that

$$\mathsf{W}_{\mathbb{F}_q}^r((\boldsymbol{t}, \zeta\boldsymbol{t})^T)|u_{\boldsymbol{s}}^{\zeta}\rangle = \omega_{\mathbb{F}}^{\boldsymbol{s}\cdot\boldsymbol{t}}|u_{\boldsymbol{s}}^{\zeta}\rangle, \quad \boldsymbol{t} \in \mathbb{F}_q^r. \tag{3.18}$$

For example, when $\zeta = 0$, the subgroup N is $\{\mathsf{X}_{\mathbb{F}}(t_1) \otimes \cdots \otimes \mathsf{X}_{\mathbb{F}}(t_r)\}_{(t_1,\ldots,t_r)}$. Hence, the stabilizer state is given as (Exercise 3.3)

$$|u_{\boldsymbol{s}}^{0}\rangle = \mathsf{Z}_{\mathbb{F}}(s_1) \otimes \cdots \otimes \mathsf{Z}_{\mathbb{F}}(s_r)\frac{1}{q^{r/2}}\sum_{\boldsymbol{x}\in\mathbb{F}_q^r}|\boldsymbol{x}\rangle. \tag{3.19}$$

Further, we can restrict stabilizer states as discussed in the following lemma.

Lemma 3.1 *Given a strictly self-orthogonal subgroup N and an element $x \in N^*$, there exist a symmetric matrix ζ and a vector $\boldsymbol{s} \in \mathbb{F}_q^r$ such that $|u_{\boldsymbol{s}}^{\zeta}\rangle$ is equivalent to $|v_x^N\rangle$ as an entangled state and all of diagonal elements of ζ are zero.*

Before the proof, we prepare three elements M_b, Q_a, and P_a of $\mathrm{Sp}(2, \mathbb{F}_q)$ for $a, b \in \mathbb{F}_q$ as

$$M_b := \begin{pmatrix} b^{-1} & 0 \\ 0 & b \end{pmatrix}, \quad Q_a := \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}, \quad P_a := \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}. \tag{3.20}$$

Proof Even though the upper half of a generating matrix A of N is not invertible, there exists an element $h \in \mathrm{GL}(r, \mathbb{F}_q)$ such that the upper half of Ah is diagonal and all of its diagonal elements are 1 or 0. Since Ah is a generating matrix of a strictly self-orthogonal subgroup, when the (j, j) entry of the upper half of Ah is zero, the (j, j) entry a_j of the lower half of Ah is not zero. Hence, the (j, j) entry of the upper half of $J_{r,j}Ah$ is a_j, where

$$J_{r,j} := -|j\rangle\langle j+r| + |j+r\rangle\langle j| + \sum_{k\neq j}(|k\rangle\langle k| + |k+r\rangle\langle k+r|). \tag{3.21}$$

Repeating this procedure, we obtain a generating matrix A of N such that the upper half g of A is invertible and diagonal. Hence, the upper half of another generating matrix Ag^{-1} of N is the unit matrix. Thus, there exists a symmetric matrix ζ and a vector $\boldsymbol{s} \in \mathbb{F}_q^r$ such that $|u_{\boldsymbol{s}}^{\zeta}\rangle$ is equivalent to $|v_x^N\rangle$ as an entangled state. Denote the j-th diagonal element of ζ by z_j. Then, we have $Q_{z_1} \oplus \cdots \oplus Q_{z_r} \in \mathrm{Sp}(2, \mathbb{F}_q)^r$. So, the upper half of $(Q_{z_1} \oplus \cdots \oplus Q_{z_r}) \begin{pmatrix} I \\ \zeta \end{pmatrix}$ is the unit matrix. Its lower half is a symmetric matrix whose diagonal elements are zero. ■

Exercise 3.3 Show (3.19).

3.3.3 Graph State

In the following, we treat the case of $\mathbb{F}_q = \mathbb{F}_2$, in which, all matrix entries are 0 or 1. Now, based on the above lemma, we choose the $r \times r$ symmetric ζ such that its diagonal elements are zero. When its (k, j) entry is 1, we connect two vertexes k and j. Then, we obtain a **graph**. In the following, we identify a graph among r vertexes and a binary $r \times r$ symmetric matrix ζ with zero diagonal entries. Hence, the stabilizer state $|u_s^\zeta\rangle$ is called a **graph state** of a graph ζ [65, 66].

Given a graph ζ and vertexes j, k, we define the element $g_{j,k}$ of $\mathrm{Sp}(2, \mathbb{F}_2)$ as

$$g_{j,k} := \begin{cases} Q_1 & \text{when } k = j \\ P_1 & \text{when } \zeta_{k,j} = 1 \\ I & \text{otherwise.} \end{cases} \tag{3.22}$$

Then, we define the unitary matrix K_j as

$$K_j := \mathsf{S}_{\mathbb{F}_2,j}(g_j), \quad g_j := g_{j,1} \oplus \cdots \oplus g_{j,r}, \tag{3.23}$$

where $\mathsf{S}_{\mathbb{F}_2,k}$ is a subgroup $\mathsf{S}_{\mathbb{F}_2}$ acting on the k-th tensor product component. Then, K_j is a local unitary.

Now, given a vertex j, we define the **local complementation** at j as the conversion of the graph ζ to the graph $\hat{\zeta}$ defined by

$$\hat{\zeta}_{k,l} := \begin{cases} \zeta_{k,l} & \text{when } \zeta_{j,k}\zeta_{j,l} = 0 \\ 1 - \zeta_{k,l} & \text{when } \zeta_{j,k}\zeta_{j,l} = 1. \end{cases} \tag{3.24}$$

Lemma 3.2 *The state $K_j|u_s^\zeta\rangle$ is a graph state of the conversion of the graph ζ by the local complementation at the vertex j. That is, when two graphs are converted by local complementation, their graph states are equivalent to each other.*

Proof It is sufficient to show that the vector space generated by the matrix $g_j \begin{pmatrix} I \\ \zeta \end{pmatrix}$ is the same as the vector space generated by the matrix $\begin{pmatrix} I \\ \hat{\zeta} \end{pmatrix}$. Consider the case when $r = 3$. When $g_1 = Q_1 \oplus P_1 \oplus P_1$,

$$\zeta = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \hat{\zeta} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

with a simple calculation, we find that the vector space spanned by $\begin{pmatrix} I \\ \zeta \end{pmatrix}$ is converted to the vector space spanned by $\begin{pmatrix} I \\ \hat{\zeta} \end{pmatrix}$ by the action g_1. Hence, we obtain the desired statement with $r = 3$. Further, we can show the general case in the same way as the case of $r = 3$. ■

Now, we denote the operation to flip the (j, k) and (k, j) entries of the graph by $\mathcal{T}_{j,k}$. Here, the "flip" means the changes of entries as $1 \to 0$ or $0 \to 1$. Also, we define the matrix $\gamma_{j,l:1} := |j)(l| + \sum_{k=1}^{r} |k)(k| \in \mathrm{GL}(r, \mathbb{F}_2)$, where $\{|k)\}_{k=1}^{r}$ is a basis of vector space $\mathbb{F}_2^r$ over the finite field $\mathbb{F}_2$. The unitary $\mathsf{U}_{j,l} := \sum_{\boldsymbol{x} \in \mathbb{F}_2^r} |\gamma_{j,l:1}\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ is C-NOT operation when the control system is the l-th qubit and the target system is the j-th qubit.[1]

Then, we obtain the following lemma.

Lemma 3.3 *The unitary matrix $\mathsf{U}_{j,l}$ maps the graph state of the graph ζ to the graph state of the graph $\mathcal{T}_{j,l}(\zeta)$.*

Proof Since the proof with the case of $r = 2$ can be easily generalized to the general case, we show it only in the case of $r = 2$. When $\zeta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\hat{\zeta} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, the unitary $\mathsf{U}_{j,l}$ converts the vector space spanned by $\begin{pmatrix} I \\ \zeta \end{pmatrix}$ to the vector space spanned by $\begin{pmatrix} I \\ \hat{\zeta} \end{pmatrix}$. Then, we obtain the desired argument. ■

Lemma 3.3 gives the method to construct a graph state for a given graph ζ as follows. Firstly, we prepare the state $(\mathsf{Z}_{\mathbb{F}}(s_1) \otimes \cdots \otimes \mathsf{Z}_{\mathbb{F}}(s_r)) \frac{1}{q^{r/2}} \sum_{\boldsymbol{x} \in \mathbb{F}_2^r} |bx\rangle$. Then, we apply the unitary $\prod_{(j,l)} \mathsf{U}_{j,l}$, where (j, l) is the pair of connected vertexes in the given graph ζ. Then, we obtain the graph state $|u_s^{\zeta}\rangle$.

Example 3.1 (*1D Cluster state*) As $\bullet - \bullet - \cdots - \bullet - \bullet$, the graph state of the graph composed of one-dimensional lattice is called a **1D Cluster state**.

Example 3.2 (*Ring state*) When the vertexes are plotted in the circle and only adjacent vertexes are connected, the graph state of the graph is called a **Ring state**.

Example 3.3 (*GHZ state*) When all vertexes are connected, the graph is called the **complete graph**. On the other hand, when one vertex j is connected to all of other vertexes and there is no other connection, the graph is called the **star graph** with

[1] As explained in [44, Sect. 8.3.2], the unitary $\mathsf{U}_{j,l}$ is given as $\mathsf{S}_{\mathbb{F}_2}^r(M_{\gamma_{j,l:1}})$ in terms of Metaplectic representation $\mathsf{S}_{\mathbb{F}_2}^r$.

center j. The complete graph and the star graph with center j are converted by local complementation at j. Especially, the graph state of the star graph of center j is called the **GHZ state**.

Example 3.4 (*2D Cluster state*) When the vertexes are plotted in the two-dimensional lattice, and adjacent vertexes are connected only in the horizontal and vertical directions, the graph of the graph is called the **2D Cluster state**.

Exercise 3.4 Give the concrete form of GHZ state with r vertexes with $\boldsymbol{s} = 0$.

Exercise 3.5 Give the concrete form of 1D Cluster state with $r = 3$ and $\boldsymbol{s} = 0$.

Exercise 3.6 Give the concrete form of Ring state with $r = 3$ and $\boldsymbol{s} = 0$.

3.4 Characterization on Bipartite System

Separable states have several known useful properties. Consider the map τ_A mapping the state ρ to the state ρ^T on the system $\mathcal{H}_A$. This map is a linear positive map, but is not quantum channel because the extended map $\tau_A \otimes id_B$ (**partial transpose** with respect to $\mathcal{H}_A$) on the bipartite composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ is not a positive map. That is, the image of a state with respect to the map $\tau_A \otimes id_B$ is not a state. However, when the state ρ is a separable state on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$, the relation $\tau_A \otimes id_B(\rho) \geq 0$ holds, which implies that $\tau_A \otimes id_B(\rho)$ is a state. The condition $\tau_A \otimes id_B(\rho) \geq 0$ is called the (**PPT**) condition, and a state ρ satisfying this condition is called a **PPT state**. This condition is equivalent to the condition with respect to the partial trace with respect to $\mathcal{H}_B$ because $\tau_A \otimes id_B(\rho) \geq 0$ if and only if $id_A \otimes \tau_B(\rho) = (\tau_A \otimes id_B(\rho))^T \geq 0$. This condition seems depend on the choice of the basis of $\mathcal{H}_A$, however, it does not depend on this choice because of the following reason. Consider the equivalent condition $id_A \otimes \tau_B(\rho) \geq 0$, which does not depend on the choice of the basis of $\mathcal{H}_A$. Since this discussion can be applied to the basis $\mathcal{H}_B$, the PPT condition does not depend on the choice of both bases.

As another property of a separable state ρ, we introduce the impossibility of generation of maximally entangled state. That is, when the n-foldtensor product state $\rho^{\otimes n}$ is prepared, whatever LOCC Λ_n cannot convert it to the maximally entangled state on $(\mathbb{C}^2)^{\otimes 2}$ even approximately. That is, the resultant state $\Lambda_n(\rho^{\otimes n})$ cannot approximate the maximally entangled state on $(\mathbb{C}^2)^{\otimes 2}$. This property is called **indistilable condition**.

In fact, separable states have other properties. Including the above properties, we can summarize them in the following theorem.

Theorem 3.2 *For a state ρ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$, we have the relations* **(1) ⇒ (2) ⇒ (3) ⇒ (4) ⇒ (5)** *for the following conditions as Fig. 3.3.*

(1) Separable condition*: ρ is a separable state.*
(2) PPT condition*: $\tau_A \otimes id_B(\rho) \geq 0$.*

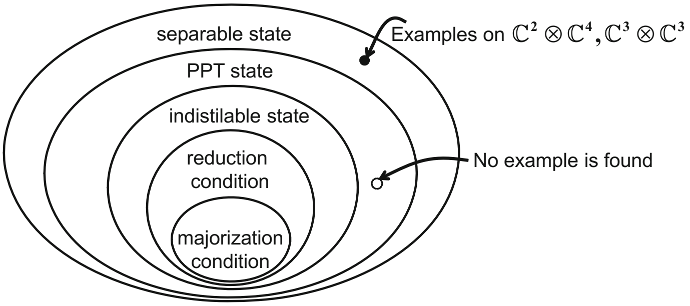


Fig. 3.3 Relation among states on composite system

(3) **Indistilable condition**: *There is no sequence of LOCC Λ_n such that $\Lambda_n(\rho^{\otimes n})$ converges to a maximally entangled state on $(\mathbb{C}^2)^{\otimes 2}$.*

(4) **Reduction condition**: $(\mathrm{Tr}_B\,\rho) \otimes I_B \geq \rho$.

(5) **Majorization condition**: $(\mathrm{Tr}_B\,\rho) \succ \rho$.

The relation **(1)** ⇒ **(2)** can be shown by calculation. The relation **(2)** ⇒ **(3)** will be shown in Sect. 3.5.3 later. The relations **(3)** ⇒ **(4)** and **(4)** ⇒ **(5)** were shown in Horodecki [74], Hiroshima [69]. It is shown in [75] that the conditions **(1)** and **(2)** are equivalent when the composite system is $\mathbb{C}^2 \otimes \mathbb{C}^2$ or $\mathbb{C}^2 \otimes \mathbb{C}^3$. However, when the composite system is $\mathbb{C}^2 \otimes \mathbb{C}^4$ or $\mathbb{C}^3 \otimes \mathbb{C}^3$, there exists a state such that the condition **(2)** holds but the condition **(1)** does not hold [73]. Such a state is called a **bound entangled state**.

3.5 Quantification of Entanglement I: Geometrical Method

3.5.1 Entanglement Monotone

The first method to quantify the amount of entanglement of a state ρ on the multipartite composite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ is to measure how different it is from the set of states with no entanglement (separable states). For this purpose, we employ information measures for two information sources given in Sect. 2.4. Such a quantification is expected to be non-increasing under the operation satisfying the locality condition. This property is called **entanglement monotone**. For example, when it is not increasing under LOCC, it is called LOCC entanglement monotone. When it is not increasing under separable operation, it is called separable entanglement monotone.

Let SEP be the set of separable states on the composite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$. Using a function f defined for two states, as in Fig. 3.4, we define the entanglement measure $E[f]_{1,\ldots,n}(\rho) := \min_{\rho_1 \in \mathrm{SEP}} f(\rho, \rho_1)$ as a quantification for the entanglement of the state ρ. When there is no possibility for confusion without expression of the

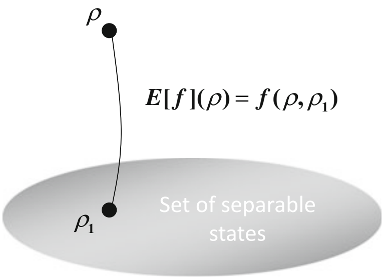


Fig. 3.4 Closet point and $E[f](\rho)$

systems, we simplify it to $E[f](\rho)$. When the relative entropy $D(\rho\|\rho_1)$ is used for the function f, $E[f](\rho)$ is denoted by $E_R(\rho)$, and is called **entanglement relative entropy** [76, 123]. When relative Rényi entropy $D_{1-s}(\rho\|\rho_1)$, logarithmic inverse square fidelity, and logarithmic inverse pseudo fidelity are used for the function f, $E[f](\rho)$ is denoted by $E_{R,s}(\rho)$, $E_{G,1}(\rho)$, and $E_{G,2}(\rho)$. When ρ is a pure state, $E_{G,1}(\rho)$ is the same as $E_{G,2}(\rho)$, and is called **Geometric measure** [116, 127].

Then, we have the following lemma.

Lemma 3.4 *When f satisfies the information processing inequality, $E[f]$ is separable entanglement monotone.*

So, since $D_{1-s}(\rho\|\rho_1)$, $D(\rho\|\rho_1)$, and logarithmic inverse square fidelity satisfy the information processing inequality, $E_R(\rho)$, $E_{R,1-s}(\rho)$, and $E_{G,1}(\rho)$ are separable entanglement monotone. On the other hand, it is unclear whether $E_{G,2}(\rho)$ is separable entanglement monotone. So, its meaning is not clear. However, it can be simplified as

$$E_{G,2}(\rho) = -\log \max_{|a\rangle\langle a|\in\mathrm{SEP}} \langle a|\rho|a\rangle. \tag{3.25}$$

Hence, it plays an important role for calculating other entanglement measures.

Proof Let Λ be a separable quantum channel. When ρ_1 is a separable state on the input system of Λ, $\Lambda(\rho_1)$ is a separable state on the output system of Λ. When ρ is a state on the input system of Λ, the information processing inequality yields that

$$\min_{\rho_1\in\mathrm{SEP}} f(\rho,\rho_1) \ge \min_{\rho_1\in\mathrm{SEP}} f(\Lambda(\rho),\Lambda(\rho_1)) \ge \min_{\rho_1'\in\mathrm{SEP}} f(\Lambda(\rho),\rho_1'),$$

which implies the desired argument. ■

Also, we have the following lemma.

Lemma 3.5 *When f satisfies joint convexity, $E[f](\rho)$ is a convex function.*

Hence, Corollary 2.2 guarantees that $E_R(\rho)$ is a convex function.

Proof Assume that $\rho = \lambda \rho_a + (1-\lambda)\rho_b$, $\rho_{a,1} := \operatorname{argmin}_{\rho_1 \in \mathrm{SEP}} f(\rho_a, \rho_1)$, and $\rho_{b,1} := \operatorname{argmin}_{\rho_1 \in \mathrm{SEP}} f(\rho_b, \rho_1)$. Then, we have

$$\begin{aligned} E[f](\rho) = & \min_{\rho_1 \in \mathrm{SEP}} f(\rho, \rho_1) \le f(\lambda\rho_a + (1-\lambda)\rho_b, \lambda\rho_{a,1} + (1-\lambda)\rho_{b,1}) \\ \le & \lambda f(\rho_a, \rho_{a,1}) + (1-\lambda) f(\rho_b, \rho_{b,1}) = \lambda E[f](\rho_1) + (1-\lambda)E[f](\rho_2), \end{aligned}$$

which implies the desired argument. ■

As other geometric methods, we introduce **logarithmic robustness** $R_L(\rho)$ [124] and **global logarithmic robustness** $R_{L,g}(\rho)$ [38], which are defined as

$$\begin{aligned} & R_{L:1,\ldots,n}(\rho) \\ := & \min\{\log(1+t) | \text{There exists } \rho' \in \mathrm{SEP} \text{ such that } \frac{\rho + t\rho'}{1+t} \in \mathrm{SEP}.\}, \\ & R_{L,g:1,\ldots,n}(\rho) \\ := & \min\{\log(1+t) | \text{There exists } \rho' \text{ such that } \frac{\rho + t\rho'}{1+t} \in \mathrm{SEP}.\}. \end{aligned}$$

When there is no possibility for confusion, we omit the subscript: $1, \ldots, n$. It follows from the definitions that $R_L(\rho) \ge R_{L,g}(\rho)$.

All of above given quantities are defined as the minimization with the separability condition. Hence, if the ranges for these minimization become larger, these quantities become smaller. The separability in the above definitions can be easily extended to the separability among n quantum systems. Alternatively, dividing the n systems into a subset $S \subset \{1, \ldots, n\}$ and its complement S^c, we define the separability with respect to two systems $\otimes_{j \in S} \mathcal{H}_j$ and $\otimes_{j \in S^c} \mathcal{H}_j$. So, the set of separable states with respect to the latter sense is larger than the set of separable states with respect to the former sense.

Using this fact, we find the following inequality [90].

$$E[f]_{1,\ldots,n}(\rho) \ge E[f]_{S,S^c}(\rho). \tag{3.26}$$

Since these entangled measures in the bipartite case have been studied better than in the multipartite case, as mentioned later, this inequality is very powerful tool for calculating these quantity in the multipartite case. A similar relation holds for $R_L(\rho)$, $R_{L,g}(\rho)$. Further, these measures are invariant with respect to local unitary. Then, we have the following lemma [58].

Lemma 3.6 *When* $1 > s_2 > 0$ *and* $0 > s_1 \ge -1$, *we have*

$$R_{L,g}(\rho) \ge E_{R,s_1}(\rho) \ge E_R(\rho) \ge E_{R,s_2}(\rho) \tag{3.27}$$

$$E_R(\rho) \ge E_{G,2}(\rho) - H(\rho), \quad E_{R,1/2}(\rho) \ge E_{G,1}(\rho). \tag{3.28}$$

Especially, when ρ is a constant times of a projection P, i.e., $\frac{P}{\operatorname{Tr} P}$, we have

$$E_{R,s_2}(\rho) \geq E_{G,2}(\rho) - \log \operatorname{Tr} P, \quad E_{G,1}(\rho) \geq E_{G,2}(\rho) - \log \operatorname{Tr} P. \tag{3.29}$$

From the above lemma, combining (3.28) and (3.29), we have a useful formula

$$R_{L,g}(\rho) \geq E_{G,1}(\rho). \tag{3.30}$$

Further, when ρ is a pure state, we have

$$R_{L,g}(\rho) \geq E_{G,2}(\rho). \tag{3.31}$$

In this case, when $E_{G,2}(\rho) = R_{L,g}(\rho)$, we have

$$E_{G,2}(\rho) = R_{L,g}(\rho) = E_{G,1}(\rho) = E_R(\rho) = E_{R,s}(\rho) \tag{3.32}$$

with $s \in (1, 0) \cup (0, -1)$.

Proof Firstly, we will show the first inequality (3.27). We choose a state ρ' and a real number $t \geq 0$ such that $\frac{\rho+t\rho'}{1+t} \in$ SEP and $\log(1+t) = R_{L,g}(\rho)$. Since $x \mapsto -x^{s_1}$ is matrix monotone, we have $\frac{\rho+t\rho'}{1+t} \leq (\frac{\rho}{1+t})^{s_1}$. Hence,

$$\operatorname{Tr} \rho^{1-s_1} (\frac{\rho + t\rho'}{1+t})^{s_1} \leq \operatorname{Tr} \rho^{1-s_1} (\frac{\rho}{1+t})^{s_1} = (1+t)^{-s_1}.$$

Thus,

$$\begin{aligned} E_{R,s_1}(\rho) \leq \frac{-1}{s_1} \log \operatorname{Tr} \rho^{1-s_1} (\frac{\rho + t\rho'}{1+t})^{s_1} \leq \frac{-1}{s_1} \log \frac{1}{(1+t)^{s_1}} \\ = \log(1+t) = R_{L,g}(\rho), \end{aligned}$$

which implies the first inequality of (3.27).

Next, we will show the first inequality of (3.28). Let ρ_1 be a separable state satisfying $E_R(\rho) = D(\rho \| \rho_1)$. Since the function $x \mapsto -\log x$ is convex, the relation (2.13) yields

$$\begin{aligned} E_R(\rho) = D(\rho \| \rho_1) = \operatorname{Tr} \rho \log \rho - \operatorname{Tr} \rho \log \rho_1 \\ \geq \operatorname{Tr} \rho \log \rho - \log \operatorname{Tr} \rho\rho_1 \geq E_{G,2}(\rho) - H(\rho), \end{aligned}$$

which implies the first inequality of (3.28).

The second and third inequalities of (3.27) follow from the second and third inequalities of (2.31), respectively. Also, the second inequality of (3.27) follows from the second inequality of (2.30).

Next, we will show the first inequality of (3.29). Let ρ_1 be a separable state. Since the function $x \mapsto x^{s_2}$ is concave, we have

$$(\mathrm{Tr}(\frac{P}{\mathrm{Tr}\,P})^{1-s_2}\rho_1^{s_2})^{\frac{1}{s_2}} = (\mathrm{Tr}\,P)(\mathrm{Tr}(\frac{P}{\mathrm{Tr}\,P})\rho_1^{s_2})^{\frac{1}{s_2}} \le (\mathrm{Tr}\,P)\,\mathrm{Tr}\,\frac{P}{\mathrm{Tr}\,P}\rho_1.$$

Taking the logarithm, we obtain the first inequality of (3.29).

Finally, we will show the second inequality of (3.29). Let ρ_1 be a separable state. Schwarz inequality $(\mathrm{Tr}\,XY)^2 \le \mathrm{Tr}\,X^2\,\mathrm{Tr}\,Y^2$ for two Hermitian matrices X and Y yields that

$$\begin{aligned}(\mathrm{Tr}\,|\sqrt{\frac{P}{\mathrm{Tr}\,P}}\sqrt{\rho_1^s}|)^2 &= \frac{1}{\mathrm{Tr}\,P}(\mathrm{Tr}\,|P\sqrt{\rho_1}|)^2 = \frac{1}{\mathrm{Tr}\,P}(\mathrm{Tr}\,\sqrt{P\rho_1 P})^2 \\ &= \frac{1}{\mathrm{Tr}\,P}(\mathrm{Tr}\,P\sqrt{P\rho_1 P})^2 \le \frac{1}{\mathrm{Tr}\,P}\mathrm{Tr}\,P\,\mathrm{Tr}\,P\rho_1 P = \mathrm{Tr}\,P\,\mathrm{Tr}\,\frac{P}{\mathrm{Tr}\,P}\rho_1.\end{aligned}$$

Taking the logarithm, we obtain the second inequality of (3.29). ■

Exercise 3.7 Show that $E_{G,2}(|\Phi\rangle\langle\Phi|) = -\log\max_i p_i$ when $|\Phi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$ by using (3.25).

Exercise 3.8 Calculate $E_{G,2}$ of the GHZ state with r vertexes with $\boldsymbol{s} = 0$ by using the concrete form obtained in Exercise 3.4.

Exercise 3.9 Calculate $E_{G,2}$ of 1D Cluster state with $r = 3$ and $\boldsymbol{s} = 0$ by using the concrete form obtained in Exercise 3.5.

Exercise 3.10 Calculate $E_{G,2}$ of Ring state with $r = 3$ and $\boldsymbol{s} = 0$ by using the concrete form obtained in Exercise 3.6.

3.5.2 Analysis with Examples

In the following, we calculate these entanglement measures in several examples. Firstly, we list several important properties. Then, using them, we calculate these measures. When ρ is a constant times of a projection P, i.e., $\frac{P}{\mathrm{Tr}\,P}$, we consider the conditions **GM1** and **GM4** for a projective unitary representation f of a compact Lie group G, and the conditions **GM2**, **GM3**, and **GM3**$'$ for f and a separable vector state $|a\rangle$.

- **GM1** For any element $g \in G$, the unitary $\mathsf{f}(g)$ is given as a tensor product of unitary matrices on individual systems.
- **GM2** $\langle a|\rho|a\rangle = \max_{|a'\rangle\langle a'|\in\mathrm{SEP}}\langle a'|\rho|a'\rangle$.
- **GM3** $\int_G \mathsf{f}(g)|a\rangle\langle a|\mathsf{f}(g)^\dagger\mu(dg) \ge \langle a|\rho|a\rangle P$.
- **GM3**$'$ There exists a subspace $\mathcal{H}'$ including the ranges of the projections P and $|a\rangle\langle a|$ such that no other irreducible subspace of $\mathcal{H}'$ is equivalent to the range of P with respect to the representation f.
- **GM4** The range of the projection P is a one-dimensional irreducible subspace of the representation f, and no other irreducible subspace of $\mathcal{H}'$ is equivalent to the range of P with respect to the representation f.

Theorem 3.3 ([59]) *Assume that a state ρ is written as $\frac{P}{\mathrm{Tr}\, P}$ by using a projection P and that there exist a projective unitary representation f of a compact Lie group G and a separable vector state $|a\rangle$ such that the conditions* **GM1** *and* **GM3** *hold. We have*

$$-\log\langle a|P|a\rangle \geq R_{L,g}(\rho). \tag{3.33}$$

Proof The condition **GM3** guarantees the existence of the state ρ' satisfying

$$\langle a|\rho|a\rangle \,\mathrm{Tr}\, P \frac{P}{\mathrm{Tr}\, P} + (1 - \langle a|\rho|a\rangle \,\mathrm{Tr}\, P)\rho' = \int_G \mathsf{f}(g)|a\rangle\langle a|\mathsf{f}(g)^\dagger \mu(dg) \in \mathrm{SEP}.$$

Hence, choosing the real number t satisfying that $\frac{1}{1+t} = \langle a|\rho|a\rangle \,\mathrm{Tr}\, P$, we obtain $-\log\langle a|P|a\rangle = \log(1+t) \geq R_{L,g}(\rho)$, which implies the desired argument. ■

Theorem 3.4 ([59]) *We assume that a state ρ is written as $\frac{P}{\mathrm{Tr}\, P}$ by using the projection P and that there exist a projective unitary representation f of a compact Lie group G and a separable vector state $|a\rangle$ such that the conditions* **GM1–GM3** *hold. For any real number $s \in (0, 1)$, the relations*

$$R_{L,g}(\rho) = E_R(\rho) = E_{R,s}(\rho) = E_{G,1}(\rho) = E_{G,2}(\rho) - \log \mathrm{Tr}\, P \tag{3.34}$$

hold.

Proof Lemma 3.6 shows that $E_{G,2}(\rho) - \log \mathrm{Tr}\, P \leq R_{L,g}(\rho)$. On the other hand, the condition **GM2** and Theorem 3.3 yield the opposite inequality. So, we obtain the desired argument. ■

Lemma 3.7 *We assume that a state ρ is written as $\frac{P}{\mathrm{Tr}\, P}$ by using the projection P, a projective unitary representation f of a compact Lie group G satisfies the condition* **GM1**, *and a separable vector state $|a\rangle$ satisfies the conditions* **GM2** *and* **GM3**′. *Then, the separable vector state $|a\rangle$ satisfies the condition* **GM3**.

Hence, when f satisfies the conditions **GM1** *and* **GM4**, *a separable vector state satisfying the condition* **GM2** *satisfies the condition* **GM3**. *So, under the same condition for f, a separable vector state satisfying the condition* **GM2** *or* **GM3** *satisfies (3.34).*

Next, we discuss the tensor product state of plural entangled states. Given a state ρ_1 on $\mathcal{H}_{1,1} \otimes \cdots \otimes \mathcal{H}_{n,1}$ and a state ρ_2 on $\mathcal{H}_{1,2} \otimes \cdots \otimes \mathcal{H}_{n,2}$, we define $E[f](\rho_1 \otimes \rho_2)$, $R_L(\rho_1 \otimes \rho_2)$, and $R_{L,g}(\rho_1 \otimes \rho_2)$ for entanglement of the tensor product state $\rho_1 \otimes \rho_2$ among the n-partite system $\mathcal{H}_{1,1} \otimes \mathcal{H}_{1,2}, \ldots, \mathcal{H}_{n,1} \otimes \mathcal{H}_{n,2}$. When f satisfies the additivity, we have $E[f](\rho_1 \otimes \rho_2) \leq E[f](\rho_1) + E[f](\rho_2)$. Similarly, we have $R_{L,g}(\rho_1 \otimes \rho_2) \leq R_{L,g}(\rho_1) + R_{L,g}(\rho_2)$. Then, as a normalized entanglement measure of ρ, we define $E[f]_\infty(\rho) := \lim_{n\to\infty} \frac{E[f](\rho^{\otimes n})}{n}$. Similarly, we define $R_{L,\infty}(\rho)$ and $R_{L,g,\infty}(\rho)$. For their calculations, we prepare the following lemma.

Lemma 3.8 *We assume that two states $\rho_1 := \frac{P_1}{\mathrm{Tr}\, P_1}$ and $\rho_2 := \frac{P_2}{\mathrm{Tr}\, P_2}$ defined by projections P_1 and P_2 satisfy (3.34). When $E_{G,2}(\rho_1 \otimes \rho_2) = E_{G,2}(\rho_1) + E_{G,2}(\rho_2)$, we have*

$$
\begin{aligned}
&R_{L,g}(\rho_1 \otimes \rho_2) = E_R(\rho_1 \otimes \rho_2) = E_{R,s}(\rho_1 \otimes \rho_2) = E_{G,1}(\rho_1 \otimes \rho_2) \\
=&E_{G,2}(\rho_1) + E_{G,2}(\rho_2) - \log \mathrm{Tr}\, P_1 - \log \mathrm{Tr}\, P_2, \quad 1 > \forall s > 0.
\end{aligned}
$$

Proof Since Lemma 3.6 yields

$$
\begin{aligned}
&R_{L,g}(\rho_1 \otimes \rho_2) \geq E_{G,2}(\rho_1) + E_{G,2}(\rho_2) - \log \mathrm{Tr}\, P_1 - \log \mathrm{Tr}\, P_2 \\
=&R_{L,g}(\rho_1) + R_{L,g}(\rho_2) \geq R_{L,g}(\rho_1 \otimes \rho_2),
\end{aligned}
$$

we have $R_{L,g}(\rho_1 \otimes \rho_2) = E_{G,2}(\rho_1) + E_{G,2}(\rho_2) - \log \mathrm{Tr}\, P_1 - \log \mathrm{Tr}\, P_2$. We can show other values. ■

Lemma 3.9 *Given states $\rho_j := \frac{P_j}{\mathrm{Tr}\, P_j}$ with projections P_j ($j = 1, \ldots n$), we assume that there exist projective unitary representations $\mathsf{f}_1, \ldots, \mathsf{f}_n$ of groups $G_1, \ldots, G_n$ such that the conditions* **GM1** *and* **GM4** *hold. Then, for $1 > s > 0$, we have*

$$
\begin{aligned}
&R_{L,g}(\otimes_{j=1}^n \rho_j) = E_R(\otimes_{j=1}^n \rho_j) = E_{R,s}(\otimes_{j=1}^n \rho_j) = E_{G,1}(\otimes_{j=1}^n \rho_j) \\
=&E_{G,2}(\otimes_{j=1}^n \rho_j) - \sum_{j=1}^n \log \mathrm{Tr}\, P_j.
\end{aligned}
$$

Proof The projective unitary representation $\mathsf{f}_1 \overline{\otimes} \cdots \overline{\otimes} \mathsf{f}_n$ of the direct group $G_1 \times \cdots \times G_n$ satisfies the condition **GM1**. This representation and the projection $P_1 \otimes \cdots \otimes P_n$ satisfy the condition **GM4**. Hence, Lemma 3.7 guarantees the desired argument. ■

The following theorems are useful for the calculation of $E_{G,2}(\rho)$.

Theorem 3.5 ([59, 77]) *We assume that $\mathcal{H}_j = \mathcal{H}$ and a vector state $|\phi\rangle$ on $\mathcal{H}^{\otimes n}$ is invariant with respect to the permutation $\sigma \in S_n$ for the order of tensor. Then, we have*

$$
\max_{|a\rangle\langle a| \in \mathrm{SEP}} |\langle a|\phi\rangle|^2 = \max_{|b\rangle \in \mathcal{H}: \|b\|=1} |\langle b^{\otimes n}|\phi\rangle|^2. \tag{3.35}
$$

Theorem 3.6 ([132]) *Given a basis $\{|e_{j,k}\rangle\}_k$ of the system $\mathcal{H}_j$, we focus on the basis $\{|e_{1,k_1}\rangle \otimes \cdots \otimes |e_{n,k_n}\rangle\}_{k_1,\ldots,k_n}$ of the composite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$. When a density matrix ρ on the composite system is composed of non-negative entries on this basis, any state ρ' satisfies $E_{G,2}(\rho \otimes \rho') = E_{G,2}(\rho) + E_{G,2}(\rho')$.*

Example 3.5 (*Maximally entangled state*) The maximally entangled state $\frac{1}{\sqrt{r}}|I\rangle\rangle$ on the bipartite system $\mathcal{H}_1 \otimes \mathcal{H}_2$ with $\mathcal{H}_1 = \mathcal{H}_2 = \mathbb{C}^r$ satisfies $E_{G,2}(\frac{1}{\sqrt{r}}|I\rangle\rangle) = \log r$. The tensor product representation of fundamental representation and its complex conjugate representation of $\mathrm{U}(r)$ satisfy the condition **GM1** and **GM4**. Hence, Lemma 3.7 guarantees (3.34). Further, Lemmas 3.6 and 3.8 imply

$$\begin{aligned} R_{L,g,\infty}(\frac{1}{\sqrt{r}}|I\rangle\rangle) &= E_{R,\infty}(\frac{1}{\sqrt{r}}|I\rangle\rangle) = E_{R,s,\infty}(\frac{1}{\sqrt{r}}|I\rangle\rangle) = E_{G,1,\infty}(\frac{1}{\sqrt{r}}|I\rangle\rangle) \\ &= E_{G,2,\infty}(\frac{1}{\sqrt{r}}|I\rangle\rangle) = E_{G,2}(\frac{1}{\sqrt{r}}|I\rangle\rangle) = \log r, \quad 1 > \forall s > 0. \end{aligned} \tag{3.36}$$

Example 3.6 (*Pure state on bipartite system*) Given a vector state $|A\rangle\rangle$ on the bipartite system $\mathcal{H}_1 \otimes \mathcal{H}_2$, let $\sum_{j=1}^k \sqrt{d_j}|v_j\rangle \otimes |u_j\rangle$ be its Schmidt decomposition. Assume that $d_1 \geq d_2 \geq \cdots \geq d_k$. Let $\rho := \sum_{j=1}^k d_j^2 |v_j\rangle\langle v_j|$ be the reduced density matrix on the system $\mathcal{H}_1$. Then, we obtain $\max_{\rho_1 \in \mathrm{SEP}} \operatorname{Tr} |A\rangle\rangle\langle\langle A|\rho_1 = d_1$. Hence, we have $E_{G,1}(|A\rangle\rangle) = E_{G,2}(|A\rangle\rangle) = -\log d_1$. Also, as will be shown later, the relations

$$\min_{\rho_1 \in \mathrm{SEP}} \operatorname{Tr} |A\rangle\rangle\langle\langle A|^{1-s} \rho_1^s = (\operatorname{Tr} \rho^{\frac{1}{1-s}})^{1-s}, \quad 0 > s > -1 \tag{3.37}$$

$$\max_{\rho_1 \in \mathrm{SEP}} \operatorname{Tr} |A\rangle\rangle\langle\langle A|^{1-s} \rho_1^s = (\operatorname{Tr} \rho^{\frac{1}{1-s}})^{1-s}, \quad 1 > s > 0 \tag{3.38}$$

hold. Thus, we have

$$E_{R,1-s}(|A\rangle\rangle) = H_{\frac{1}{1-s}}(\rho). \tag{3.39}$$

Taking the limit $s \to 0$, we have

$$E_R(|A\rangle\rangle) = H(\rho). \tag{3.40}$$

Combining the first inequality in (3.27) and the relation (3.37) with the limit $s \to -1$, we have $R_{L,g}(|A\rangle\rangle) \geq H_{\frac{1}{2}}(\rho)$. In fact, since the relation $R_L(|A\rangle\rangle) \leq H_{\frac{1}{2}}(\rho)$ holds [124], we obtain

$$R_L(|A\rangle\rangle) = R_{L,g}(|A\rangle\rangle) = H_{\frac{1}{2}}(\rho) \tag{3.41}$$

as will be shown later.

Proof of (3.41) We choose $R = (\sum_{j=1}^k \sqrt{d_j})^2$ and focus on the separable state $\rho^- := \frac{1}{R-1} \sum_{j \neq l} \sqrt{d_j}\sqrt{d_l} |v_j\rangle\langle v_j| \otimes |u_l\rangle\langle u_l|$. In the following, we show that the state

$$\begin{aligned} \rho^+ &:= \frac{|A\rangle\rangle\langle\langle A| + (R-1)\rho^-}{R} \\ &= \sum_{j,l} \frac{\sqrt{d_j d_l}}{R} |v_j\rangle\langle v_l| \otimes |u_j\rangle\langle u_l| + \sum_{j \neq l} \frac{\sqrt{d_j d_l}}{R} |v_j\rangle\langle v_j| \otimes |u_l\rangle\langle u_l| \end{aligned}$$

is separable. For this purpose, we choose $\alpha_0, \ldots, \alpha_k$ such that $\alpha_j = 2\alpha_{j-1}+1, \alpha_0 = 0$. Then, defining $|e_t\rangle := \sum_{j=1}^{k}(\frac{d_j}{R})^{1/4} e^{\frac{i2\pi}{\alpha_k}\alpha_j t}|v_j\rangle$, $|e_t^*\rangle := \sum_{j=1}^{k}(\frac{d_j}{R})^{1/4} e^{-\frac{i2\pi}{\alpha_k}\alpha_j t}|u_j\rangle$, we will show that $\rho^+ = \sum_{t=1}^{\alpha_k} |e_t\rangle\langle e_t| \otimes |e_t^*\rangle\langle e_t^*|$. We find that

$$\begin{aligned}&\langle v_j| \otimes \langle u_l| \sum_{t=1}^{\alpha_k} |e_t\rangle\langle e_t| \otimes |e_t^*\rangle\langle e_t^*||v_{j'}\rangle \otimes |u_{l'}\rangle\\ =&\frac{(d_j d_l d_{j'} d_{l'})^{1/4}}{R} \sum_{t=1}^{\alpha_k} e^{\frac{i2\pi}{\alpha_k}(\alpha_j-\alpha_l-\alpha_{j'}+\alpha_{l'})t}.\end{aligned}$$

Here, the RHS is zero if and only if the real number $\alpha_j - \alpha_l - \alpha_{j'} + \alpha_{l'}$ is not zero. This real number is zero if and only if the relations $j = l$ and $j' = l'$ or the relations $j = j'$ and $l = l'$ hold. Thus, $\sum_{t=1}^{\alpha_k} |e_t\rangle\langle e_t| \otimes |e_t^*\rangle\langle e_t^*|$ is ρ^+. That is, ρ^+ is separable. Since $R_L(|A\rangle\rangle) \le 2\psi(1/2|\rho) = H_{\frac{1}{2}}(\rho)$, we obtain (3.41). ■

Proof of (3.37) Let ρ_1 be a separable state. Theorem 3.2 yields that $\mathrm{Tr}_2\, \rho_1 \otimes I_2 \ge \rho_1$. When $0 > s > -1$, since the function $x \mapsto -x^s$ is matrix monotone, the state $\rho' := \mathrm{Tr}_2\, \rho_1$ satisfies $\rho_1^s \ge (\mathrm{Tr}_2\, \rho_1 \otimes I_2)^s = \rho'^s \otimes I_2$. Hence, (2.23) implies that

$$\begin{aligned}\mathrm{Tr}\, \rho_1^s |A\rangle\rangle\langle\langle A|^{1-s} &= \mathrm{Tr}\, \rho_1^s |A\rangle\rangle\langle\langle A| \ge \mathrm{Tr}\, \rho'^s \otimes I_2 |A\rangle\rangle\langle\langle A|\\ &= \mathrm{Tr}\, \rho'^s \rho \ge (\mathrm{Tr}\, \rho^{\frac{1}{1-s}})^{1-s}.\end{aligned}$$

We diagonalize the state $\frac{\rho^{\frac{1}{1-s}}}{\mathrm{Tr}\, \rho^{\frac{1}{1-s}}}$ as $\sum_j d_j' |v_j\rangle\langle v_j|$, and define $v_j' := \frac{Av_j}{\|Av_j\|}$. Then, the equality in the above inequalities holds when $\rho_1 = \sum_j d_j' |v_j\rangle\langle v_j| \otimes |v_j'\rangle\langle v_j'|$. So, we obtain (3.37). ■

Proof of (3.38) Let ρ_1 be a separable state. Theorem 3.2 shows that $\mathrm{Tr}_2\, \rho_1 \otimes I_2 \ge \rho_1$. When $1 > s > 0$, since the function $x \mapsto x^s$ is matrix monotone, the state $\rho' := \mathrm{Tr}_2\, \rho_1$ satisfies $\rho_1^s \le (\mathrm{Tr}_2\, \rho_1 \otimes I_2)^s = \rho'^s \otimes I_2$. Hence, (2.24) guarantees that

$$\begin{aligned}\mathrm{Tr}\, \rho_1^s |A\rangle\rangle\langle\langle A|^{1-s} &= \mathrm{Tr}\, \rho_1^s |A\rangle\rangle\langle\langle A| \le \mathrm{Tr}\, \rho'^s \otimes I_2 |A\rangle\rangle\langle\langle A|\\ &= \mathrm{Tr}\, \rho'^s \rho \le (\mathrm{Tr}\, \rho^{\frac{1}{1-s}})^{1-s}.\end{aligned}$$

Similar to the proof of (3.37), the equality in the above inequalities holds when $\rho_1 = \sum_j d_j' |v_j\rangle\langle v_j| \otimes |v_j'\rangle\langle v_j'|$. So, we obtain (3.38). ■

Example 3.7 (*Dicke state [59, 132]*) Assume that $\mathcal{H}_i = \mathbb{C}^r$. When a non-negative integer-valued vector $\boldsymbol{n} = (n_1, \ldots, n_r)$ has the sum n, we define **Dicke state**

$$|\boldsymbol{n}; \mathrm{Dicke}\rangle := \sum_{(\omega_1,\ldots,\omega_n)\in T_{\boldsymbol{n}}} \frac{1}{\sqrt{|T_{\boldsymbol{n}}|}} |\omega_1\rangle \otimes \cdots \otimes |\omega_n\rangle,$$

where $T_{\boldsymbol{n}}$ is the set of elements of $\{1, \ldots, r\}^n$ such that n_j is the number of indexes whose entry is j for $j = 1, \ldots, r$.

In the following, we calculate $E_{G,2}(|\boldsymbol{n}\rangle)$. Due to Theorem 3.5, this calculation is reduced to the calculation of the RHS of (3.35). Further, the vector appearing in the RHS of (3.35) can be restricted to vector with non-negative entries under the standard basis due to the property of the maximization problem. Given a vector $\boldsymbol{p} = (p_1, \ldots, p_r)$ forming a probability distribution, the vector state $|\boldsymbol{p}\rangle := \sum_{j=1}^{r} \sqrt{p_j}|j\rangle$ satisfies

$$\begin{aligned}\langle \boldsymbol{p}|^{\otimes n}|\boldsymbol{n}; \text{Dicke}\rangle &= \sqrt{|T_{\boldsymbol{n}}|}\prod_{j=1}^{r}\sqrt{p_j}^{n_j} = \sqrt{|T_{\boldsymbol{n}}|}e^{\frac{n}{2}\sum_{j=1}^{r}\frac{n_j}{n}\log p_j}\\ =&\sqrt{|T_{\boldsymbol{n}}|}e^{\frac{n}{2}(-H(\frac{\boldsymbol{n}}{n})-D(\frac{\boldsymbol{n}}{n}\|\boldsymbol{p}))} \le \sqrt{|T_{\boldsymbol{n}}|}e^{\frac{n}{2}-H(\frac{\boldsymbol{n}}{n})}.\end{aligned}$$

Since the equality in the above inequality holds when $\boldsymbol{p} = \frac{\boldsymbol{n}}{n}$, we obtain $E_{G,2}(|\boldsymbol{n}; \text{Dicke}\rangle) = nH(\frac{\boldsymbol{n}}{n}) - \log|T_{\boldsymbol{n}}|$.

Next, showing that Theorem 3.4 can be applied, we will prove (3.34). We embed $\mathrm{U}(1)^r$ in $\mathrm{U}(r)$ as a subgroup $\{\sum_{j=1}^{r} e^{i\theta_j}|j\rangle\langle j|\}$. Let f be the fundamental representation of $U(r)$. Under the restriction of the representation $\mathsf{f}^{\otimes n}_{(1,0,\ldots,0)}$ to the subgroup $U(1)^r$, the condition **GM1** holds. Then, the vector state $|\frac{\boldsymbol{n}}{n}\rangle$ satisfies the condition **GM2**, and additionally, the condition **GM3′** on the symmetric tensor product space. Thus, Lemma 3.7 yields (3.34).

Further, Lemma 3.6 implies $E_{G,2}(|\boldsymbol{n}; \text{Dicke}\rangle^{\otimes m}) = m(nH(\frac{\boldsymbol{n}}{n}) - \log|T_{\boldsymbol{n}}|)$. Thus, Lemma 3.8 guarantees that the vector state $|\boldsymbol{n}\rangle$ satisfies (3.36).

Example 3.8 (*Stabilizer state [59]*) Consider discrete Heisenberg representation $\mathsf{W}^r_{\mathbb{F}_q}$ of $\mathbb{F}_q^{2r}$ on the composite system $(\mathbb{C}^q)^{\otimes r} := \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_r$. The stabilizer state of a strictly self-orthogonal subgroup N of $\mathbb{F}_q^{2r}$ and the restriction of discrete Heisenberg representation $\mathsf{W}^r_{\mathbb{F}_q}$ to N satisfy the conditions **GM1** and **GM4**. Hence, due to Lemma 3.7, the stabilizer state of N satisfies Theorem 3.4.

Example 3.9 (*Antisymetric state [59, 132]*) Consider the case when $\mathcal{H}_i = \mathbb{C}^r$ with $n \le r$. Let P be the projection to the antisymetric tensor space. Then, the state $\frac{P}{\mathrm{Tr}\,P}$ is called the **antisymetric state**. The property of the antisymetric tensor space guarantees that $\|P|a_1 \otimes \cdots \otimes a_n\rangle\| \le \frac{1}{\sqrt{n!}}\|a_1 \otimes \cdots \otimes a_n\|$. The equality holds only when the vectors $a_1, \ldots, a_n$ are orthogonal to each other. Hence, we obtain $\max_{|a\rangle\langle a| \in \mathrm{SEP}}\langle a|P|a\rangle = \max_{|a\rangle\langle a| \in \mathrm{SEP}}\langle a|P^2|a\rangle = \frac{1}{n!}$.

Further, the n-foldtensor product representation of $\mathrm{U}(r)$ satisfies the condition **GM1** and **GM4**. Lemma 3.7 implies that

$$R_{L,g}(\frac{P}{\mathrm{Tr}\,P}) = E_R(\frac{P}{\mathrm{Tr}\,P}) = E_{R,s}(\frac{P}{\mathrm{Tr}\,P}) = E_{G,1}(\frac{P}{\mathrm{Tr}\,P}) = \log n!$$

for $1 > \forall s > 0$.

However, the projection P contains negative entries, we cannot apply Theorem 3.6, and need to calculate $E_{G,2}(\frac{P^{\otimes 2}}{(\mathrm{Tr}\,P)^2})$ without use of Theorem 3.6. Since the projection $P^{\otimes 2}$ is invariant with respect to the action of permutation group, we can apply Theorem 3.5, and obtain $\max_{|a\rangle\langle a| \in \mathrm{SEP}}\langle a|P^{\otimes 2}|a\rangle = \max_{\mathrm{Tr}\,A^\dagger A=1}\langle\langle A^{\otimes n}|P^{\otimes 2}|A^{\otimes n}\rangle\rangle$.

Let $a_1, \dots, a_r$ be eigenvalues of $A^\dagger A$. Since P is commutative with $A^{\dagger\otimes n}$, we have $\langle\langle A^{\otimes n}|P^{\otimes 2}|A^{\otimes n}\rangle\rangle = \mathrm{Tr}\, A^{\dagger \otimes n} P A^{\otimes n} P^T = \mathrm{Tr}\, P A^\dagger A^{\otimes n} = \sum_{1 \le j_1 < \cdots < j_n \le r} a_{j_1} \cdots a_{j_n}$. Since this value is a Schur function for $a_1, \dots, a_r$ and Example 2.1 guarantees that this function is a Schur concave function, the maximum value of this value under the condition $\sum_{j=1}^r a_j = 1$ is attained when $a_j = \frac{1}{r}$. Hence, $\max_{|a\rangle\langle a| \in \mathrm{SEP}} \langle a|P^{\otimes 2}|a\rangle = \frac{r!}{r^n n!(r-n)!}$. Applying Lemma 3.8, we have

$$\begin{aligned} & R_{L,g}(\frac{P^{\otimes 2}}{(\mathrm{Tr}\, P)^2}) = E_R(\frac{P^{\otimes 2}}{(\mathrm{Tr}\, P)^2}) = E_{R,s}(\frac{P^{\otimes 2}}{(\mathrm{Tr}\, P)^2}) = E_{G,1}(\frac{P^{\otimes 2}}{(\mathrm{Tr}\, P)^2}) \\ = & \log \frac{r^n n!(r-n)!}{r!} < 2 \log n!, \quad 1 > \forall s > 0. \end{aligned}$$

Then, the additivity of $R_{L,g}$, E_R, $E_{R,s}$, and $E_{G,1}$ do not hold for the antisymmetric state.

3.5.3 Extension Based on PPT States

In the bipartite case, i.e., the case when $r = 2$, we can change the set SEP of separable states to the set PPT of PPT states as the range of the minimization for $E[f]_{1,2}$. Then, we obtain new entanglement measure, which is denoted by $E[f]_{1,2}^{\mathrm{PPT}}$. A quantum channel Λ on the composite system is called a **PPT quantum channel** when $\Lambda(\rho)$ is a PPT state for any PPT input state ρ. Similar to the proof of Lemma 3.4, we can show that a PPT quantum channel Λ satisfies

$$E[f]_{1,2}^{\mathrm{PPT}}(\Lambda(\rho)) \le E[f]_{1,2}^{\mathrm{PPT}}(\rho). \tag{3.42}$$

Then, we see that a separable quantum channel Λ is a PPT quantum channel. Hence, a separable quantum channel Λ satisfies (3.42). That is, $E[f]_{1,2}^{\mathrm{PPT}}$ is separable entanglement monotone. Similarly, we can replace SEP by PPT in the definitions of $R_{L:1,2}$ and $R_{L,g:1,2}$. Such modified measures are denoted by $R_{L:1,2}^{\mathrm{PPT}}$ and $R_{L,g:1,2}^{\mathrm{PPT}}$, and are separable entanglement monotone. Further, an entanglement measure is called **PPT monotone** when the inequality (3.42) holds for any PPT quantum channel Λ.

Proof of (2) ⇒ (3) in Theorem 3.2 We focus on $E[f]_{1,2}^{\mathrm{PPT}}$ when f is relative entropy. Then, $E[f]_{1,2}^{\mathrm{PPT}}(\rho) = 0$ if and only if ρ is a PPT state. Also, the function $\rho \mapsto E[f]_{1,2}^{\mathrm{PPT}}(\rho)$ is continuous. We choose a PPT state ρ and an LOCC Λ_n on n-fold tensor product system. Then, we have $E[f]_{1,2}^{\mathrm{PPT}}(\Lambda_n(\rho^{\otimes n})) = 0$. When $\Lambda_n(\rho^{\otimes n})$ converges to a state ρ_0, the continuity of $E[f]_{1,2}^{\mathrm{PPT}}$ implies $E[f]_{1,2}^{\mathrm{PPT}}(\rho_0) = 0$. On the other hand, the maximally entangled state $|\Phi\rangle\langle\Phi|$ on the system $(\mathbb{C}^2)^{\otimes 2}$ satisfies $E[f]_{1,2}^{\mathrm{PPT}}(|\Phi\rangle\langle\Phi|) > 0$. Hence, we obtain **(3)**. ■

3.6 Quantification of Entanglement II: Operational Method

3.6.1 Bipartite System

In the above discussion, we quantified the amount of entanglement by measuring the difference from separable states. However, such a discussion cannot directly characterize the ability of the given entangled state as a resource. One solution for this problem is to measure the number of extracted maximally entangled states quantitatively via a distillation protocol.

In this method, given an entangled state ρ on the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$, we optimize our LOCC so that the resultant state is sufficiently close to as many copies of maximally entangled state as possible. However, it is difficult to analyze this quantity with one copy state ρ. To resolve this difficulty, we discuss the n-fold tensor product state $\rho^{\otimes n}$ on the bipartite system $(\mathcal{H}_A^{\otimes n}) \otimes (\mathcal{H}_B^{\otimes n})$ and treat the asymptotics $n \to \infty$. In this scenario, a maximally entangled state $|X_n\rangle := \frac{1}{\sqrt{d_n}}|I\rangle\rangle$ on the bipartite system $\mathcal{H}_{A,n} \otimes \mathcal{H}_{B,n}$ is given as the target, where we assume that both dimensions of $\mathcal{H}_{B,n}$ and $\mathcal{H}_{A,n}$ are d_n. Now, we apply an LOCC Λ_n whose input and output systems are the composite system $\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$ and the composite system $\mathcal{H}_{A,n} \otimes \mathcal{H}_{B,n}$, respectively when the initial state is $\rho^{\otimes n}$. This protocol is described by $\Phi_n := (|X_n\rangle, \Lambda_n)$, and its performance is characterized by the size $|\Phi_n| := d_n$ and the error $\epsilon(\Phi_n) := 1 - \langle X_n|\Lambda_n(\rho^{\otimes n})|X_n\rangle$, which is determined by the fidelity between the target and the resultant state. Hence, under the condition that the error $\epsilon(\Phi_n)$ converges to 0, we maximize the generation rate $\lim_{n\to\infty} \frac{\log|\Phi_n|}{n}$ of maximally entangled state per one copy. This maximum is denoted by $E_D(\rho)$ [4], and is called **entanglement distillation** of ρ. As in Fig. 3.5, this quantity is mathematically defined as

$$E_D(\rho) := \sup_{\{\Phi_n\}} \left\{ \limsup_{n\to\infty} \frac{\log|\Phi_n|}{n} \middle| \lim_{n\to\infty} \epsilon(\Phi_n) = 0 \right\}.$$

The entanglement distillation $E_D(\rho)$ is a quantity to measure the amount of entanglement in the entangled state ρ as a resource generating maximally entangled state.

Next, we consider how large maximally entangled state is required to generate a specific entangled state via LOCC. In contrast to entanglement distillation, we consider the n-fold tensor product state $\rho^{\otimes n}$ as our target when n goes to infinity. Firstly, we set a maximally entangled state $|X_n\rangle$ on the composite system $\mathcal{H}_{A,n} \otimes \mathcal{H}_{B,n}$ with size d_n as the initial state. Then, using an LOCC Λ_n from the composite system $\mathcal{H}_{A,n} \otimes \mathcal{H}_{B,n}$ to the composite system $\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$, we convert the initial state $|X_n\rangle$ to

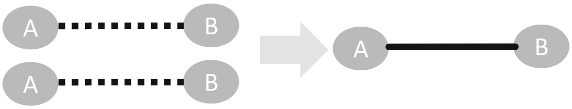


Fig. 3.5 Entanglement distillation

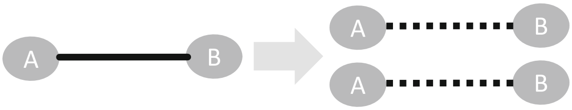


Fig. 3.6 Entanglement dilution

plural copies of a specific state ρ, as in Fig. 3.6. This protocol is called an entanglement dilution, and is described by $\Psi_n := (|X_n\rangle, \Lambda_n)$. Its performance is characterized by the size $|\Psi_n| := d_n$ and the error $\epsilon(\Psi_n) := 1 - F(\Lambda_n(|X_n\rangle\langle X_n|), \rho^{\otimes n})$, which is determined by the fidelity between the target and the resultant state. Hence, under the condition that the error $\epsilon(\Phi_n)$ goes to 0 as $n \to \infty$, we minimize the conversion rate $\lim_{n\to\infty} \frac{\log|\Phi_n|}{n}$ to generate the entangled state ρ from maximally entangled state. This minimum is denoted by $E_C(\rho)$ [4], and is called the **entanglement cost** of ρ. That is, the entanglement cost $E_C(\rho)$ is mathematically defined as

$$E_C(\rho) := \inf_{\{\Psi_n\}} \left\{ \liminf_{n\to\infty} \frac{\log|\Psi_n|}{n} \middle| \lim_{n\to\infty} \epsilon(\Psi_n) = 0 \right\}.$$

The entanglement cost $E_C(\rho)$ is the quantity to measure the size of maximally entangled state demanded to generate the entangled state ρ per one copy. Considering the operational meaning of the entanglement cost $E_C(\rho)$ and the entanglement distillation $E_D(\rho)$, we obtain

$$E_D(\rho) \le E_C(\rho). \tag{3.43}$$

This relation can be derived by considering the following protocol. Firstly, we prepare the maximally entangled state with a specific size. Then, we convert it to the state $\rho^{\otimes n}$ by the LOCC attaining the optimal rate of the entanglement cost. Next, we convert the resultant state to the maximally entangled state by the LOCC attaining the optimal rate of the entanglement distillation. Since the concatenated protocol of both protocols is an LOCC, the size of the resultant maximally entangled state is less than the size of the initial maximally entangled state. Hence, we obtain (3.43).

Further, it is known that [26, 27] [45, (8.123)]

$$E_D(\rho) \le E_R(\rho) \le E_C(\rho). \tag{3.44}$$

When ρ is a pure state, $E_R(|X\rangle)$ is $H(\rho_A)$, where ρ_A is the reduced density matrix of $|X\rangle$ on the system $\mathcal{H}_A$. In fact, the following relation is known [4].

$$E_D(|X\rangle) = E_R(|X\rangle) = E_C(|X\rangle) = H(\rho_A). \tag{3.45}$$

The relation $E_D(|X\rangle) \ge H(\rho_A)$ can be shown by the protocol given in Sect. 6.5. For the proof of the relation $E_C(|X\rangle) \le H(\rho_A)$ [4], see [45, Sect. 8.6].

3.6.2 Multipartite System

In the multipartite system, we have only a limited number of methods to quantify the amount of entanglement operationally. We have the following relation between quantification of entanglement and state distinguishability under the locality condition. We define the following quantity for an entangled state ρ on the multipartite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_r$.

$$L(\rho) := \max_{L} \{L | \exists L, \exists \rho_1, \ldots, \rho_L, \exists M = \{M_j\}_{j=1}^{L} \text{ s.t. } \operatorname{Tr} M_j \rho_j = 1\},$$

where M is restricted to a separable POVM, and ρ_j is restricted to an entangled state equivalent to the entangled state ρ as an entangled state.

Theorem 3.7 ([58]) *Let D be the dimension of the composite system $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_r$. Then, the inequality* $\log \frac{D}{L(\rho)} \geq R_{L,g}(\rho)$ *holds.*

Proof We set $L := L(\rho)$. Then, we choose L states $\rho_1, \ldots, \rho_L$ and the separable POVM $M = \{M_j\}_{j=1}^{L}$ such that the condition of definition of $L(\rho)$ holds. Then, there exists an integer j such that the condition $\operatorname{Tr} M_j \leq \frac{D}{L(\rho)}$ holds. When $1 + t = \operatorname{Tr} M_j$, the matrix $\rho_j' := \frac{M_j - \rho_j}{t}$ is a state. Then, the state $\frac{M_j}{\operatorname{Tr} M_j} = \frac{\rho_j + t\rho_j'}{1+t}$ is a separable state. So, we obtain the desired argument. ■

Especially, when ρ is the stabilizer state of a strictly self-orthogonal subgroup N, we focus on a subgroup N' of N such that all elements of N' can be simultaneously measured by a separable POVM. When L is the number of stabilizer states of N that can be distinguished by all of elements N', we have $L(\rho) \geq L$. In particular, in the case of a graph state, when the vertexes are classified into two colors (white and black) such that the same color vertexes are not connected to each other, such a graph is called two-colorable and $L(\rho)$ is lower bounded as follows when the number of vertexes is r [90]. When the vertex j is white, we define the vector $e_j \in \mathbb{F}_2^{2r}$ such that the j-th entry is 1 and other entries are 0. When the vertex j is black, we define the vector $e_j \in \mathbb{F}_2^{2r}$ such that the $r + j$-th entry is 1 and other entries are 0. Since the all of elements of the subgroup N'' of $\mathbb{F}_2^{2r}$ generated by $e_1, \ldots, e_r$ are commutative with each other and can be measured by a separable POVM, the all of elements of the intersection $N' := N \cap N''$ are can be simultaneously measured by a separable POVM. Let m be the maximum of the number of black vertexes and the number of white vertexes. Then, the number of element N' is 2^m. Hence, the number of locally distinguishable stabilizer states by N' is 2^m. So, $L(\rho)$ is lower bounded by 2^m. That is, Theorem 3.7 guarantees that $R_{L,g}(\rho)$ is upper bounded by $\log \frac{2^r}{2^m} = (r - m) \log 2$.

When we need more than three colors for color coding, our situation is more complicated. First, we choose the first color such that the number of vertexes of the first color is largest. Then, we define e_j for the first color in the same way as white. Second, we choose the second color such that the number of vertexes of the second color is second largest. Then, we define e_j for the second color in the same

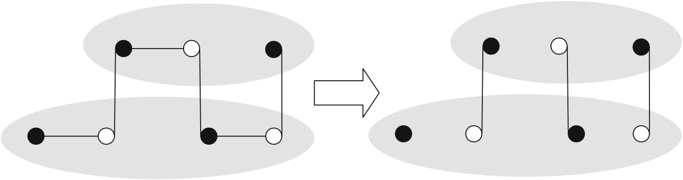


Fig. 3.7 1D cluster state

way as black. Let m be the number of vertexes of the first color. Due to the same discussion as the above, we can define the subgroup N', and the number of locally distinguishable stabilizer states by N' is 2^m.

In the pure state case, once we have $R_{L,g} = E_{G,2}$, (3.32) guarantees that all of other entanglement measures take the same value. Hence, Using this fact, and (3.26), (3.30), we can calculate $R_{L,g}$ and $E_{G,2}$ for several graph states as follows.

Example 3.10 (*1D Cluster state [90]*) In the 1D Cluster state ρ of r vertexes, the graph is coded with two colors as the left figure in Fig. 3.7. The above value m is $\lceil \frac{r}{2} \rceil$. Hence, due to Theorem 3.7, $R_{L,g}(\rho)$ is upper bounded by $\log \frac{2^r}{2^{\lceil \frac{r}{2} \rceil}} = \lfloor \frac{r}{2} \rfloor \log 2$. On the other hand, we divide the r vertexes in to two groups S and S^c as the left figure in Fig. 3.7. When we focus on the entanglement of the bipartite system composed of these two groups, The C-NOT operations inside of both groups are regarded as local unitaries. Hence, applying the C-NOT operations along the edges inside of both groups like the right figure of Fig. 3.7, we obtain $E_{G,2:S,S^c}(\rho) = \lfloor \frac{r}{2} \rfloor \log 2$, which is a lower bound of $E_{G,2}(\rho)$ due to (3.26). Therefore, using (3.31), we obtain $R_{L,g}(\rho) = E_{G,2}(\rho) = \lfloor \frac{r}{2} \rfloor \log 2$.

Example 3.11 (*Ring state [90]*) When r is an even number and r vertexes form the graph of the ring state ρ, the graph is coded with two colors as in the first figure of Fig. 3.8. Then, the above value m is $\frac{r}{2}$. When r is an odd number and r vertexes form the graph of the ring state ρ, the graph is coded with two colors as in the first figure of Fig. 3.9. Then, the above value m is $\lfloor \frac{r}{2} \rfloor$. Hence, due to Theorem 3.7, $R_{L,g}(\rho)$ is upper bounded by $\log \frac{2^r}{2^{\lfloor \frac{r}{2} \rfloor}} = \lceil \frac{r}{2} \rceil \log 2$.

When r is even, the r vertexes can be divided into two groups S and S^c as the first figure of Fig. 3.8. That is, we focus on the bipartite system composed of these two groups. Then, according to arrow 1, we apply C-NOT operations inside of groups. According to arrow 2, we apply the local complementation. Finally, according to arrow 3, we apply C-NOT operations inside of groups. Then, we obtain $E_{G,2:S,S^c}(\rho) = \frac{r}{2} \log 2$, which is a lower bound of $E_{G,2}(\rho)$ due to (3.26). Therefore, using (3.31), we obtain $R_{L,g}(\rho) = E_{G,2}(\rho) = \frac{r}{2} \log 2$.

When r is odd, the r vertexes can be divided into two groups S and S^c as the first figure of Fig. 3.9. That is, we focus on the bipartite system composed of these two groups. Then, similar to the even case, by converting the graph state according to the arrows, we obtain $E_{G,1:S,S^c}(\rho) = \lfloor \frac{r}{2} \rfloor \log 2$, which is a lower bound of $E_{G,2}(\rho)$ due

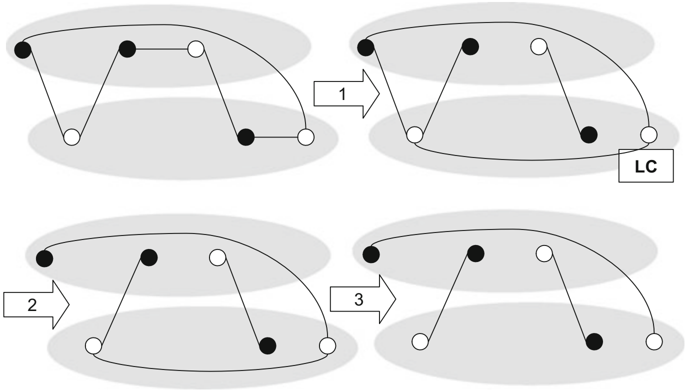


Fig. 3.8 Ring state with even number of vertexes: *LC* shows the local complementation

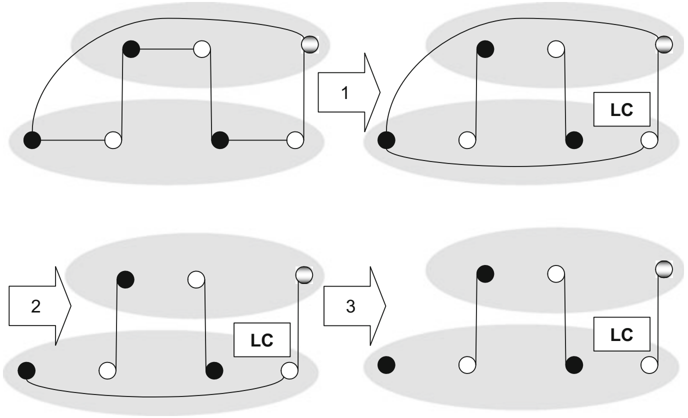


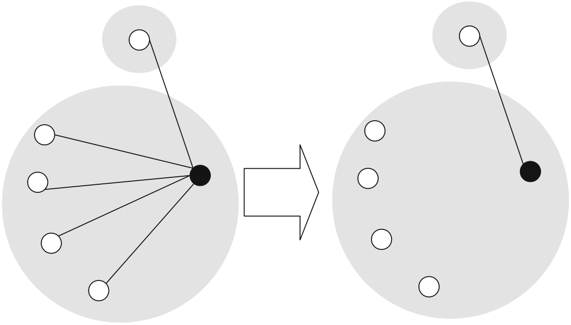
Fig. 3.9 Ring state with odd number of vertexes

to (3.26). Unfortunately, the upper bound $\lceil \frac{r}{2} \rceil \log 2$ is different from the lower bound $\lfloor \frac{r}{2} \rfloor \log 2$. However, when $r = 3$, as calculated in Exercise 3.10, $E_{G,2}(\rho)$ equals $2 \log 2$, which is the lower bound.

Example 3.12 (*GHZ state [90]*) Consider the GHZ state ρ composed of r vertexes, which is given by the star graph. While Exercise 3.8 discusses $E_{G,2}(\rho)$ with the concrete form of the state, we discuss it without use of the concrete form, now. This graph is coded with two colors as the left figure in Fig. 3.10. Then, the above value m is $r - 1$. So, $R_{L,g}(\rho)$ is upper bounded by $\log 2$.

On the other hand, the r vertexes can be divided into two groups S and S^c as the left figure of Fig. 3.10. Then, applying C-NOT operation inside of the two groups, we have $E_{G,1:S,S^c}(\rho) = \log 2$, which is a lower bound of $E_{G,2}(\rho)$ due to (3.26). Therefore, using (3.31), we obtain $R_{L,g}(\rho) = E_{G,2}(\rho) = \log 2$.

Fig. 3.10 GHZ state



It is known that we can calculate $R_{L,g}(\rho)$ and $E_{G,2}(\rho)$ of 2D Cluster state in the same way [90].

3.7 Quantification of Entanglement III: Convex Decomposition Method

The discussion in Sect. 3.6 shows that it is suitable to use the von Neumann entropy of the reduced density matrix as the criterion of the entropy for bipartite pure states. In the following, we discuss only the bipartite case. Then, we define a function $E(\rho)$ only for the case when ρ is a pure state to be $H(\mathrm{Tr}_B\,\rho)$. Here, we notice that the equality in (3.44) does not necessarily hold for a mixed state. Hence, we need to choose another criterion. As one solution, we employ the convex roof defined in Sect. 2.3.1. That is, we define the **entanglement formation** $E_f(\rho) := \mathrm{CR}(E)(\rho)$ as the convex roof of $E(\rho)$ [7]. Then, since $E_R(\rho)$ is a convex function due to Lemma 3.5, Lemma 2.2 guarantees that $E_R(\rho) \le E_f(\rho)$. Now, we choose the purification $|\sqrt{\rho}\rangle\rangle$ of ρ with the reference system $\mathcal{H}_C$. Using (1.4), we have an expression of $E_f(\rho)$ with use of a POVM as

$$E_f(\rho) = \min_{\{M_j\}:\mathrm{POVM}} \sum_j (\mathrm{Tr}\, M_j\rho) E(\frac{\sqrt{\rho} M_j \sqrt{\rho}}{\mathrm{Tr}\, M_j\rho}). \tag{3.46}$$

Since $E(\rho)$ satisfies the additivity, the definition of convex roof implies that

$$E_f(\rho_1 \otimes \rho_2) \le E_f(\rho_1) + E_f(\rho_2). \tag{3.47}$$

It had been an open problem in quantum information theory for a long time whether the equality in this inequality holds in general. It was negatively solved by Hastings [40]. That is, the equality in (3.47) does not hold in general.

Then, the following theorem is known.

Theorem 3.8 ([4, 64])

$$\lim_{n\to\infty}\frac{1}{n}E_f(\rho^{\otimes n}) = E_C(\rho). \tag{3.48}$$

On the other hand, E_f is closely related to the **Holevo capcity** $\chi(\Lambda)$ [71] of a quantum channel Λ.

$$\chi(\Lambda) := \sup_\rho H(\Lambda(\rho)) - \min_{(p_j,\rho_j):\sum_j p_j\rho_j=\rho}\sum_j p_j H(\Lambda(\rho_j)). \tag{3.49}$$

Now, we choose Stinespring representation of Λ such that the environment system is $\mathcal{E}$ and the isometry is U. The second term of (3.49) equals the entanglement formation of the state $U\rho U^\dagger$ on the composite system $\mathcal{K}\otimes\mathcal{E}$ [94]. That is, we have

$$\chi(\Lambda) = \sup_\rho H(\Lambda(\rho)) - E_f(U\rho U^\dagger). \tag{3.50}$$

When the state has a symmetry with respect to a projective unitary representation, the entanglement formation $E_f(\rho)$ can be simply calculated as follows.

Lemma 3.10 *Let f_A and f_B be projective unitary representations of the group G on the systems $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Then, we choose an irreducible subspace $\mathcal{H}_C$ of the composite system $\mathcal{H}_A\otimes\mathcal{H}_B$ of the tensor product representation $\mathsf{f}_A\otimes\mathsf{f}_B$. The completely mixed state $\rho_{\mathrm{mix},C}$ on $\mathcal{H}_C$ satisfies*

$$E_f(\rho_{\mathrm{mix},C}) = \min_{|x\rangle\in\mathcal{H}_C,\ \|x\|=1} H(\mathrm{Tr}\,|x\rangle\langle x|). \tag{3.51}$$

Proof Firstly, we obtain the inequality $\geq$ of (3.51) from the definition of convex roof.

Further, any vector state $|x\rangle\in\mathcal{H}_C$ and any element $g\in G$ satisfy

$$H(\mathrm{Tr}\,|x\rangle\langle x|) = H(\mathsf{f}_A(g)\,\mathrm{Tr}\,|x\rangle\langle x|\mathsf{f}_A(g^{-1})).$$

We also find that $\int_G \mathsf{f}_A\otimes\mathsf{f}_B(g)|x\rangle\langle x|\mathsf{f}_A\otimes\mathsf{f}_B(g^{-1})\mu(dg)$ equals $\rho_{\mathrm{mix},C}$. Hence, we obtain the inequality $\leq$ of (3.51). ■

The logarithmic inverse square fidelity $E_{G,1}(\rho)$ is given as a function of $E[1-F^2](\rho)$. Although $E_{G,1}(\rho)$ is not self convex roof, we have the following theorem for $E[1-F^2](\rho)$.

Theorem 3.9 ([121]) *$E[1-F^2]$ is self convex roof. That is,* $E[1-F^2](\rho) = \mathrm{CR}(E[1-F^2]_{\mathcal{P}(\mathcal{H})})$, *where $\mathcal{P}(\mathcal{H})$ is the set of pure states on $\mathcal{H}$.*

Chapter 4
Group Covariance and Optimal Information Processing

Abstract In quantum information processing, we need to decrease the amount of error as much as possible. To achieve this requirement, we mathematically formulate the error as a function, and seek the information processing to minimize (optimize) it. This chapter gives optimal scheme including optimal measurement under several setting with the group covariance. This approach can be applied to so many topics in quantum information, the quantum state estimation, the estimation of unknown unitary action, and the approximate quantum state cloning. The problem to optimize the estimation procedure cannot be exactly solved with the finite resource setting, in general, however, it can be exactly solved only when the problem has group symmetric property. Hence, we can say that group symmetric property is essential for state estimation. This idea can be applied to the estimation of unknown unitary action. This problem is much more deeply related to group representation because this problem can be investigated by using Fourier transform in the sense of group representation theory. Further, we deal with the approximate quantum state cloning in a similar way because we can say the same thing for approximate quantum state cloning. That is, all of solved examples of approximate quantum state cloning are based on the group symmetry.

4.1 Covariant State Family and Optimal Measurement

4.1.1 Covariant State Family and Covariant Measurement

Let us consider the case when the information of our interest takes values in the set Θ and the set Θ is closed with respect to the action of the group G [70]. When the group G transitively acts on Θ, there exists a subgroup H of G such that $\Theta \cong G/H$. How does the density matrix ρ reflect such information $\theta \in \Theta$ with group covariant structure? To satisfy this requirement, the following structure is needed.

Given a (projective) unitary representation f of a group G on a quantum system $\mathcal{H}$, a state family (a parametric set of states) $\{\rho_\theta | \theta \in \Theta\}$ is called **covariant** with respect to the representation f when

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8_4

$$\mathsf{f}(g)\rho_\theta \mathsf{f}(g)^\dagger = \rho_{g\theta}$$

for $\theta \in \Theta$ and $g \in G$. Now, we fix a point $\theta_0 \in \Theta$, which is called the **origin**. Then, since the action of G is transitive, for any $\theta \in \Theta$, there exists an element $g(\theta) \in G$ such that $\theta = g(\theta)\theta_0$. Hence, the above condition is changed to $\rho_\theta = \mathsf{f}(g(\theta))\rho_{\theta_0}\mathsf{f}(g(\theta))^\dagger$. Given a stabilizer H, we choose a density matrix ρ_0 such that

$$\mathsf{f}(h)\rho_0 \mathsf{f}(h)^\dagger = \rho_0, \quad \forall h \in H. \tag{4.1}$$

Then, the covariant state family with respect to the representation f is given by

$$\rho_\theta = \mathsf{f}(g(\theta))\rho_0 \mathsf{f}(g(\theta))^\dagger. \tag{4.2}$$

As an example of covariant state family, when the group G is $\mathrm{SU}(2)$, we consider a covariant state family on the irreducible space $\mathcal{U}_\lambda$ with highest weight λ. Let f_λ be the irreducible representation on $\mathcal{U}_\lambda$. Using the element $F_1^z := \frac{1}{2}\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$ of $\mathfrak{su}(2)$, we choose the subgroup H to be $H = \{\exp(tF_1^z)\}_{t\in\mathbb{R}}$. Let $|\lambda : \zeta\rangle$ be the coherent vector in the irreducible representation with highest weight λ characterized by the complex number ζ and Θ be the Riemann sphere $\hat{\mathbb{C}}$. The coherent state $\rho_\zeta := |\lambda : \zeta\rangle\langle\lambda : \zeta|$ is a covariant state family with stabilizer H with respect to $\mathrm{SU}(2)$. Then, we have $\mathsf{f}_\lambda(g(\zeta)) = U_{\zeta,t}$.

Similarly, when G is $\mathrm{SU}(\tilde{1}, 1)$, we consider a covariant state family on the irreducible space $\mathcal{U}_\lambda$ with highest weight λ. Then, using the element $F_{-1}^z := \frac{1}{2}\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$ of $\mathfrak{su}(1, 1)$, we choose the subgroup H to be $H = \{\exp(tF_{-1}^z)\}_{t\in\mathbb{R}}$. Let $|\lambda : \zeta\rangle$ be the coherent vector characterized by the complex number ζ and Θ be the unit disk D. The coherent state $\rho_\zeta := |\lambda : \zeta\rangle\langle\lambda : \zeta|$ is a covariant state family with stabilizer H with respect to. Similar to the case of $\mathrm{SU}(2)$, we have $\mathsf{f}_\lambda(g(\zeta)) = U_{\zeta,t}$. For the detail notations of these two examples, see Sect. 4.2 of [44].

When G is the Heisenberg group $H(2, \mathbb{R})$, letting $|\zeta\rangle$ be the coherent vector of the one-mode Bosonic system $L^2(\mathbb{R})$, we find that the set of coherent states $\rho_\zeta := |\zeta\rangle\langle\zeta|$ forms a covariant state family with respect to $H(2, \mathbb{R})$.

In the case of general Lie group G, we consider a covariant state family in a similar way. Firstly, we focus on the irreducible representation space $\mathcal{U}_{\boldsymbol{\lambda}}$ characterized by a highest weight, which is given as a vector $\boldsymbol{\lambda}$. The highest weight vector $|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle$ has the weight $\boldsymbol{\lambda}$. Its orbit by G is the set of coherent states $\rho_\zeta = |\boldsymbol{\lambda} : \zeta\rangle\langle\boldsymbol{\lambda} : \zeta|$, where $|\boldsymbol{\lambda} : \zeta\rangle$ is the coherent vector characterized by the complex-valued vector ζ. Then, the set of coherent state forms a covariant state family with stabilizer $H = \{g \in G | \mathsf{f}(g)|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|\mathsf{f}(g)^\dagger = |\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|\}$. For the detail notations of these two examples, see Sects. 6.7, 6.9, and 6.10 of [44].

Next, under the assumption that the unknown state ρ on the system $\mathcal{H}$ belongs to the above covariant state family $\{\rho_\theta | \theta \in \Theta\}$, we discuss the problem to estimate the unknown parameter θ, as in Fig. 4.1. In this case, we employ a POVM that takes

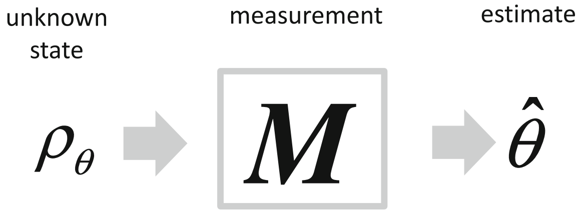


Fig. 4.1 State estimation with $\{\rho_\theta | \theta \in \Theta\}$

values in Θ. We denote the set of such POVMs by $\mathcal{M}(\Theta)$. Since the parameter space Θ has the continuous cardinality, we adopt the conditions (**M1**), (**M2**), and (**M3**) for the definition of our POVM. When we observe our estimate $\hat{\theta}$ different from the true value θ, we employ an **error function** $R(\theta, \hat{\theta})$ to quantify the amount of error. In the following, we seek an estimating method to decrease the expectation $\mathcal{D}_{R,\theta}(M) := \int_\Theta R(\theta, \hat{\theta}) \operatorname{Tr} M(d\hat{\theta})\rho_\theta$ of the error function. In a case, we often seek a method to increase the expectation of a **merit function** that has the opposite sign of the error function. In general, when we increase th expectation of R, R is called an error function, and when we decrease th expectation of R, R is called a merit function.

Since we do not know the true value θ in Θ, we cannot measure the performance of our POVM M by the quantity that depends on the true parameter θ. As a solution, we focus on the worst $\mathcal{D}_R(M) := \max_{\theta\in\Theta} \mathcal{D}_{R,\theta}(M)$ when R is an error function. That is, to seek an optimal POVM M, we minimize the worst value $\mathcal{D}_R(M)$, which is called the mini-max method. On the other hand, when the unknown parameter θ is assumed to be subject to a Bayesian prior distribution ν on Θ, we can minimize the expectation $\mathcal{D}_{R,\nu}(M) := \int_\Theta \mathcal{D}_{R,\theta}(M)\nu(d\hat{\theta})$ of $\mathcal{D}_{R,\theta}(M)$, which is called the Bayesian method. Then, we have

$$\mathcal{D}_R(M) \geq \mathcal{D}_{R,\nu}(M). \tag{4.3}$$

Now, given two POVMs M_0, M_1 and $p \in (0, 1)$, we define the POVM M_p as $M_p(B) := (1 - p)M_0(B) + pM_1(B)$. Then, the relation $\mathcal{D}_{R,\theta}(M_p) = (1 - p)\mathcal{D}_{R,\theta}(M_0) + p\mathcal{D}_{R,\theta}(M_1)$ holds. Taking the expectation with respect to the distribution ν, we have

$$\mathcal{D}_{R,\nu}(M_p) = (1 - p)\mathcal{D}_{R,\nu}(M_0) + p\mathcal{D}_{R,\nu}(M_1). \tag{4.4}$$

Since Θ is a homogeneous space, it is natural to consider that the difference between θ and $\hat{\theta}$ is the same as that between $g\theta$ and $g\hat{\theta}$. Hence, we can naturally assume the following **group invariance** for our error function R:

$$R(\theta, \hat{\theta}) = R(g\theta, g\hat{\theta}). \tag{4.5}$$

On the other hand, when a POVM M on the system $\mathcal{H}$ takes values in Θ and satisfies the condition $M(gB) = \mathsf{f}(g)M(B)\mathsf{f}(g)^\dagger$, it is called **covariant** with respect to the representation f, where $gB := \{g\theta | \theta \in B\}$. This condition is equivalent to the condition $M(d\hat{\theta}') = \mathsf{f}(g)M(d\hat{\theta})\mathsf{f}(g)^\dagger$, where $\hat{\theta}' := g\hat{\theta}$. Then, we denote the set of covariant POVMs taking values in Θ by $\mathcal{M}_{\mathrm{cov}}(\Theta)$, and the set of extremal points of this set by $\mathcal{M}_{\mathrm{cov,ex}}(\Theta)$. When the error function R satisfies the invariance (4.5) and the POVM M is covariant, we have

$$\mathcal{D}_{R,g\theta}(M) = \int_\Theta R(g\theta, \hat{\theta}) \operatorname{Tr} M(d\hat{\theta})\rho_{g\theta} = \int_\Theta R(g\theta, \hat{\theta}) \operatorname{Tr} M(d\hat{\theta})\mathsf{f}(g)\rho_\theta \mathsf{f}(g)^\dagger$$
$$= \int_\Theta R(g\theta, g\hat{\theta}') \operatorname{Tr} M(d\hat{\theta}')\rho_\theta = \mathcal{D}_{R,\theta}(M),$$

where $\hat{\theta}' := g^{-1}\hat{\theta}$. Hence, $\mathcal{D}_{R,\theta}(M) = \mathcal{D}_R(M) = \mathcal{D}_{R,\nu}(M)$. Further, the following theorem holds.

Theorem 4.1 (Quantum Hunt-Stein Theorem [70]) *When Θ is compact,*

$$\min_{M\in\mathcal{M}(\Theta)} \mathcal{D}_R(M) = \min_{M\in\mathcal{M}(\Theta)} \mathcal{D}_{R,\mu_\Theta}(M) = \min_{M\in\mathcal{M}_{\mathrm{cov}}(\Theta)} \mathcal{D}_R(M)$$
$$= \min_{M\in\mathcal{M}_{\mathrm{cov}}(\Theta)} \mathcal{D}_{R,\mu}(M) = \min_{M\in\mathcal{M}_{\mathrm{cov,ex}}(\Theta)} \mathcal{D}_{R,\mu_\Theta}(M),$$

where μ_Θ is the invariance measure. (For the definition of the invariant measure, see [44, Sect. 3.7].) That is, the optimal value of the mini-max method is the same as the optimal value of the Bayesian method with respect to the invariant measure μ_Θ. The optimal value can be attained by the same covariant POVM in the both framework. Further, any POVM attaining the optimal value of the mini-max method is covariant.

Notice that a POVM attaining the optimal value of the Bayesian method is not necessarily covariant. In fact, there exists a non-covariant POVM that attains the optimal value of the Bayesian method.

Proof Due to (4.3), it is sufficient to show that for any POVM M, there exists covariant POVM $\bar{M}$ satisfying the following condition.

$$\mathcal{D}_R(M) \geq \mathcal{D}_{R,\mu_\Theta}(M) = \mathcal{D}_{R,\mu_\Theta}(\bar{M}).$$

For this purpose, for an element $g \in G$, we define the POVM M_g as $M_g(B) := \mathsf{f}(g)^\dagger M(gB)\mathsf{f}(g)$. Then, $\mathcal{D}_{R,\theta}(M_g) = \mathcal{D}_{R,g\theta}(M)$. Taking the integral with respect to θ, we have $\mathcal{D}_{R,\mu_\Theta}(M_g) = \mathcal{D}_{R,\mu_\Theta}(M)$ due to the invariance μ_Θ. Since the POVM $\bar{M} := \int_G M_g \mu_G(dg)$ is covariant, the relation (4.4) yields that $\mathcal{D}_{R,\mu_\Theta}(\bar{M}) = \int_G \mathcal{D}_{R,\mu_\Theta}(M_g)\mu_G(dg) = \mathcal{D}_{R,\mu_\Theta}(M)$, which implies the desired argument. Also, (4.4) guarantees that the above value equals $\min_{M\in\mathcal{M}_{\mathrm{cov,ex}}(\Theta)} \mathcal{D}_{R,\mu_\Theta}(M)$. ■

When Θ is not compact, the full measure of the invariant measure μ_Θ is not finite. Hence, the above theorem does not hold in the non-compact case. In particular, although the above argument for the Bayesian method cannot be extend to the non-compact case, the above argument for the mini-max method can be extend to the non-compact case as follows.

Theorem 4.2 ([11, 104]) *Assume that Θ is locally compact. Any covariant POVM M satisfies $\mathcal{D}_R(M) = \mathcal{D}_{R,\theta}(M)$. Further, we have $\min_{M\in\mathcal{M}(\Theta)} \mathcal{D}_R(M) = \min_{M\in\mathcal{M}_{\mathrm{cov}}(\Theta)} \mathcal{D}_R(M) = \min_{M\in\mathcal{M}_{\mathrm{cov,ex}}(\Theta)} \mathcal{D}_R(M)$. So, the optimal value can be attained by a covariant measurement.*

Remark 4.1 In a general framework of statistical estimation, it is not necessarily required to estimate the whole information for θ to characterize the state. It is often required to estimate a partial information θ. In this case, we set the set of parameter to be estimated to another set Θ_1 and assume an group action of G on Θ_1. Then, we can impose the invariance for the error function $R(\theta, \hat{\theta})$ between the true value $\theta_1 \in \Theta_1$ and the estimate $\hat{\theta}_1 \in \Theta_1$ as

$$R(\theta, \hat{\theta}) = R(g\theta, g\hat{\theta}), \quad g \in G. \tag{4.6}$$

Under the above generalization, we have the same argument as Theorem 4.1.

In such a generalized situation, the sets Θ and Θ_1 have group actions by the group G, they usually are not homogeneous. For example, when an element of Θ_1 is stabilized by the action of the group G, i.e.,

$$g\hat{\theta} = \hat{\theta}, \quad \forall g \in G, \ \forall\hat{\theta} \in \Theta_1, \tag{4.7}$$

the set Θ_1 is not homogeneous unless the set Θ_1 consists of one point. In this example, a covariant POVM $\{M_{\hat{\theta}}\}_{\hat{\theta}\in\Theta_1}$ is given as a POVM satisfying

$$\mathsf{f}(g) M_{\hat{\theta}} \mathsf{f}(g)^\dagger = M_{\hat{\theta}}, \quad \forall g \in G, \ \forall\hat{\theta} \in \Theta_1. \tag{4.8}$$

4.1.2 Characterization of Covariant POVM

We consider how to characterize a covariant POVM when Θ is a homogeneous space. For this purpose, we focus on the element $T := M_{\theta_0}$ of the POVM M that corresponds to the origin $\theta_0 \in \Theta$. Then, we have

$$M_\theta = \mathsf{f}(g(\theta)) T \mathsf{f}(g(\theta))^\dagger. \tag{4.9}$$

So, the matrix T decides the covariant POVM M. Next, we investigate the condition for the matrix T. When a subgroup H of G is the stabilizer of Θ at θ_0, the relation

$$\mathsf{f}(h) T \mathsf{f}(h)^\dagger = T, \quad \forall h \in H \tag{4.10}$$

holds. Since the total integral is the unit matrix, we have

$$\int_\Theta \mathsf{f}(g(\theta)) T \mathsf{f}(g(\theta))^\dagger \mu_\Theta(d\theta) = \int_G \mathsf{f}(g) T \mathsf{f}(g)^\dagger \mu_G(dg) = I. \tag{4.11}$$

That is, a matrix $T \geq 0$ gives a covariant POVM via (4.9) if and only if a matrix T satisfies the conditions (4.10) and (4.11). In the following, a covariant POVM is denoted by M_T when it is generated by a matrix T satisfying the conditions (4.10) and (4.11).

4.1.2.1 Irreducible System

We consider the case when the representation space of our interest is an irreducible space $\mathcal{U}_\lambda$. We denote the set of labels of irreducible unitary representation of G by $\hat{G}$, and denote its element by $\lambda \in \hat{G}$. Then, the corresponding irreducible representation is denoted by f_λ. In particular, when G is a Lie group, the highest weight $\boldsymbol{\lambda}$ uniquely identifies the irreducible representation, which is denoted by $\mathsf{f}_{\boldsymbol{\lambda}}$. In this case, the corresponding element of $\hat{G}$ is also denoted by $\boldsymbol{\lambda}$. However, when the weight is described by a real number like SU(2) and SU(1, 1), we denote the element of $\hat{G}$ by λ.

When the representation of the stabilizer H on the space $\mathcal{U}_\lambda$ is completely reducible and has its irreducible decomposition $\oplus_m \mathcal{H}_m \otimes \mathbb{C}^{n_{\lambda,m}}$, the condition (4.10) for $T \geq 0$ giving a covariant measurement is equivalent to the existence of a positive semi definite matrix T_m on $\mathbb{C}^{n_{\lambda,m}}$ satisfying

$$T = \bigoplus_m I_m \otimes T_m, \tag{4.12}$$

where I_m is the unit matrix on $\mathcal{H}_m$. This fact can be shown by applying Schur's lemma [44, Lemma 2.7] with the completely reducible case to the representation of H.

In the following, we consider the case when G is compact and the space $\mathcal{U}_\lambda$ has finite dimension d_λ. Since μ_G is the normalized invariant measure, the other condition (4.11) of $T \geq 0$ giving a covariant measurement can be replaced by

$$\operatorname{Tr} T = d_\lambda. \tag{4.13}$$

This fact follows from the fact that $\int_G \mathsf{f}_\lambda(g) T \mathsf{f}_\lambda(g)^\dagger \mu_G(dg)$ is commutative with $\mathsf{f}_\lambda(g)$ and is a constant times of the unit matrix I due to Schur's lemma.

On the other hand, when G is unimodular but is not compact, we assume the existence of normalized vectors u, v satisfying

$$d_\lambda^{-1} := \int_G |\langle v|\mathsf{f}_\lambda(g)|u\rangle|^2 \mu_G(dg) < \infty. \tag{4.14}$$

Then, the real number d_λ is called the formal dimension (For the detail, see [44, Sect. 3.7].) and the condition (4.11) is replaced by $\operatorname{Tr} T = d_\lambda$. Indeed, when G is compact and μ_G is the invariant measure, the condition (4.14) is automatically satisfied and the real number d_λ is the dimension of the space $\mathcal{U}_\lambda$. When the stabilizer

H consists of the unit element, the condition (4.12) is trivial, so it suffices to check the condition (4.13).

4.1.2.2 Completely Reducible System

Next, we deal with the case when the representation f_λ of the group G is completely reducible, i.e., the representation space $\mathcal{H}$ has the irreducible decomposition $\oplus_\lambda \mathcal{U}_\lambda \otimes \mathbb{C}^{n_\lambda}$ by dividing it into two cases. For this purpose, we denote the projection to $\mathcal{U}_\lambda \otimes \mathbb{C}^{n_\lambda}$ by P_λ, and assume that each irreducible component $\mathcal{U}_\lambda$ satisfies the condition (4.14). Firstly, we consider the case when the stabilizer H consists of the unit element. Then, the condition (4.10) is a trivial condition, so, it is sufficient to check the other condition (4.11). Hence, the condition (4.11) for $T \geq 0$ giving a covariant POVM is equivalent to the condition

$$\mathrm{Tr}_{\mathcal{U}_\lambda} P_\lambda T P_\lambda = d_\lambda I_{n_\lambda}, \tag{4.15}$$

where I_{n_λ} is the unit matrix on the space $\mathbb{C}^{n_\lambda}$. Secondly, we consider the case when the stabilizer H is a general subgroup. The representation of the subgroup H on the representation space $\mathcal{H}$ has the irreducible decomposition $\oplus_m \mathcal{H}_m \otimes \mathbb{C}^{\sum_\lambda n_\lambda n_{\lambda,m}}$. Then, the condition (4.10) for $T \geq 0$ giving a covariant POVM is equivalent to the existence of a positive semi definite matrix T_m on $\mathbb{C}^{\sum_\lambda n_\lambda n_{\lambda,m}}$ satisfying

$$T = \bigoplus_m I_m \otimes T_m. \tag{4.16}$$

This condition is the same as the condition (4.12). Hence, the condition (4.11) is equivalent to the above given condition (4.15).

In fact, the above conditions can be discussed in terms of the Lie algebra $\mathfrak{g}$ instead of the Lie group G when the Lie group G is connected. Lie algebra $\mathfrak{g}$ often gives a simpler discussion than Lie group G. So, as necessary, we employ Lie algebra $\mathfrak{g}$ to discuss the above conditions.

4.1.3 Optimization of Covariant Measurement

Next, we derive the optimal measurement under a concrete framework with a Lie group G when we have a quantum system whose true state belongs to the state family $\{\rho_\theta\}$, which is often called the input family. For this purpose, we consider how to choose the error function $R(\theta, \hat{\theta})$. To clarify our error function, we focus on another state family $\{\rho_{1,\theta}\}$, which is often called the target family. Our aim is to estimate the state $\rho_{1,\theta}$ belonging to the state family $\{\rho_{1,\theta}\}$ when our system has the true state ρ_θ, as in Fig. 4.2. That is, the input and target states are parametrized by the same parameter θ. Since our interest is to know what density matrix in the state family $\{\rho_{1,\theta}\}$ is true,

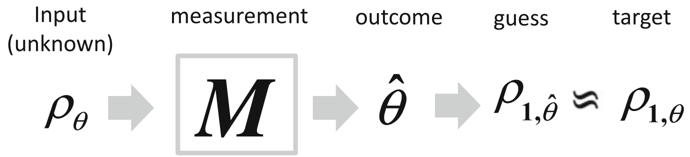


Fig. 4.2 State estimation with $\{\rho_\theta\}$ and $\{\rho_{1,\theta}\}$

the error function $R(\theta,\hat\theta)$ should be given as a function of the state family $\{\rho_{1,\theta}\}$. In the following discussion, the state family $\{\rho_{1,\theta}\}$ giving the error function $R(\theta,\hat\theta)$ is not necessarily the same as the physically prepared state family $\{\rho_\theta\}$.

We often employ the square of Bures distance $1-\mathrm{Tr}\,|\sqrt{\rho_\theta}\sqrt{\rho_{\hat\theta}}|$ as the error function. However, since the calculation of the average error is not easy under the above choice, we sometimes employ $1-\mathrm{Tr}\,\rho_\theta\rho_{\hat\theta}$ instead of the above choice. As another choice, we focus on the ϵ**-error probability**, i.e., the probability that the measurement outcome belongs to the set $\{\hat\theta|\ \mathrm{Tr}\,\rho_\theta\rho_{\hat\theta}\le\epsilon\}$. That is, the error function $R(\theta,\hat\theta)$ takes the value 1 when the outcome $\hat\theta$ belongs to the set, and takes the value 0 otherwise.

In the following, we address the case when the input and target state families $\{\rho_{1,\theta}\}$ and $\{\rho_\theta\}$ are the sets of coherent states of irreducible representations with highest weights $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}$, respectively. We denote the ϵ-error probability with the optimal measurement by $P_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda},\epsilon}$. Then, the optimal measurement is given as follows.

Theorem 4.3 *Assume that the Lie algebra $\mathfrak{g}$ is given as $\mathfrak{su}(2)$ or $\mathfrak{su}(1,1)$ and the system of our interest is given as the representation space $\mathcal{U}_\lambda$. Additionally, we assume that $\lambda<-\frac{1}{2}$ when $\mathfrak{g}=\mathfrak{su}(1,1)$. We give the error function $R(\zeta,\hat\zeta)$ on the coherent state family $\{\rho_\zeta:=|\lambda:\zeta\rangle\langle\lambda:\zeta|\}_\zeta$ as a monotone decreasing function of the fidelity $|\langle\lambda:\zeta|\lambda:\hat\zeta\rangle|^2$. Then, the optimal measurement is given as the covariant POVM decided by $T=d_\lambda|\lambda;\lambda\rangle\langle\lambda;\lambda|$.*

On the other hand, when $\mathfrak{g}$ is $\mathfrak{h}(2,\mathbb{R})$ and the error function $R(\zeta,\hat\zeta)$ of the coherent state family $\{\rho_\zeta:=|\zeta\rangle\langle\zeta|\}_\zeta$ is given as a monotone decreasing function of the fidelity $|\langle\zeta|\hat\zeta\rangle|^2$, the optimal measurement is given as the covariant POVM decided by $T=|0\rangle\langle 0|$, where $|\zeta\rangle$ is the coherent vector in the one-mode Bosonic system.

Proof Firstly, we choose the monotone decreasing function f satisfying that $R(\zeta,\hat\zeta)=\mathsf{f}(|\langle\lambda:\zeta|\lambda:\hat\zeta\rangle|^2)$. Due to Theorem 4.1, it is sufficient to show

$$\begin{aligned}&\int \mathsf{f}(|\langle\lambda;\lambda|\lambda:\hat\zeta\rangle|^2)\langle\lambda;\lambda|M(d\hat\zeta)|\lambda;\lambda\rangle\\ \ge&\int \mathsf{f}(|\langle\lambda;\lambda|\lambda:\hat\zeta\rangle|^2)\langle\lambda;\lambda|M_{d_\lambda|\lambda;\lambda\rangle\langle\lambda;\lambda|}(d\hat\zeta)|\lambda;\lambda\rangle\end{aligned}$$

for any element $M\in\mathcal{M}_{\mathrm{cov,ex}}(\Theta_s)$. Hence, to show the desired argument, it suffices to show the inequality

$$\int_{R_c} \langle \lambda; \lambda | M(d\hat{\zeta}) | \lambda; \lambda \rangle \geq \int_{R_c} \langle \lambda; \lambda | M^{d_\lambda |\lambda; \lambda\rangle\langle\lambda;\lambda|}(d\hat{\zeta}) | \lambda; \lambda \rangle \tag{4.17}$$

for the subset $\mathcal{R}_c := \{\zeta || \langle \lambda : \zeta | \lambda; \lambda \rangle|^2 \geq c\}$ defined by an arbitrary real $c \in [0, 1]$.

In the following, we show the case with $\mathfrak{su}(2)$ or $\mathfrak{su}(1, 1)$. Let $|\lambda; m\rangle$ be the vector with weight m in the representation space $\mathcal{U}_\lambda$. Due to (4.12), it is sufficient to show (4.17) with $M = M_{d_\lambda|\lambda;m\rangle\langle\lambda;m|}$. Since $|\langle \lambda : \zeta | \lambda; m \rangle|^2$ depends only on $x := |\hat{\zeta}|^2$, using the function $h(x) := \frac{|\langle\lambda;m|\lambda:\sqrt{x}\rangle|^2}{|\langle\lambda;\lambda|\lambda:\sqrt{x}\rangle|^2}$ and the real number x_c satisfying $c = (1 + sx_c)^{-2\lambda} = |\langle \lambda : \zeta | \lambda; \lambda \rangle|^2$, we obtain

$$\begin{aligned}
&\int_{\mathcal{R}_c} d_\lambda |\langle \lambda; m | U_{\hat{\zeta},s} | \lambda; \lambda \rangle|^2 \mu_{\Theta_s}(d\hat{\zeta}) = \int_{\mathcal{R}_c} d_\lambda |\langle \lambda; m | \lambda : \hat{\zeta} \rangle|^2 \mu_{\Theta_s}(d\hat{\zeta}) \\
= &\int_0^{x_c} d_\lambda |\langle \lambda; m | \lambda : \sqrt{x} \rangle|^2 \frac{da}{(1+sx)^2} = \int_0^{x_c} h(x) d_\lambda |\langle \lambda; \lambda | \lambda : \sqrt{x} \rangle|^2 \frac{dx}{(1+sx)^2}.
\end{aligned}$$

Hence, it is enough to show

$$\int_0^{x_c} d_\lambda |\langle \lambda; \lambda | \lambda : \sqrt{x} \rangle|^2 \frac{dx}{(1+sx)^2} \geq \int_0^{x_c} h(x) d_\lambda |\langle \lambda; \lambda | \lambda : \sqrt{x} \rangle|^2 \frac{dx}{(1+sx)^2}. \tag{4.18}$$

Since

$$\begin{aligned}
h(x) = &\frac{|\langle \lambda; m | \lambda : \sqrt{x} \rangle|^2}{|\langle \lambda; \lambda | \lambda : \sqrt{x} \rangle|^2} = |\langle \lambda; m | \frac{(\sqrt{x}\mathsf{K}_{-,s})^{\lambda - m}}{(\lambda - m)!} | \lambda; \lambda \rangle|^2 \\
= &x^{\lambda - m} |\langle \lambda; m | \frac{(\mathsf{K}_{-,s})^{\lambda - m}}{(\lambda - m)!} | \lambda; \lambda \rangle|^2,
\end{aligned}$$

$h(x)$ is monotone increasing with respect to x. Thus, we obtain (4.18) when $h(x_c) \leq 1$.

On the other hand, even when $h(x_c) > 1$, using

$$x_m := \begin{cases} \infty & \text{when } s = 0, 1 \\ 1 & \text{when } s = -1, \end{cases}$$

we obtain

$$\int_{x_c}^{x_m} d_\lambda |\langle \lambda; \lambda | \lambda : \sqrt{x} \rangle|^2 \frac{dx}{(1+sx)^2} \leq \int_{x_c}^{x_m} h(x) d_\lambda |\langle \lambda; \lambda | \lambda : \sqrt{x} \rangle|^2 \frac{dx}{(1+sx)^2}.$$

Therefore, since the total integral $\int_0^\infty d_\lambda |\langle \lambda; \lambda | \lambda : \sqrt{x} \rangle|^2 \frac{dx}{(1+sx)^2}$ is 1, we obtain (4.18).

On the other hand, when the Lie algebra $\mathfrak{g}$ is Lie algebra $\mathfrak{h}(2, \mathbb{R})$ of Heisenberg group, we assume that the highest weight is (λ_Z, λ_F) and choose x_c as $c = e^{x_c} =$

$|\langle\sqrt{\lambda_Z}\zeta|0\rangle|^2$. Since the weight vector is the number vector $|m\rangle$, replacing the vector $|\lambda; m\rangle$ by the vector $|m\rangle$ in the above proof, we define the function $h(x)$ as

$$h(x) := \frac{|\langle m|\sqrt{\lambda_Z}\sqrt{x}\rangle|^2}{|\langle 0|\sqrt{\lambda_Z}\sqrt{x}\rangle|^2} = |\langle m|\frac{(\sqrt{\lambda_Z}\sqrt{x}\mathsf{a}^\dagger)^m}{m!}|0\rangle|^2 = x^m\lambda_Z^m|\langle m|\frac{(\mathsf{a}^\dagger)^m}{m!}|0\rangle|^2$$

Since $h(x)$ is monotone increasing with respect to x, we can show the desired argument for Lie algebra $\mathfrak{h}(2,\mathbb{R})$ in the same way as the above. ■

Example 4.1 Let us calculate the ϵ-error probability $P_{\lambda_1,\lambda,\epsilon}$ when the optimal measurement is applied to the case with $\mathfrak{g} = \mathfrak{su}(2), \mathfrak{su}(1,1)$. When $\lambda < \frac{-1}{2}$ or a positive half integer, putting $\zeta = re^{i\theta}$ and $x = (1 + sr^2)$ and choose r_0 satisfying $(1 + sr_0^2)^{-2\lambda_1} = \epsilon$, we have

$$\begin{aligned} P_{\lambda_1,\lambda,\epsilon} &= s(2\lambda+1)\int_{\{(1+s|\zeta|^2)^{-2\lambda_1}\le\epsilon\}}(1+s|\zeta|^2)^{-2\lambda}\mu_{\Theta_s}(d\zeta) \\ &= (2\lambda+1)\int_{r_0}^{a}(1+sr^2)^{-2\lambda}\frac{2\pi s r dr}{\pi(1+sr^2)^2} \\ &= (2\lambda+1)\int_{\epsilon^{\frac{-1}{2\lambda_1}}}^{b} x^{-2\lambda-2}dx = -[x^{-2\lambda-1}]_{\epsilon^{\frac{-1}{2\lambda_1}}}^{b} = \epsilon^{\frac{2\lambda+1}{2\lambda_1}}, \end{aligned}$$

where $a = \infty$ and $b = \infty$ when $s = 1$, and $a = 1$ and $b = 0$ when $s = -1$. In particular, when $\lambda = n\lambda_1$, the ϵ-error probability $P_{\lambda_1,\lambda,\epsilon}$ is $\epsilon^{n+\frac{1}{2\lambda_1}}$, which exponentially goes to zero as n increases.

Example 4.2 Let us calculate the ϵ-error probability $P_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda},\epsilon}$ when the optimal measurement is applied to the case with $\mathfrak{g} = \mathfrak{h}(2,\mathbb{R}) \rtimes \mathfrak{u}(1)$. Putting $\boldsymbol{\lambda}_1 = (\lambda_{1,Z}, \lambda_{1,F})$, $\boldsymbol{\lambda} = (\lambda_Z, \lambda_F)$, $\zeta = re^{i\theta}$, and choose r_0 satisfying $e^{-\lambda_{1,Z}r_0^2} = \epsilon$, we have

$$\begin{aligned} P_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda},\epsilon} &= \lambda_Z\int_{\{e^{-\lambda_{1,Z}|\zeta|^2}\le\epsilon\}} e^{-\lambda_Z|\zeta|^2}\mu_{\mathbb{C}}(d\zeta) = \lambda_Z\int_{r_0}^{\infty} e^{-\lambda_Z r^2}\frac{2\pi r dr}{\pi} \\ &= -[e^{\lambda_Z r^2}]_{r_0}^{\infty} = \epsilon^{\frac{\lambda_Z}{\lambda_{1,Z}}}. \end{aligned}$$

In particular, when $\boldsymbol{\lambda} = n\boldsymbol{\lambda}_1$, the ϵ-error probability $P_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda},\epsilon}$ is ϵ^n, which exponentially goes to zero as n increases.

Theorem 4.3 can be generalized as follows.

Theorem 4.4 *Given a Lie algebra $\mathfrak{g}$, we focus on the skew-Hermitian representation $\mathsf{f}_{\boldsymbol{\lambda}}$ on the irreducible space $\mathcal{U}_{\boldsymbol{\lambda}}$ with highest weight $\boldsymbol{\lambda}$. Then, the coherent state family $\{\rho_\zeta := |\boldsymbol{\lambda} : \zeta\rangle\langle\boldsymbol{\lambda} : \zeta|\}_\zeta$ is given as the orbit of the origin $|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|$. Let $\mathfrak{h}_1$ be the Lie algebra corresponding to the invariant subgroup H_1 that stabilizes the origin $|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|$. (For the definition of the coherent state of a general Lie algebra, see Sects. 6.7, 6.9, and 6.10 in [44].) Then, $\mathfrak{g}$ has the direct sum decomposition $\mathfrak{h}_1 \oplus \mathfrak{p}$ and*

$\mathfrak{p}$ is closed with respect to the natural action by $\mathfrak{h}_1$. We assume that $\mathfrak{h}_1$ is $\mathfrak{su}(d)$ or $\mathfrak{u}(d)$ and that the representation of $\mathfrak{h}_1$ on $\mathfrak{p}$ is complexifiable and is isomorphic to the natural action of $\mathfrak{su}(d)$ or $\mathfrak{u}(d)$ on $\mathbb{C}^d$. Then, the coherent state family $\{\rho_\zeta := |\boldsymbol{\lambda} : \zeta\rangle\langle\boldsymbol{\lambda} : \zeta|\}_\zeta$ is parametrized by an element of $\mathbb{C}^d$. When the error function $R(\zeta, \hat{\zeta})$ is given as a monotone decreasing function of the fidelity $|\langle\boldsymbol{\lambda} : \zeta|\boldsymbol{\lambda} : \hat{\zeta}\rangle|^2$, an optimal measurement is given as a covariant POVM $M_{d_{\boldsymbol{\lambda}}|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}|}$. Here, we need to assume the existence of the formal dimension $d_{\boldsymbol{\lambda}}$ when $\mathfrak{g}$ is not compact.

When the Lie group G is $\mathrm{SU}(d)$ and the highest weight is $\boldsymbol{\lambda} = [m, 0, \ldots, 0]$, the Lie algebra $\mathfrak{h}_1$ of the stabilizer of the the highest weight vector is $\mathfrak{u}(d-1)$ and the condition of Theorem 4.4 holds [46]. As another example, when the Lie group G is the semi direct product of Heisenberg group and the unitary group $\mathrm{H}(2r, \mathbb{R}) \rtimes \mathrm{U}(r)$ and the representation is the Heisenberg representation on $L^2(\mathbb{R}^r)$ and the Lie algebra $\mathfrak{h}_1$ of the stabilizer of the vacuum state, the condition of Theorem 4.4 holds [46]. (See Sect. 7.3 of [44].)

Proof Firstly, we choose the monotone decreasing function f is given as $R(\zeta, \hat{\zeta}) = f(|\langle\boldsymbol{\lambda} : \zeta|\boldsymbol{\lambda} : \hat{\zeta}\rangle|^2)$. Due to Theorem 4.1, it is sufficient to show that

$$\begin{aligned}&\int \mathsf{f}(|\langle\boldsymbol{\lambda} : \zeta|\boldsymbol{\lambda} : \hat{\zeta}\rangle|^2)\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|M_T(d\hat{\zeta})|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle \\ \geq &\int \mathsf{f}(|\langle\boldsymbol{\lambda} : \zeta|\boldsymbol{\lambda} : \hat{\zeta}\rangle|^2)\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|M^{d_{\boldsymbol{\lambda}}|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}|}(d\hat{\zeta})|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle\end{aligned}$$

for any $M \in \mathcal{M}_{\mathrm{cov,ex}}(\Theta)$. So, due to (4.12), it is sufficient to show (4.17) to the case when T is $d_{\boldsymbol{\lambda}}$ times of the one-dimensional projection to the eigenspace with weight m. Here, using the unitary U_ζ defined in the sentence after (6.38), we have

$$\begin{aligned}\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|M_{d_{\boldsymbol{\lambda}}|v_m\rangle\langle v_m|}(d\hat{\zeta})|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle &= d_{\boldsymbol{\lambda}}|\langle v_m|U_{-\hat{\zeta}}|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle|^2\mu_\Theta(d\hat{\zeta}) \\ &= d_{\boldsymbol{\lambda}}|\langle v_m|\boldsymbol{\lambda} : -\hat{\zeta}\rangle|^2\mu_\Theta(d\hat{\zeta}).\end{aligned}$$

Let μ_0 be the normalized invariant measure on the set S of unit vectors so that the full measure of μ_0 is 1. Choosing a function $p(x)$ for $x = \|\zeta\|^2$, we can deform the integral $\mu_\Theta(d\zeta)$ to the integral $\mu_0(d\zeta_0)p(x)dx$.

Given an arbitrary real number $c \in [0, 1]$, we have a unique constant $a_c > 0$ satisfying $\phi(x) := |\langle\boldsymbol{\lambda} : \sqrt{x_{c,\zeta_0}}\zeta_0|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle|^2 = c$ independently of ζ_0 because $\{\rho_{\sqrt{c}\zeta_0}\}_\zeta$ is closed with respect to the action of the invariant subgroup H_1. Hence, it is sufficient to show the inequality

$$\begin{aligned}&\int_0^{x_c}\int_S d_{\boldsymbol{\lambda}}|\langle\boldsymbol{\lambda}; \boldsymbol{\lambda}|\boldsymbol{\lambda} : \sqrt{x}\zeta_0\rangle|^2\mu_0(d\zeta_0)p(x)dx = \int_0^{x_c} d_{\boldsymbol{\lambda}}\phi(x)p(x)dx \\ \geq &\int_0^{x_c}\int_S d_{\boldsymbol{\lambda}}|\langle v_m|\boldsymbol{\lambda} : \sqrt{x}\zeta_0\rangle|^2\mu_0(d\zeta_0)p(x)dx.\end{aligned}$$

Using $h(x) := \frac{\int_S |\langle v_m|\boldsymbol{\lambda}{:}\sqrt{x}\zeta_0\rangle|^2\mu_0(d\zeta_0)}{\int_S |\langle \boldsymbol{\lambda};\boldsymbol{\lambda}|\boldsymbol{\lambda}{:}\sqrt{x}\zeta_0\rangle|^2\mu_0(d\zeta_0)} = \frac{\int_S |\langle v_m|\boldsymbol{\lambda}{:}\sqrt{x}\zeta_0\rangle|^2\mu_0(d\zeta_0)}{\phi(x)}$, we have

$$\int_0^{x_c}\int_S d_{\boldsymbol{\lambda}}|\langle v_m|\boldsymbol{\lambda}, \sqrt{x}\zeta_0\rangle|^2\mu_0(d\zeta_0)p(x)dx = \int_0^{x_c} d_{\boldsymbol{\lambda}}h(x)\phi(x)p(x)dx.$$

Since $|\langle \boldsymbol{\lambda}; \boldsymbol{\lambda}|\boldsymbol{\lambda} : \sqrt{x}\zeta_0\rangle|^2 = \phi(x)$, there exists an integer n_0 such that

$$\begin{aligned}
h(x) =& \int_S \frac{|\langle v_m|\boldsymbol{\lambda} : \sqrt{x}\zeta_0\rangle|^2}{|\langle \boldsymbol{\lambda}; \boldsymbol{\lambda}|\boldsymbol{\lambda} : \sqrt{x}\zeta_0\rangle|^2}\mu_0(d\zeta_0) \\
=& \int_S |\langle v_m|\sum_{n=0}^{\infty}\frac{(\sqrt{x}\sum_{\alpha\in\Phi_{\boldsymbol{\lambda}}^+}\zeta_{0,\alpha}\mathsf{K}_{-\alpha})^n}{n!}|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle|^2\mu_0(d\zeta_0) \\
=& x^{n_0}\int_S |\langle v_m|\frac{(\sum_{\alpha\in\Phi_{\boldsymbol{\lambda}}^+}\zeta_{0,\alpha}\mathsf{K}_{-\alpha})^{n_0}}{n_0!}|\boldsymbol{\lambda}; \boldsymbol{\lambda}\rangle|^2\mu_0(d\zeta_0).
\end{aligned}$$

So, $h(x)$ is monotone increasing for x. Here, $\Phi_{\boldsymbol{\lambda}}^+$ is defined as the subset of $\{\boldsymbol{\alpha} \in \Phi^+|(\boldsymbol{\alpha}, \boldsymbol{\lambda}) > 0\}$ the set Φ^+ of positive roots. Therefore, the desired argument can be shown in the same way as the proof of Theorem 4.3. ■

Exercise 4.1 Calculate $\max_M \int |\langle\lambda_1; \lambda_1|\lambda_1 : \hat{\zeta}\rangle|^2\langle\lambda; \lambda|M(d\hat{\zeta})|\lambda; \lambda\rangle$, where the maximum is taken among covariant measurement and $\mathfrak{g}$ is $\mathfrak{su}(2)$ or $\mathfrak{su}(1, 1)$.

4.2 Generation of Approximating State

4.2.1 General Theory for Generation of Approximating State

In this section, we focus on two covariant state families ρ_θ and $\rho_{1,\theta}$ with respect to the representations of the same group G on two systems $\mathcal{H}$ and $\mathcal{H}_1$, respectively. Then, we apply a measurement corresponding to the POVM $M = \{M_\omega\}$ on the system $\mathcal{H}$, and prepare the input state ρ_ω according to the measurement outcome ω so that the target state ρ_ω is close to $\rho_{1,\theta}$, as in Fig. 4.3. We denote this operation by $\hat{M} := (M, \{\rho_\omega\})$, and call it **generation of approximating state** or measurement and state preparation (MSP) [108].

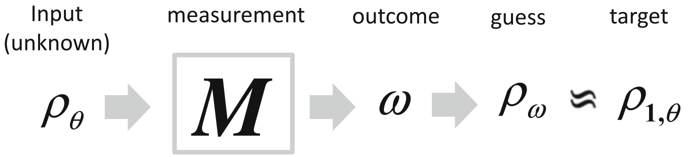


Fig. 4.3 Generation of approximating state (measurement and state preparation)

Given a function $f(\rho_{1,\theta},\rho_\omega)$ that measures the error of the approximating state ρ_ω to the target state $\rho_{1,\theta}$, the performance of the operation $\hat{M}$ is evaluated by $\mathcal{D}_{f,\theta}(\hat{M}) := \int_\Omega f(\rho_{1,\theta},\rho_\omega)\operatorname{Tr}\rho_\theta M_\omega d\omega$. Similar to the discussion in the previous section, we can minimize the quantities $\mathcal{D}_f(\hat{M}) := \max_{\theta\in\Theta}\mathcal{D}_{f,\theta}(\hat{M})$ and $\mathcal{D}_{f,\nu}(\hat{M}) := \int_\Theta \mathcal{D}_{f,\theta}(\hat{M})\nu(d\theta)$.

In the following, when f and f_1 are unitary representations of a connected Lie group G on the systems $\mathcal{H}$ and $\mathcal{H}_1$, respectively, ρ_θ and $\rho_{1,\theta}$ are assumed to be covariant state families of the parameter space $\Theta = G/H$ with respect to the representations f and f_1, respectively. We assume that the Lie group G is compact or satisfies the condition (4.14). We say that an MSP $\hat{M} := (M,\{\rho_\omega\})$ is covariant when the following conditions hold;

(1) M is a covariant POVM taking values in G, i.e., the relation $M_{g\omega} = \mathsf{f}(g)M_\omega\mathsf{f}(g)^\dagger$ holds for ω and $g\in G$.
(2) The relation $\rho_{1,g\omega} = \mathsf{f}_1(g)\rho_{1,\omega}\mathsf{f}_1(g)^\dagger$ holds for ω and $g\in G$.

Then, any covariant MSP is given by using a state ρ on $\mathcal{H}$ and a state ρ_1 on $\mathcal{H}_1$ as follows

$$M(dg) = d_\lambda \mathsf{f}(g)\rho\mathsf{f}(g)^\dagger \mu_G(dg), \quad \rho_{1,g} = \mathsf{f}_1(g)\rho_1\mathsf{f}_1(g)^\dagger.$$

In the following, we denote the above covariant MSP by $\hat{M}_{\rho,\rho_1}$. Even when the set of measurement outcomes of a MSP $\hat{M} := (M,\{\rho_\omega\})$ is not the group G itself but is its homogeneous space G/H, if M and $\{\rho_\omega\}$ and covariant, the MSP can be treated as the covariant MSP $\hat{M}' := (M',\{\rho'_\omega\})$ whose set of measurement outcomes is the group G as $M'(g) = cM([g])$ and $\rho'_g = \rho_{[g]}$ with a normalizing constant c. In the following discussion, the error function $f(\rho_{1,\theta},\rho_\omega)$ measuring the error of the approximation is assumed to satisfy the following invariance under the representation f;

$$f(\rho_{1,\theta},\rho_\omega) = f(\mathsf{f}(g)\rho_{1,\theta}\mathsf{f}(g)^\dagger, \mathsf{f}(g)\rho_\omega\mathsf{f}(g)^\dagger), \quad \forall g\in G. \tag{4.19}$$

Then, similar to the discussion in the previous section, when the Lie group G is compact, we can assume that the invariant measure μ_Θ has the full measure 1. Then, we have the following theorem, which is similar to Theorem 4.1.

Theorem 4.5 *When the Lie group G is compact and the function f measuring the error of the approximation satisfies (4.19), the relations*

$$\min_{\hat{M}}\mathcal{D}_f(\hat{M}) = \min_{\rho,\rho_1}\mathcal{D}_f(\hat{M}_{\rho,\rho_1}) = \min_{\hat{M}}\mathcal{D}_{f,\mu_\Theta}(\hat{M}) = \min_{\rho,\rho_1}\mathcal{D}_{f,\mu_\Theta}(\hat{M}_{\rho,\rho_1})$$

hold. Even though G is not compact, when the other condition holds, the first equation holds.

Proof We assume that the POVM M is written as $M_\omega\nu(d\omega)$ by using a probability measure $\nu(d\omega)$. Then, we have

$$
\begin{aligned}
&\int_G \mathcal{D}_{f,g\theta_0}(\hat{M})\mu(dg) = \int_G \int_\Omega f(\rho_{1,g\theta_0}, \rho_\omega) \operatorname{Tr} \rho_{g\theta} M_\omega \nu(d\omega)\mu(dg) \\
= &\int_G \int_\Omega f(\rho_{1,\theta_0}, \mathsf{f}_1(g)^\dagger \rho_\omega \mathsf{f}_1(g)) \operatorname{Tr} \rho_\theta \mathsf{f}(g)^\dagger M_\omega \mathsf{f}(g) \nu(d\omega)\mu(dg) \\
= &\int_\Omega \frac{\operatorname{Tr} M_\omega}{d_\lambda} d_\lambda \int_G f(\rho_{1,\theta_0}, \mathsf{f}_1(g)^\dagger \rho_\omega \mathsf{f}_1(g)) \operatorname{Tr} \rho_\theta \mathsf{f}(g)^\dagger \frac{M_\omega}{\operatorname{Tr} M_\omega} \mathsf{f}(g)\mu(dg)\nu(d\omega) \\
= &\int_\Omega \mathcal{D}_f(\hat{M}_{\frac{M_\omega}{\operatorname{Tr} M_\omega}, \rho_\omega}) \frac{\operatorname{Tr} M_\omega}{d_\lambda} \nu(d\omega).
\end{aligned}
$$

Since $\frac{\operatorname{Tr} M_\omega}{d_\lambda}\nu(d\omega)$ is a probability distribution on Ω, we have

$$
\min_{\hat{M}} \mathcal{D}_{f,\mu_\Theta}(\hat{M}) \geq \min_{\rho,\rho_1} \mathcal{D}_{f,\mu_\Theta}(\hat{M}_{\rho,\rho_1}) = \min_{\rho,\rho_1} \mathcal{D}_f(\hat{M}_{\rho,\rho_1}).
$$

Since the opposite inequality

$$
\min_{\rho,\rho_1} \mathcal{D}_f(\hat{M}_{\rho,\rho_1}) \geq \min_{\hat{M}} \mathcal{D}_f(\hat{M}) \geq \min_{\hat{M}} \mathcal{D}_{f,\mu_\Theta}(\hat{M})
$$

holds, we obtain the desired argument. ■

4.2.2 General Theory for Generation of Approximating State with Puseudo Fidelity

It is not easy to treat a general function $f(\rho_{1,\theta_0}, \rho_\omega)$ expressing the error of the approximation. In the following, we maximize the average of puseudo fidelity $F(\rho_{1,\theta_0}, \rho_\omega) = \operatorname{Tr} \rho_{1,\theta_0}\rho_\omega$. Given an arbitrary irreducible representation space $\mathcal{H}$ in the representation space $\mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathcal{U}_{\boldsymbol{\lambda}_1}$ of the Lie group G and the input and target states ρ_{θ_0} and ρ_{1,θ_0}, we define $c(\mathcal{H}) := \frac{\operatorname{Tr} P(\mathcal{H})\rho_{\theta_0}\otimes\rho_{1,\theta_0}}{\dim \mathcal{H}}$, where $P(\mathcal{H})$ is the projection to $\mathcal{H}$. Then, the highest weight of the irreducible subrepresentation $\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}$ is maximum among irreducible subrepresentations in $\mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathcal{U}_{\boldsymbol{\lambda}_1}$ and has multiplicity 1. Under these notations, we have the following theorem.

Theorem 4.6 *When the relation*

$$
c_0 := c(\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}) \geq c(\mathcal{H}) \tag{4.20}
$$

holds for any irreducible representation space $\mathcal{H}$ of the representation space $\mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathcal{U}_{\boldsymbol{\lambda}_1}$, the relation

$$
\max_{\rho,\rho_1} \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) = \mathcal{D}_F(\hat{M}_{|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}|,|\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1\rangle\langle\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1|}) = d_{\boldsymbol{\lambda}} c_0
$$

holds. That is, the optimal measurement is the covariant measurement generated by $|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}|$ *and the guess is* $|\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1\rangle\langle\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1|$ *when the observed measurement outcome corresponds to the POVM element* $|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}|$.

When G is not compact, $\dim\mathcal{H}$ *and* $d_{\boldsymbol{\lambda}}$ *are the formal dimensions.*

Proof When our MSP is characterized by ρ and ρ_1, $\mathcal{D}_{F,\theta}(\hat{M})$ is calculated as

$$
\begin{aligned}
&\mathcal{D}_{F,\theta_0}(\hat{M}_{\rho,\rho_1})\\
=&\int_G (\mathrm{Tr}_{\mathcal{H}_1}\,\rho_{1,\theta_0}\mathsf{f}_{\boldsymbol{\lambda}_1}(g)\rho_1\mathsf{f}_{\boldsymbol{\lambda}_1}(g)^\dagger)(\mathrm{Tr}_{\mathcal{H}}\,\rho_{\theta_0}d_{\boldsymbol{\lambda}}\mathsf{f}_{\boldsymbol{\lambda}}(g)\rho\mathsf{f}_{\boldsymbol{\lambda}}(g)^\dagger)\mu_G(dg)\\
=&d_{\boldsymbol{\lambda}}\int_G \mathrm{Tr}_{\mathcal{H}_1\otimes\mathcal{H}}(\rho_{1,\theta_0}\otimes\rho_{\theta_0})(\mathsf{f}_{\boldsymbol{\lambda}_1}(g)\otimes\mathsf{f}_{\boldsymbol{\lambda}}(g))(\rho_1\otimes\rho)(\mathsf{f}_{\boldsymbol{\lambda}_1}(g)\otimes\mathsf{f}_{\boldsymbol{\lambda}}(g))^\dagger\mu_G(dg)\\
=&d_{\boldsymbol{\lambda}}\,\mathrm{Tr}_{\mathcal{H}_1\otimes\mathcal{H}}[\int_G(\mathsf{f}_{\boldsymbol{\lambda}_1}(g)\otimes\mathsf{f}_{\boldsymbol{\lambda}}(g))^\dagger(\rho_{1,\theta_0}\otimes\rho_{\theta_0})(\mathsf{f}_{\boldsymbol{\lambda}_1}(g)\otimes\mathsf{f}_{\boldsymbol{\lambda}}(g))\mu_G(dg)](\rho_1\otimes\rho).
\end{aligned}
$$

Due to Schur's lemma, there exist irreducible subspaces $\mathcal{H}_1,\ldots\mathcal{H}_k$ such that the integral $[\int_G(\mathsf{f}_{\boldsymbol{\lambda}_1}(g)\otimes\mathsf{f}_{\boldsymbol{\lambda}}(g))^\dagger(\rho_{1,\theta_0}\otimes\rho_{\theta_0})(\mathsf{f}_{\boldsymbol{\lambda}_1}(g)\otimes\mathsf{f}_{\boldsymbol{\lambda}}(g))\mu_G(dg)]$ is written as $\sum_{j=1}^k c(\mathcal{H}_j)P(\mathcal{H}_j)$. Hence, due to the condition (4.20), the quantity $d_{\boldsymbol{\lambda}}\,\mathrm{Tr}_{\mathcal{H}_1\otimes\mathcal{H}}\sum_{j=1}^k c(\mathcal{H}_j)P(\mathcal{H}_j)(\rho_1\otimes\rho)$ takes the maximum value $d_{\boldsymbol{\lambda}}c(\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1})$ when $\rho_1\otimes\rho$ is included in $\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}$. For example, when $\rho_1\otimes\rho$ is equal to $|\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1\rangle\langle\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1|\otimes|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}| = |\boldsymbol{\lambda}_1+\boldsymbol{\lambda};\boldsymbol{\lambda}_1+\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda}_1+\boldsymbol{\lambda};\boldsymbol{\lambda}_1+\boldsymbol{\lambda}|$, the maximum value $d_{\boldsymbol{\lambda}}c(\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1})$ is realized. ■

The following lemma holds for the condition (4.20).

Lemma 4.1 *Given two Lie groups* $G'\subset G$ *and two irreducible representations* f *and* f_1 *of* G*, we assume that their subrepresentations* $\mathsf{f}_{\boldsymbol{\lambda}}$ *and* $\mathsf{f}_{\boldsymbol{\lambda}_1}$ *of* G' *satisfy the condition (4.20). When the representation space* $\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}$ *of* G' *whose highest weight is maximal is closed with respect to the representation of G, the condition (4.20) holds even for G.*

There is a relation between the generation of approximating state and the state estimation discussed in the previous section as follows. When an error function f for the state estimation is extended to a function defined for two arbitrary density matrices, the generation of approximating state with the error function has better performance than the state estimation with the error function. Then, even when a measurement achieves the optimal performance in the state estimation, it does not necessarily attain the optimal performance in the generation of approximating state in general because we can use a state that does not belong to the parametric state family as our guess in the generation of approximating state. Theorem 4.6 shows that the measurement $\hat{M}_{|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}|,|\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1\rangle\langle\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1|}$ is still optimal even for the generation of approximating state when we employ the function $1-F(\rho_{1,\theta},\rho_\omega)$ as the error function.

Since the obtained optimal measurement has continuous measurement outcome, its realization is not so easy. Next, we consider how to realize the same performance with discrete measurement outcome. For this issue, a measure ν on Θ is called a $\boldsymbol{\lambda}$-design when

$$d_{\boldsymbol{\lambda}} \int_{\Theta} |\boldsymbol{\lambda} : \theta\rangle\langle\boldsymbol{\lambda} : \theta| \nu(d\theta) = I_{\boldsymbol{\lambda}}. \tag{4.21}$$

Here, in the infinite-dimensional representation, we assume the existence of the formal dimension. Further, it is called an ϵ approximate $\boldsymbol{\lambda}$-design with matrix norm when

$$\left\| d_{\boldsymbol{\lambda}} \int_{\Theta'} |\boldsymbol{\lambda} : \theta\rangle\langle\boldsymbol{\lambda} : \theta| \nu(d\theta) - I_{\boldsymbol{\lambda}} \right\| \leq \epsilon \tag{4.22}$$

For its detail, see [44, Sect. 4.5.2]. Then, we have the following theorem.

Theorem 4.7 *Consider the case when $\mathcal{H}_1 = \mathcal{U}_{\boldsymbol{\lambda}_1}$ and $\mathcal{H} = \mathcal{U}_{\boldsymbol{\lambda}}$. Given a $\boldsymbol{\lambda}$-design measure ν, we can define the measurement $M_\nu(d\hat{\theta}) := d_{\boldsymbol{\lambda}} |\boldsymbol{\lambda} : \hat{\theta}\rangle\langle\boldsymbol{\lambda} : \hat{\theta}|\nu(d\hat{\theta})$. When the measure ν is $\boldsymbol{\lambda} + \boldsymbol{\lambda}_1$-design measure, we can define the operation $\hat{M}_\nu$ that prepares the state $|\boldsymbol{\lambda}_1 : \hat{\theta}\rangle\langle\boldsymbol{\lambda}_1 : \hat{\theta}|$ by following the measurement outcome $\hat{\theta}$ of the measurement M_ν. Then, the relation $\mathcal{D}_{F,\theta}(\hat{M}_\nu) = d_{\boldsymbol{\lambda}} c_0$ holds. In particular, when the condition (4.20) holds, the operation $\hat{M}_\nu$ attains the optimal performance of the generation of approximating state.*

Proof Similar to the first relation of Theorem 4.6, we have the following relation;

$$\mathcal{D}_{F,\theta}(\hat{M}_\nu) = \mathrm{Tr}(\rho_{1,\theta_0} \otimes \rho_{\theta_0}) \int_{\Theta'} d_{\boldsymbol{\lambda}} |\boldsymbol{\lambda} + \boldsymbol{\lambda}_1 : \hat{\theta}\rangle\langle\boldsymbol{\lambda} + \boldsymbol{\lambda}_1 : \hat{\theta}|\nu(d\hat{\theta})$$
$$= \mathrm{Tr}(\rho_{1,\theta_0} \otimes \rho_{\theta_0}) I_{\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}} \frac{d_{\boldsymbol{\lambda}}}{d_{\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}}} = d_{\boldsymbol{\lambda}} c_0.$$

■

Given an ϵ approximate $\boldsymbol{\lambda}$-design measure p with matrix norm, using $Y_{\boldsymbol{\lambda},p} := \int_{\Theta} d_{\boldsymbol{\lambda}} |\boldsymbol{\lambda} : \hat{\theta}\rangle\langle\boldsymbol{\lambda} : \hat{\theta}|p(d\hat{\theta})$, we define the POVM $M_p(d\hat{\theta}) := Y_{\boldsymbol{\lambda},p}^{-1/2} d_{\boldsymbol{\lambda}} |\boldsymbol{\lambda} : \hat{\theta}\rangle\langle\boldsymbol{\lambda} : \hat{\theta}| Y_{\boldsymbol{\lambda},p}^{-1/2}$. Then, the following theorem holds.

Theorem 4.8 *When p is an ϵ approximate $\boldsymbol{\lambda}$-design measure with matrix norm and is an ϵ_1 approximate $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}$-design measure with matrix norm, the measurement M_p and the operation $\hat{M}_p$ are defined in the same way as Theorem 4.7 and satisfy $|\mathcal{D}_{F,\theta}(\hat{M}_p) - d_{\boldsymbol{\lambda}} c_0| \leq (\epsilon_1 + 2\epsilon) d_{\boldsymbol{\lambda}} c_0$.*

Proof Similar to Theorem 4.6, we can show

$$\mathcal{D}_{F,\theta}(\hat{M}_p) = \frac{d_{\boldsymbol{\lambda}}}{d_{\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}}} \mathrm{Tr}(\rho_{1,\theta_0} \otimes \rho_{\theta_0})(I \otimes Y_{\boldsymbol{\lambda},p}^{-1/2}) Y_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1,p} (I \otimes Y_{\boldsymbol{\lambda},p}^{-1/2}).$$

Since $(1 - (\epsilon + \epsilon_1))I \leq (I \otimes Y_{\boldsymbol{\lambda},p}^{-1/2}) Y_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1,p} (I \otimes Y_{\boldsymbol{\lambda},p}^{-1/2}) \leq (1 + (2\epsilon + \epsilon_1))I$, we obtain the desired argument. ■

4.2.3 Generation of Approximating State for Pure States

Next, we consider the case when the state families $\{\rho_{\theta_0}\}$ and $\{\rho_{1,\theta_0}\}$ are coherent state families as follows.

$$\begin{aligned}\rho_{\theta_0} &= \mathsf{f}_{\boldsymbol{\lambda}}(g(\theta))|\boldsymbol{\lambda};\boldsymbol{\lambda}\rangle\langle\boldsymbol{\lambda};\boldsymbol{\lambda}|\mathsf{f}_{\boldsymbol{\lambda}}(g(\theta))^{\dagger}\\ \rho_{1,\theta_0} &= \mathsf{f}_{\boldsymbol{\lambda}_1}(g(\theta))|\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1\rangle\langle\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_1|\mathsf{f}_{\boldsymbol{\lambda}_1}(g(\theta))^{\dagger}.\end{aligned}$$

This scenario contains the case with Bosonic coherent states. In this case, since puseudo fidelity is fidelity, we employ the fidelity as the criterion in the following. Then, $c(\mathcal{H})$ is 0 when $\mathcal{H}$ is not $\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}$, and it is $1/d_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}$ and satisfies the condition (4.20) when $\mathcal{H}$ is $\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}$. Hence, we have

$$\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) = \frac{d_{\boldsymbol{\lambda}}}{d_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}}. \tag{4.23}$$

Now, we compare the averages of the puseudo fidelity $\mathrm{Tr}\,\rho_\theta\rho_{\hat{\theta}}$ of state estimation and generation of approximating state. The latter has a better average precision than the former in the general case because the latter has a larger choice than the former. However, in the above case of pure states, the guess to achieve the optimal precision is a pure state belonging to the state family $\{\rho_\theta\}$. So, both optimal averaged precision coincide with each other.

Example 4.3 (*Case of* $\boldsymbol{\lambda} = n\boldsymbol{\lambda}_1$) Consider the case when $\boldsymbol{\lambda} = n\boldsymbol{\lambda}_1$, as mentioned in Sect. 4.1. The optimal averaged precision is calculated to be $\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) = \frac{d_{n\boldsymbol{\lambda}_1}}{d_{(n+1)\boldsymbol{\lambda}_1}}$.

When $\mathfrak{g} = \mathfrak{su}(2), \mathfrak{su}(1,1)$, the highest weight λ_1 is a negative real number or a half integer. The optimal averaged precision is calculated to be $\frac{d_{n\lambda_1}}{d_{(n+1)\lambda_1}} = \frac{2n\lambda_1+1}{2(n+1)\lambda_1+1} = 1 - \frac{1}{n+1+\frac{1}{2\lambda}} \cong 1 - \frac{1}{n} + (1+\frac{1}{2\lambda})\frac{1}{n^2}$.

When $\mathfrak{g} = \mathfrak{h}(2,\mathbb{R}) \rtimes \mathfrak{u}(1)$, the highest weight is given as $\boldsymbol{\lambda}_1 = (\lambda_Z, \lambda_F)$. The optimal averaged precision is calculated to be $\frac{d_{n\boldsymbol{\lambda}_1}}{d_{(n+1)\boldsymbol{\lambda}_1}} = \frac{n\lambda_Z}{(n+1)\lambda_Z} = 1 - \frac{1}{(n+1)} \cong 1 - \frac{1}{n} + \frac{1}{n^2}$.

When $\mathfrak{g} = \mathfrak{su}(r)$, we consider the case when the highest weight is $\boldsymbol{\lambda}_1 = [1,0,\ldots,0]$. The optimal averaged precision is calculated to be $\frac{d_{n\boldsymbol{\lambda}_1}}{d_{(n+1)\boldsymbol{\lambda}_1}} = \frac{(n+r-1)!(r-1)!(n+1)!}{(r-1)!n!(n+1+r-1)!} = \frac{n+1}{n+1+r-1} = 1 - \frac{r-1}{n+r} \cong 1 - \frac{r-1}{n} + r(r-1)\frac{1}{n^2}$.

When $\mathfrak{g} = \mathfrak{h}(2r,\mathbb{R}) \rtimes \mathfrak{u}(r)$, the highest weight is given as $\boldsymbol{\lambda}_1 = (\lambda_Z, \lambda_F\boldsymbol{v})$. The optimal averaged precision is calculated to be $\frac{d_{n\boldsymbol{\lambda}_1}}{d_{(n+1)\boldsymbol{\lambda}_1}} = \frac{n^r\lambda_Z^r}{(n+1)^r\lambda_Z^r} = (1 - \frac{1}{(n+1)})^r \cong 1 - \frac{r}{n} + \frac{r(r+1)}{2}\frac{1}{n^2}$.

Example 4.4 (*One-mode squeezed state under Group* $\mathrm{H}(2,\mathbb{R}) \times \mathrm{U}(1)$ *[103]*) We assume that the representations f and f_1 of the group $G = \mathrm{H}(2,\mathbb{R}) \times \mathrm{U}(1)$ is the same Heisenberg representation on $L^2(\mathbb{R}) = \mathcal{U}_{1,0}$. (For its detail, see Sect. 7.1 of [44].) We also assume that the parametric space Θ is the group G and that ρ_{θ_0} and

ρ_{1,θ_0} are the squeezed state $|\zeta,\zeta'\rangle\langle\zeta,\zeta'|$, where ζ is the shift parameter and ζ' is the squeezing parameter [44, Sect. 7.5]. Applying a suitable squeezing operation, without loss of generality, we can assume that ρ_{θ_0} and ρ_{1,θ_0} are $|\zeta,0\rangle\langle\zeta,0|$.

To show that the condition (4.20) in Theorem 4.6 holds, we focus on the projection to the subspace with the highest weight $-n$, which is given as $\mathsf{U}_{1,2}^\dagger(I\otimes|n\rangle\langle n|)\mathsf{U}_{1,2}$ by using the unitary $\mathsf{U}_{1,2}$ defined in p. 243 of [44]. The vector $|\zeta,0\rangle\otimes|\zeta,0\rangle$ is the coherent vector with respect to the tensor product representation of the Lie group SU(1, 1) on the space $L^2(\mathbb{R})\otimes L^2(\mathbb{R})$, whose highest weight is $\lambda=-1/2$. This representation satisfies $\mathsf{K}_{-,-1}=\frac{1}{2}((\mathsf{a}_1^\dagger)^2+(\mathsf{a}_2^\dagger)^2)=\mathsf{U}(g_0)^\dagger\frac{1}{2}((\mathsf{a}_1^\dagger)^2+(\mathsf{a}_2^\dagger)^2)\mathsf{U}(g_0)$. Since

$$\begin{aligned}&|\zeta,0\rangle\otimes|\zeta,0\rangle=(1-|\zeta|^2)^{1/2}\exp(\zeta((\mathsf{a}_1^\dagger)^2+(\mathsf{a}_2^\dagger)^2))|0,0\rangle\\
=&\mathsf{U}(g_0)^\dagger(1-|\zeta|^2)^{1/2}\exp(\zeta((\mathsf{a}_1^\dagger)^2+(\mathsf{a}_2^\dagger)^2))\mathsf{U}(g_0)|0,0\rangle\\
=&\mathsf{U}(g_0)^\dagger(1-|\zeta|^2)^{1/2}\exp(\zeta((\mathsf{a}_1^\dagger)^2+(\mathsf{a}_2^\dagger)^2))|0,0\rangle=\mathsf{U}(g_0)^\dagger|\zeta,0\rangle\otimes|\zeta,0\rangle,\end{aligned}$$

we obtain

$$\begin{aligned}&\operatorname{Tr}\mathsf{U}(g_0)^\dagger(I\otimes|n\rangle\langle n|)\mathsf{U}(g_0)|\zeta,0\rangle\langle\zeta,0|\otimes|\zeta,0\rangle\langle\zeta,0|\\
=&\operatorname{Tr}\mathsf{U}(g_0)^\dagger(I\otimes|n\rangle\langle n|)\mathsf{U}(g_0)\mathsf{U}(g_0)^\dagger|\zeta,0\rangle\langle\zeta,0|\otimes|\zeta,0\rangle\langle\zeta,0|\mathsf{U}(g_0)^\dagger\\
=&\operatorname{Tr}(I\otimes|n\rangle\langle n|)|\zeta,0\rangle\langle\zeta,0|\otimes|\zeta,0\rangle\langle\zeta,0|=|\langle n|\zeta,0\rangle|^2.\end{aligned}$$

When n is odd, this value is zero. When n is even, this value is calculated as

$$|\langle 2n|\zeta,0\rangle|^2=\left|(1-|\zeta|^2)^{1/4}\langle 2n|\frac{\zeta^n(\mathsf{a}_2^\dagger)^{2n}}{2^n n!}|0\rangle\right|^2=(1-|\zeta|^2)^{1/2}\frac{|\zeta|^{2n}(2n)!}{(2^n n!)^2}.$$

Since $|\zeta|^2<1$ and $\frac{(2n)!}{(2^n n!)^2}$ is monotone decreasing for n, it takes the maximum value $(1-|\zeta|^2)^{1/2}$ when $n=0$. Hence, the condition (4.20) in Theorem 4.6 holds.

Now, we apply in Theorem 4.6. Since the formal dimension of the irreducible representation of G in the tensor product space is 2, we have $c_0=(1-|\zeta|^2)^{1/2}/2$, which implies that

$$\max_{\hat{M}}\mathcal{D}_F(\hat{M})=\mathcal{D}_F(\hat{M}_{\rho,\rho_1})=\frac{(1-|\zeta|^2)^{1/2}}{2}.$$

Example 4.5 (*r mode squeezed state under Group* $\mathrm{H}(2r,\mathbb{R})\times\mathrm{U}(r)$\*) We assume that the representations f and f_1 are given as representations of the group $G=\mathrm{H}(2r,\mathbb{R})\times\mathrm{U}(r)$ on the space $L^2(\mathbb{R})^{\otimes r}$ and that the parameter space Θ is G. Then, we additionally assume that the states ρ_{θ_0} and ρ_{1,θ_0} are the r mode-squeezed state $|\frac{-1}{2}\boldsymbol{v},\zeta\rangle\langle\frac{-1}{2}\boldsymbol{v},\zeta|$. (For the definition of $|\frac{-1}{2}\boldsymbol{v},\zeta\rangle\langle\frac{-1}{2}\boldsymbol{v},\zeta|$, see (7.81) of [44].) Given a complex symmetric matrix ζ, we can choose a unitary g such that $g\zeta g^T$ is the diagonal matrix with the diagonal elements $\zeta_1, \ldots, \zeta_r$. Hence, without loss of generality, we can assume that $|\frac{-1}{2}\boldsymbol{v}, \zeta\rangle = |\zeta_1, 0\rangle \otimes \cdots \otimes |\zeta_r, 0\rangle$.

In the following, we apply Lemma 4.1 to the case when G' is $\mathrm{H}(2r, \mathbb{R}) \times \mathrm{U}(1)^r$, where each subgroup $\mathrm{U}(1)$ of $\mathrm{U}(1)^r$ acts on each mode independently. That is, $\mathrm{U}(1)^r$ is the group composed of the diagonal elements. Due to the same reason as the previous example, the condition (4.20) holds for G'. An irreducible space with respect to $\mathrm{H}(2r, \mathbb{R})$ is also irreducible with respect to $\mathrm{H}(2r, \mathbb{R}) \times \mathrm{U}(r)$ when it is closed for the action of $\mathrm{H}(2r, \mathbb{R}) \times \mathrm{U}(r)$. So, we can apply Lemma 4.1 to the representation and the condition (4.20) holds for $\mathrm{H}(2r, \mathbb{R}) \times \mathrm{U}(r)$. Hence, we obtain

$$\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \max_{\rho, \rho_1} \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) = \prod_{j=1}^{r} \frac{(1 - |\zeta_j|^2)^{1/2}}{2} = \sqrt{\det \frac{I - \overline{\zeta}\zeta}{2}}.$$

In particular, since the RHS is invariant constant for the action of $\mathrm{U}(r)$, the relation

$$\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \max_{\rho, \rho_1} \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) = \sqrt{\det \frac{I - \overline{\zeta}\zeta}{2}}$$

holds for any r-mode squeezed state.

Exercise 4.2 Calculate the asymptotic behavior of the optimal averaged precision with the limit $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \to \infty$ when $\mathfrak{g} = \mathfrak{su}(2)$.

Exercise 4.3 Calculate the asymptotic behavior of the optimal averaged precision with the limit $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \to \infty$ when $\mathfrak{g} = \mathfrak{su}(1, 1)$.

Exercise 4.4 Calculate the asymptotic behavior of the optimal averaged precision when $\mathfrak{g} = \mathfrak{h}(2, \mathbb{R}) \rtimes \mathfrak{u}(1)$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 = (\lambda_Z, \lambda_F)$.

Exercise 4.5 Calculate the asymptotic behavior of the optimal averaged precision with the limit $n \to \infty$ when $\mathfrak{g} = \mathfrak{su}(r)$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 = [n, 0, \ldots, 0]$. That is, both systems are symmetric subspaces.

Exercise 4.6 Calculate the asymptotic behavior of the optimal averaged precision when $\mathfrak{g} = \mathfrak{h}(2r, \mathbb{R}) \rtimes \mathfrak{u}(r)$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 = (\lambda_Z, \lambda_F \boldsymbol{v})$.

4.2.4 Generation of Approximating State for Mixed States: Puseudo Fidelity

We discuss the generation of approximating state when the precision criterion is puseudo fidelity, G is a Lie group, and the states ρ_0 and $\rho_{1,0}$ are given as $ce^{i\mathfrak{f}(X)}$ and $c_1 e^{i\mathfrak{f}_1(X)}$ by using constants $c, c_1 > 0$ and an element X of maximal Cartan subalgebra of Lie algebra $\mathfrak{g}$, respectively. Hence, the state families are given as

$$\rho_\theta := c\mathsf{f}(g(\theta))e^{i\mathsf{f}(X)}\mathsf{f}(g(\theta))^\dagger, \quad \rho_{1,\theta} := c_1\mathsf{f}_1(g(\theta))e^{i\mathsf{f}_1(X)}\mathsf{f}_1(g(\theta))^\dagger.$$

When the tensor product representation $\mathsf{f}' \otimes \mathsf{f}_1'$ has the irreducible decomposition $\mathcal{H} \otimes \mathcal{H}_1 = \oplus_{\boldsymbol{\lambda}} \mathcal{U}_{\boldsymbol{\lambda}}$, we have $c(\mathcal{U}_{\boldsymbol{\lambda}}) = cc_1 \frac{\operatorname{Tr} e^{i\mathsf{f}_{\boldsymbol{\lambda}}(X)}}{\dim \mathcal{U}_{\boldsymbol{\lambda}}}$. Therefore, when the condition (4.20) of Theorem 4.6 holds, i.e., the relation $c(\mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}) > c(\mathcal{U}_{\boldsymbol{\lambda}})$ holds for $\boldsymbol{\lambda}$ that appears in the above irreducible decomposition, we have

$$\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \max_{\rho,\rho_1} \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) = \dim \mathcal{U}_{\boldsymbol{\lambda}} cc_1 \frac{\operatorname{Tr} e^{i\mathsf{f}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}(X)}}{\dim \mathcal{U}_{\boldsymbol{\lambda}+\boldsymbol{\lambda}_1}}.$$

In the following, we examine several examples by using this fact.

Example 4.6 We assume that $\mathfrak{g} = \mathfrak{su}(2)$, $\mathcal{H} = \mathcal{U}_\lambda$, and $\mathcal{H}_1 = \mathcal{U}_{\lambda_1}$. Then, by using $t, c, c_1 > 0$, the states ρ_0 and $\rho_{1,0}$ are assumed to be written as $\rho_0 = ce^{t\mathsf{E}_{0,1}}$ and $\rho_{1,0} = c_1 e^{t\mathsf{E}_{0,1}}$. Then, we have $c = \frac{e^t-1}{e^{t(\lambda+1)}-e^{-t\lambda}}$ and $c_1 = \frac{e^t-1}{e^{t(\lambda_1+1)}-e^{-t\lambda_1}}$. Given a weight $\lambda' > 0$, we can calculate as $c(\mathcal{U}_{\lambda'}) = cc_1 \frac{e^{t(\lambda'+1)}-e^{-t\lambda'}}{(e^t-1)(2\lambda'+1)}$. Thus, since the relation $c(\mathcal{U}_{\lambda'}) > c(\mathcal{U}_{\lambda''})$ holds for two weights λ' and λ'' satisfying $\lambda' > \lambda'' > 0$, the condition (4.20) holds. Hence, we obtain

$$\begin{aligned}
&\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \max_{\rho,\rho_1} \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) = (2\lambda+1)cc_1 \frac{e^{t(\lambda+\lambda_1+1)} - e^{-t(\lambda+\lambda_1)}}{(e^t-1)(2(\lambda+\lambda_1)+1)} \\
=&\frac{(2\lambda+1)(e^t-1)(e^{t(\lambda+\lambda_1+1)} - e^{-t(\lambda+\lambda_1)})}{(2(\lambda+\lambda_1)+1)(e^{t(\lambda_1+1)} - e^{-t\lambda_1})(e^{t(\lambda+1)} - e^{-t\lambda})}. \qquad (4.24)
\end{aligned}$$

When t goes to infinity, the above values approaches the optimum value $\frac{2\lambda+1}{2(\lambda+\lambda_1)+1}$ with the pure states case.

Example 4.7 We assume that $\mathfrak{g} = \mathfrak{su}(1,1)$, $\mathcal{H} = \mathcal{U}_\lambda$, and $\mathcal{H}_1 = \mathcal{U}_{\lambda_1}$. Then, by using $t, c, c_1 > 0$, the states ρ_0 and $\rho_{1,0}$ are assumed to be written as $\rho_0 = ce^{t\mathsf{E}_{0,-1}}$ and $\rho_{1,0} = c_1 e^{t\mathsf{E}_{0,-1}}$. Then, we have $c = \frac{1-e^{-t}}{e^{t\lambda}}$ and $c_1 = \frac{1-e^{-t}}{e^{t\lambda_1}}$. Given a weight $\lambda' < 0$, we can calculate as $c(\mathcal{U}_{\lambda'}) = cc_1 \frac{e^{t\lambda'}}{-(2\lambda'+1)(1-e^{-t})}$. Thus, since the relation $c(\mathcal{U}_{\lambda'}) > c(\mathcal{U}_{\lambda''})$ holds for two weights λ' and λ'' satisfying $0 > \lambda' > \lambda''$, the condition (4.20) holds. Hence, we obtain

$$\begin{aligned}
&\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \max_{\rho,\rho_1} \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) \\
=&-(2\lambda+1)cc_1 \frac{e^{t(\lambda+\lambda_1)}}{-(2(\lambda+\lambda_1)+1)(1-e^{-t})} = \frac{(2\lambda+1)(1-e^{-t})}{2(\lambda+\lambda_1)+1}. \qquad (4.25)
\end{aligned}$$

When t goes to infinity, the above values approaches the optimum value $\frac{2\lambda+1}{2(\lambda+\lambda_1)+1}$ with the pure states case.

Example 4.8 We assume that $\mathfrak{g} = \mathfrak{h}(2,\mathbb{R}) \rtimes \mathfrak{u}(1)$, $\mathcal{H} = \mathcal{U}_{\lambda_Z,\lambda_F}$, and $\mathcal{H}_1 = \mathcal{U}_{\lambda_{Z,1},\lambda_{F,1}}$. Then, by using $t, c, c_1 > 0$, the states ρ_0 and $\rho_{1,0}$ are assumed to be written as

$\rho_0 = ce^{-t\mathsf{N}}$ and $\rho_{1,0} = c_1 e^{-t\mathsf{N}}$. Then, we have $c = \frac{1-e^{-t}}{e^{t\lambda_F}}$ and $c_1 = \frac{1-e^{-t}}{e^{t\lambda_{F,1}}}$. Since the formal dimension of the tensor product space is 2, we have $c(\mathcal{U}_{\lambda_Z+\lambda_{Z,1},\lambda'_F}) = cc_1 \frac{e^{t\lambda'_F}}{\lambda_Z+\lambda_{Z,1}(1-e^{-t})}$. Thus, since the relation $c(\mathcal{U}_{\lambda_Z+\lambda_{Z,1},\lambda'_F}) > c(\mathcal{U}_{\lambda_Z+\lambda_{Z,1},\lambda''_F})$ holds for two integers λ'_F and λ''_F satisfying $\lambda'_F > \lambda''_F$, the irreducible decomposition [45, (7.29)] guarantees the condition (4.20). Hence, we obtain

$$\begin{aligned}
&\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \max_{\rho,\rho_1} \mathcal{D}_F(\hat{M}_{\rho,\rho_1}) \\
=&\lambda_Z cc_1 \frac{e^{t(\lambda_F+\lambda_{F,1})}}{(\lambda_Z+\lambda_{Z,1})(1-e^{-t})} = \frac{\lambda_Z(1-e^{-t})}{\lambda_Z+\lambda_{Z,1}}.
\end{aligned} \tag{4.26}$$

When t goes to infinity, the above values approaches the optimum value $\frac{\lambda_Z}{\lambda_Z+\lambda_{Z,1}}$ with the pure states case.

Exercise 4.7 Consider the case discussed in Example 4.6. Calculate the asymptotic behavior of the optimal averaged precision with the limit $\lambda = \lambda_1 \to \infty$.

Exercise 4.8 Consider the case discussed in Example 4.7. Calculate the asymptotic behavior of the optimal averaged precision with the limit $\lambda = \lambda_1 \to \infty$.

Exercise 4.9 Consider the case discussed in Example 4.8. Calculate the optimal averaged precision when $\lambda = \lambda_1$.

4.2.5 Generation of Approximating State for Mixed States: Fidelity

Next, choosing the square of the fidelity as the criterion of the precision, we consider the generation of approximating state for mixed states when G is a Lie group. However, this case is much more complicated than the case with pseudo fidelity. Hence, we discuss only several special cases.

Firstly, we consider the case when $\mathfrak{g} = \mathfrak{su}(2)$, $\mathcal{H} = \mathcal{U}_{n/2}$, and $\mathcal{H}_1 = \mathcal{U}_{1/2}$. Since Theorem 4.6 guarantees that the optimal performance can be attained by a covariant operation, we restrict our operation to an operation with the form $\hat{M}_{\rho,\rho_1}$. There exists a suitable unitary $g \in \mathrm{SU}(2)$ such that $g^\dagger \rho_1 g$ is diagonal. Hence, without loss of generality, we can assume that the state ρ_1 is a diagonal matrix $\rho_{(r',0,0)}$.

Similar to Example 4.6, by using $t, c, c_1 > 0$, the states ρ_0 and $\rho_{1,0}$ are assumed to be written as $\rho_0 = ce^{t\mathsf{f}(E_{0,1})}$ and $\rho_{1,0} = c_1 e^{tE_{0,1}}$. Then, we choose r such that $\rho_{1,0} = \rho_{(r,0,0)}$. We focus on the homogeneous space S^2, which is the quotient space of $\mathrm{SU}(2)$ divided by the one-dimensional subgroup H composed on diagonal elements. Using $T(\rho) := \int_H g\rho g \mu_H(dg)$, we have

$$
\begin{aligned}
&\max_{r',\rho} d_{\frac{n}{2}} \int_{\mathrm{SU}(2)} F(\rho_{r,0,0}, \rho_{r',0,0})^2 \operatorname{Tr} c e^{t\mathsf{f}(E_{0,1})} \mathsf{f}_{n/2}(g) \rho \mathsf{f}_{n/2}(g)^\dagger \mu(dg) \\
=& \max_{r',\rho} d_{\frac{n}{2}} \int_{S^2} F(\rho_{r,0,0}, \rho_{r'\boldsymbol{x}}) \operatorname{Tr} c e^{t\mathsf{f}(E_{0,1})} \mathsf{f}_{n/2}(g(\boldsymbol{x})) T(\rho) \mathsf{f}_{n/2}(g(\boldsymbol{x}))^\dagger \mu_{S^2}(d\boldsymbol{x}), \quad (4.27)
\end{aligned}
$$

where the element $\mathfrak{g}(\boldsymbol{x}) \in \mathrm{SU}(2)$ is chosen as $g(\boldsymbol{x})\rho_{(r,0,0)}g(\boldsymbol{x})^\dagger = \rho_{\boldsymbol{x}}$. Then, by letting $\boldsymbol{x}_0 := (1, 0, 0)$, the relation (2.50) implies that

$$
\begin{aligned}
F(\rho_{r,0,0}, \rho_{r'\boldsymbol{x}}) =& \frac{1 + rr'\boldsymbol{x}_0 \cdot \boldsymbol{x} + \sqrt{1-r^2}\sqrt{1-r'^2}}{2} \\
=& r' \frac{1 + r\boldsymbol{x}_0 \cdot \boldsymbol{x}}{2} + \frac{1 - r' + \sqrt{1-r^2}\sqrt{1-r'^2}}{2}. \quad (4.28)
\end{aligned}
$$

In the RHS, only the first term is a random variable. By using the relation $ce^{t\mathsf{E}_{0,1}} = \rho_{(r,0,0)}$, i.e., $e^t = \frac{1+r}{1-r}$, the discussion of Example 4.6 evaluates the expectation as

$$
\begin{aligned}
& d_{\frac{n}{2}} \int_{S^2} \frac{1 + r\boldsymbol{x}_0 \cdot \boldsymbol{x}}{2} \operatorname{Tr} c e^{t\mathsf{f}(E_{0,1})} \mathsf{f}_{n/2}(g(\boldsymbol{x})) T(\rho) \mathsf{f}_{n/2}(g(\boldsymbol{x}))^\dagger \mu_{S^2}(d\boldsymbol{x}) \\
=& d_{\frac{n}{2}} \int_{S^2} (\operatorname{Tr} \rho_{r,0,0}\rho_{\boldsymbol{x}}) \operatorname{Tr} c e^{t\mathsf{f}(E_{0,1})} \mathsf{f}_{n/2}(g(\boldsymbol{x})) T(\rho) \mathsf{f}_{n/2}(g(\boldsymbol{x}))^\dagger \mu_{S^2}(d\boldsymbol{x}) \\
\le& \frac{(n+1)(e^t - 1)(e^{(\frac{n+1}{2}+1)t} - e^{-\frac{n+1}{2}t})}{(n+2)(e^{\frac{3}{2}t} - e^{-\frac{1}{2}t})(e^{(\frac{n}{2}+1)t} - e^{-\frac{n}{2}t})} \\
=& \frac{(n+1)((1+r)^{n+2} - (1-r)^{n+2})}{2(n+2)((1+r)^{n+1} - (1-r)^{n+1})}, \quad (4.29)
\end{aligned}
$$

where the equality is attained when $T(\rho)$ is the highest weight state. Hence, the relation (4.28) yields that

$$
\begin{aligned}
& \max_{\rho} d_{\frac{n}{2}} \int_{S^2} F(\rho_{r,0,0}, \rho_{r'\boldsymbol{x}}) \operatorname{Tr} c e^{t\mathsf{f}(E_{0,1})} \mathsf{f}_{n/2}(g(\boldsymbol{x})) T(\rho) \mathsf{f}_{n/2}(g(\boldsymbol{x}))^\dagger \mu_{S^2}(d\boldsymbol{x}) \\
=& r' \frac{(n+1)((1+r)^{n+2} - (1-r)^{n+2})}{2(n+2)((1+r)^{n+1} - (1-r)^{n+1})} + \frac{1 - r' + \sqrt{1-r^2}\sqrt{1-r'^2}}{2} \\
=& \frac{1}{2} + r' \left(\frac{(n+1)((1+r)^{n+2} - (1-r)^{n+2})}{2(n+2)((1+r)^{n+1} - (1-r)^{n+1})} - \frac{1}{2} \right) + \sqrt{1-r'^2} \frac{\sqrt{1-r^2}}{2} \\
& \quad (4.30) \\
\le& \frac{1}{2} + \sqrt{ \left(\frac{(n+1)((1+r)^{n+2} - (1-r)^{n+2})}{2(n+2)((1+r)^{n+1} - (1-r)^{n+1})} - \frac{1}{2} \right)^2 + \frac{1-r^2}{4} }. \quad (4.31)
\end{aligned}
$$

Since the equality of the inequality (4.31) holds with a suitable r', the maximum of (4.27) equals the RHS of (4.31).

4.2.6 Generation of Approximating State: Relative Entropy

Next, we consider the generation of approximating state for the mixed state when the precision criterion is the relative entropy. To treat this criterion, we need more complicated discussion than the criterion based on the fidelity even when the state is restricted to a pure state. In this case, given a state family ρ_θ, $\rho_{1,\theta}$ on the system $\mathcal{H}$, $\mathcal{H}_1$, we minimize $\int_\Theta \int_\Omega D(\rho_{1,\theta}\|\rho_\omega) \operatorname{Tr} \rho_\theta M_\omega d\omega \mu(d\theta)$ with respect to our operation $\hat{M} = (M, \{\rho_\omega\})$. Using $\rho[M]_\omega := \int_\Theta \rho_{1,\theta}(\operatorname{Tr} \rho_\theta M_\omega)\mu(d\theta)$, we can calculate this value as

$$\begin{aligned}
&\int_\Theta \int_\Omega D(\rho_{1,\theta}\|\rho_\omega) \operatorname{Tr} \rho_\theta M_\omega d\omega \mu(d\theta) \\
=& \int_\Theta H(\rho_{1,\theta})\mu(d\theta) - \int_\Omega \operatorname{Tr} \int_\Theta \rho_{1,\theta}(\operatorname{Tr} \rho_\theta M_\omega)\mu(d\theta) \log \rho_\omega d\omega \\
=& \int_\Theta H(\rho_{1,\theta})\mu(d\theta) \\
&+ \int_\Omega \operatorname{Tr} \rho[M]_\omega \Big(H\Big(\frac{\rho[M]_\omega}{\operatorname{Tr} \rho[M]_\omega}\Big) + D\Big(\frac{\rho[M]_\omega}{\operatorname{Tr} \rho[M]_\omega} \Big\| \rho_\omega\Big)\Big) d\omega. \qquad (4.32)
\end{aligned}$$

So, when we apply POVM $M = \{M_\omega\}$, the optimal guessing state ρ_ω with the measurement outcome ω is $\frac{\rho[M]_\omega}{\operatorname{Tr} \rho[M]_\omega}$. Thus, the remaining problem is to find the POVM M to minimize

$$\int_\Theta H(\rho_{1,\theta})\mu(d\theta) + \int_\Omega \operatorname{Tr} \rho[M]_\omega H\big(\frac{\rho[M]_\omega}{\operatorname{Tr} \rho[M]_\omega}\big) d\omega. \qquad (4.33)$$

Here, using the reference system $\mathcal{R}$, we choose a purification $|x\rangle$ of the state $\int_\Theta \operatorname{Tr} \rho_\theta \otimes \rho_{1,\theta} \nu(d\theta)$ on the system $\mathcal{H} \otimes \mathcal{H}_1$. Due to the expression (3.46) for $\mathcal{D}_F(\rho)$, the second term (4.33) equals $\mathcal{D}_F(\operatorname{Tr}_{\mathcal{H}} |x\rangle\langle x|)$. However, it is not so easy to calculate $\mathcal{D}_F(\operatorname{Tr}_{\mathcal{H}} |x\rangle\langle x|)$ in general. Hence, we discuss a special case in the following.

Firstly, we consider the case when $\mathfrak{g} = \mathfrak{su}(2)$, $\mathcal{H} = \mathcal{U}_{n/2}$, and $\mathcal{H}_1 = \mathcal{U}_{1/2}$. Since Theorem 4.6 guarantees that the optimal average precision can be attained by a covariant operation, we restrict our operation into the operation with the form $\hat{M}_{\rho,\rho_1}$. Given a state ρ_1, there exists a unitary $g \in \mathrm{SU}(2)$ such that $g^\dagger \rho_1 g$ is diagonal. In the following discussion, we fix a state ρ_1 so that our guess ρ_ω is $\frac{\rho[M]_\omega}{\operatorname{Tr} \rho[M]_\omega}$. Under this constraint, we optimize our POVM M. We will show that an optimal POVM is given as a covariant POVM generated by the highest weight vector. Hence, we consider the case when ρ_1 is a diagonal matrix $\rho_{(r',0,0)}$.

Similar to Example 4.6, using $t, c, c_1 > 0$, we can write ρ_0 and $\rho_{1,0}$ as $\rho_0 = ce^{t\mathsf{f}(E_{0,1})}$ and $\rho_{1,0} = c_1 e^{tE_{0,1}}$. Then, we choose r such that $\rho_{1,0} = \rho_{(r,0,0)}$. We focus on the homogeneous space S^2, which is the quotient space of SU(2) divided by the one-dimensional subgroup H composed on diagonal elements. Using $T(\rho) := \int_H g\rho g \mu_H(dg)$, we have

$$\begin{aligned}
&\min_{r',\rho} d_{\frac{n}{2}} \int_{\mathrm{SU}(2)} D(\rho_{r,0,0}\|\rho_{r',0,0}) \operatorname{Tr} c e^{t\mathsf{f}(E_{0,1})} \mathsf{f}_{n/2}(g)\rho \mathsf{f}_{n/2}(g)^{\dagger}\mu(dg) \\
=& \min_{r',\rho} d_{\frac{n}{2}} \int_{S^2} D(\rho_{r,0,0}\|\rho_{r'\boldsymbol{x}}) \operatorname{Tr} c e^{t\mathsf{f}(E_{0,1})} \mathsf{f}_{n/2}(g(\boldsymbol{x}))T(\rho) \mathsf{f}_{n/2}(g(\boldsymbol{x}))^{\dagger}\mu_{S^2}(d\boldsymbol{x}),
\end{aligned}$$

where we choose $\mathfrak{g}(\boldsymbol{x}) \in \mathrm{SU}(2)$ such that $g(\boldsymbol{x})\rho_{(r,0,0)}g(\boldsymbol{x})^{\dagger} = \rho_{\boldsymbol{x}}$. Due to (4.33), using $\boldsymbol{x}_0 := (1, 0, 0)$, we obtain

$$\begin{aligned}
&D(\rho_{r,0,0}\|\rho_{r'\boldsymbol{x}}) \\
=&h(\frac{1+r}{2}) - \frac{1+r\boldsymbol{x}_0\cdot\boldsymbol{x}}{2}\log(\frac{1+r'}{2}) - \frac{1-r\boldsymbol{x}_0\cdot\boldsymbol{x}}{2}\log(\frac{1-r'}{2}) \\
=&h(\frac{1+r}{2}) - \log(\frac{1-r'}{2}) - \frac{1+r\boldsymbol{x}_0\cdot\boldsymbol{x}}{2}\log(\frac{1+r'}{1-r'}).
\end{aligned}$$

Notice that only the final term is a random variable. The expectation is evaluated by (4.29). The equality in (4.29) holds when $T(\rho)$ is a highest weight state. Hence, we can restrict our POVM M to a covariant POVM generated by a highest weight vector. Then, due to (4.32), the optimal ρ_1 is $d_{\frac{n}{2}}\int_{\Theta}\operatorname{Tr}\rho_{1,\theta}\langle\frac{n}{2};\frac{n}{2}|\rho_\theta|\frac{n}{2};\frac{n}{2}\rangle\mu(d\theta)$.

When ρ_θ and $\rho_{1,\theta}$ are coherent states, i.e., pure states, by using the formula [44, (4.30)] for Clebsh-Gordan coefficient, this can be calculated to be

$$\begin{aligned}
&d_{\frac{n}{2}}\int_{\Theta}\operatorname{Tr}\rho_{1,\theta}\langle\frac{n}{2};\frac{n}{2}|\rho_\theta|\frac{n}{2};\frac{n}{2}\rangle\mu(d\theta) \\
=&\frac{d_{\frac{n}{2}}}{d_{\frac{n+1}{2}}}|\frac{1}{2};\frac{1}{2}\rangle\langle\frac{1}{2};\frac{1}{2}|\langle\frac{1}{2},\frac{n}{2};\frac{1}{2},\frac{n}{2}|\frac{n+1}{2};\frac{n+1}{2}\rangle\langle\frac{n+1}{2};\frac{n+1}{2}|\frac{1}{2},\frac{n}{2};\frac{1}{2},\frac{n}{2}\rangle \\
&+\frac{d_{\frac{n}{2}}}{d_{\frac{n+1}{2}}}|\frac{1}{2};-\frac{1}{2}\rangle\langle\frac{1}{2};-\frac{1}{2}|\langle\frac{1}{2},\frac{n}{2};-\frac{1}{2},\frac{n}{2}|\frac{n+1}{2};\frac{n-1}{2}\rangle \\
&\cdot\langle\frac{n+1}{2};\frac{n-1}{2}|\frac{1}{2},\frac{n}{2};-\frac{1}{2},\frac{n}{2}\rangle \\
=&\frac{n+1}{n+2}|\frac{1}{2};\frac{1}{2}\rangle\langle\frac{1}{2};\frac{1}{2}|+\frac{1}{n+2}|\frac{1}{2};-\frac{1}{2}\rangle\langle\frac{1}{2};-\frac{1}{2}|.
\end{aligned}$$

Hence, the minimum value of (4.33) is calculated to be $h(\frac{1}{n+2})$.

4.3 Identification of Group Action

4.3.1 Formulation of Problem

In this section, given a (projective) representation f of the group G on the representation space $\mathcal{H}$, when the physical operation $\mathsf{f}(g)$ based on the unknown element $g \in G$

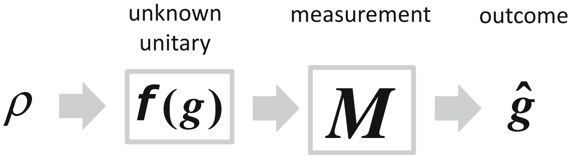


Fig. 4.4 Identification of group action

acts on the physical system $\mathcal{H}$, we discuss the problem to identify the element g. In this case, we prepare an initial state ρ on the system $\mathcal{H}$. Then, the resultant state is given as $\mathsf{f}(g)\rho\mathsf{f}(g)^\dagger$. So, applying a POVM M taking values in G, we estimate the unknown element $g \in G$. In this discussion, the (projective) representation f is not necessarily irreducible. As explained later, we need to utilize the correlation among different irreducible components in the representation space $\mathcal{H}$.

In this problem, our operation is given as a pair $\mathcal{I} := (\rho, M)$ of an initial state ρ on the system $\mathcal{H}$ and a POVM M on $\mathcal{H}$ that output the outcome $\hat{g}$ in G, as Fig. 4.4. When we evaluate the error between the true $\mathsf{f}(g)$ and our guess $\mathsf{f}(\hat{g})$ by $R(g, \hat{g})$, the performance of our operation $\mathcal{I}$ with the true operation $\mathsf{f}(g)$ is given as

$$\mathcal{D}_{R,g}(\mathcal{I}) := \int_G R(g, \hat{g}) \operatorname{Tr} M(dg)\mathsf{f}(g)\rho\mathsf{f}(g)^\dagger. \tag{4.34}$$

Then, similar to the discussion in Sect. 4.1.1, we define the mini-max error $\mathcal{D}_R(\mathcal{I})$ and Bayesian error $\mathcal{D}_{R,\nu}(\mathcal{I})$ with the prior ν as follows.

$$\mathcal{D}_R(\mathcal{I}) := \max_{g\in G} \mathcal{D}_{R,g}(\mathcal{I}), \quad \mathcal{D}_{R,\nu}(\mathcal{I}) := \int_G \mathcal{D}_{R,g}(\mathcal{I})\nu(dg).$$

When we focus on the Bayesian method with the normalized invariant measure μ_G or the mini-max method, and when we fix the initial state ρ and apply Theorem 4.1 to the optimization for M, the optimal performance $\min_{\mathcal{I}} \mathcal{D}_R(\mathcal{I}) = \min_{\mathcal{I}} \mathcal{D}_{R,\mu_G}(\mathcal{I})$ can be attained by a covariant measurement taking values in G. When $\rho = \sum_i p_i |x_i\rangle\langle x_i|$, since $\mathcal{D}_{R,g}(M, \rho) = \sum_i p_i \mathcal{D}_{R,g}(M, |x_i\rangle\langle x_i|)$, the optimal performance $\min_{(M,\rho)} \mathcal{D}_R(M, \rho) = \min_{(M,\rho)} \mathcal{D}_{R,\mu_G}(M, \rho)$ can be attained by a pure state ρ. Hence, we assume that our POVM M is a covariant measurement and our initial state is a pure state $|x\rangle\langle x|$. Then, we denote the representation space $\mathcal{H}$ by $\oplus_{\lambda\in S}\mathcal{U}_\lambda \otimes \mathbb{C}^{n_\lambda}$, where $S \subset \hat{G}$ is the set of irreducible representations appearing in $\mathcal{H}$. For simplicity, we simplify $\mathcal{D}_{R,g}(M_T, \rho)$ to $\mathcal{D}_{R,g}(T, \rho)$.

Due to the symmetric structure of our problem, it is enough to discuss the case when the true element g is the unit element. Then, the group invariance (4.5) yields that

$$R(e, g^{-1}\hat{g}g) = R(g^{-1}, g^{-1}\hat{g}) = R(e, \hat{g}). \tag{4.35}$$

Since $R(e, \hat{g})$ is a real number, the error $R(e, \hat{g})$ can be written as a sum of irreducible characters as $\sum_\lambda a_\lambda(\chi_\lambda(\hat{g}) + \chi_{\lambda^*}(\hat{g}))$ (see [44, Lemma 2.16]).

4.3.2 Case of Commutative Group

When the group G is commutative, any irreducible representation is a one-dimensional representation. Hence, denoting the basis of the one-dimensional space corresponding to $\lambda \in \hat{G}$ by $|\hat{e}_\lambda\rangle$, we can write any pure initial state as $|\varphi\rangle := \sum_{\lambda \in S} \varphi_\lambda |\hat{e}_\lambda\rangle$, which belongs to the square integrable space $L^2(\hat{G})$ on $\hat{G}$. Since $\dim \mathcal{U}_\lambda = 1$, it is enough to the case when $n_\lambda = 1$ for $\lambda \in S$.

Then, a matrix $T \geq 0$ describes a covariant POVM M_T, and is written as $T = \sum_k |\eta_k\rangle\langle\eta_k|$ in general. When $|\eta_k\rangle = \sum_{\lambda \in S} \sqrt{q_{\lambda,k}} e^{i\theta_{\lambda,k}} |\hat{e}_\lambda\rangle$ and the true element is the unit element e, the probability distribution of the estimate $\hat{g}$ is $\sum_k |\sum_{\lambda \in S} \mathsf{f}_\lambda(\hat{g}) e^{-i\theta_{\lambda,k}} \sqrt{q_{\lambda,k}} \varphi_\lambda|^2 \mu_G(d\hat{g})$.

This value is the same as the average error when the measurement is generated by $T = |\eta_0\rangle\langle\eta_0|$ with $|\eta_0\rangle := \sum_{\lambda \in S} |\hat{e}_\lambda\rangle$ and the initial state is the mixed state $\sum_k |\phi_k\rangle\langle\phi_k|$ with $|\phi_k\rangle := \sum_\lambda e^{-i\theta_{\lambda,k}} \sqrt{q_{\lambda,k}} |\hat{e}_\lambda\rangle$. Hence, without loss of generality, we can assume that our measurement is restricted to the POVM generated by $T = |\eta_0\rangle\langle\eta_0|$. That is, our operation to estimate the unknown element g is identified with the initial state $\varphi \in L^2(\hat{G})$. When the true element g is the unit element e, the distribution of the estimate $\hat{g}$ is characterized by using the inverse Fourier transform $\mathcal{F}^{-1}[\varphi] := \sum_{\lambda \in S} \mathsf{f}_\lambda(\hat{g}^{-1}) \varphi_\lambda$ of φ (see [44, Sect. 3.8.1] for the precise discussion of inverse Fourier transform) as

$$\left|\sum_{\lambda \in S} \mathsf{f}_\lambda(\hat{g}^{-1}) \varphi_\lambda\right|^2 \mu_G(d\hat{g}) = |\mathcal{F}^{-1}[\varphi](\hat{g}^{-1})|^2 \mu_G(d\hat{g}),$$

where $\sum_{\lambda \in S}$ is replaced by an integral when G is not compact.

When the gain function $\tilde{R}(g, \hat{g})$ is given as $\sum_{\lambda \in \hat{G}} a_\lambda \mathsf{f}_\lambda(g^{-1}\hat{g})$ by using a non-negative valued function a on $\hat{G}$, for a vector state $\phi \in \mathcal{H}$ and the vector $|I\rangle := \sum_{\lambda \in S} |\hat{e}_\lambda\rangle$, we have

$$\mathcal{D}_{\tilde{R}}(\varphi) := \mathcal{D}_{\tilde{R}}(\varphi, M_{|I\rangle\langle I|}) = \sum_{\lambda, \lambda' \in S} a_{\lambda'} \varphi_\lambda \overline{\varphi_{\lambda+\lambda'}}. \tag{4.36}$$

Example 4.9 When $G = \mathbb{R}$, we have $L^2(\hat{G}) = L^2(\mathbb{R})$. So, the initial state is given as a vector state $\varphi(\lambda)$ on $L^2(\mathbb{R})$. Then, we have $\mu_G(d\hat{g}) = \frac{1}{\sqrt{2\pi}} d\hat{g}$ and $\mu_{\hat{G}}(d\lambda) = \frac{1}{\sqrt{2\pi}} d\lambda$, which implies that $\int_{\mathbb{R}} |\varphi(\lambda)|^2 \frac{1}{\sqrt{2\pi}} d\lambda = 1$. When the true parameter is 0, the distribution of the estimate $\hat{g} \in \mathbb{R}$ is given as $|\mathcal{F}^{-1}[\varphi](-\hat{g})|^2 \frac{1}{\sqrt{2\pi}} d\hat{g}$ by using the inverse Fourier transform $\mathcal{F}^{-1}[\varphi]$ of φ.

For example, when we impose the constraint $\int \lambda^2 |\varphi(\lambda)|^2 \frac{1}{\sqrt{2\pi}} d\lambda \leq E^2$ with a constant E to the initial state φ, the mean square error $\int \hat{g}^2 |\mathcal{F}^{-1}[\varphi](-\hat{g})|^2 d\hat{g}$ is minimized by the vector state $\varphi(\lambda) = \frac{1}{\sqrt{E}} e^{-\frac{\lambda^2}{4E^2}}$. Its inverse Fourier transform is $\sqrt{2E} e^{-E^2\lambda^2}$ and attains the minimum value $\frac{1}{4E^2}$ of the mean square error.

When we restrict the support of $\varphi(\lambda)$ to the closed interval $[-1, 1]$, the vector state φ minimizing the mean square error is $(2\pi)^{1/4} \sin \frac{\pi(1+\lambda)}{2}$. Then, the inverse Fourier transform is $-\frac{\pi^{3/4}}{2^{1/4}} \frac{\cos \hat{g}}{\hat{g}^2 - \pi^2/4}$ and attains the minimum value $\frac{\pi^2}{4}$.

Here, we prepare the following theorem for the next example.

Theorem 4.9 ([23]) *The $n \times n$ matrix $P_n := \sum_{k=1}^{n-1} |k\rangle\langle k+1| + |k+1\rangle\langle k|$ has the eigenvector $x^j := \sum_{k=1}^{n} \sin \frac{jk\pi}{n+1} |k\rangle$ associated with the eigenvalue $2\cos\frac{j\pi}{n+1}$ for $j = 1, \ldots, n$.*

Example 4.10 (*[13, 89, 130]*) When $G = U(1) = \{e^{i\theta} | \theta \in [0, 2\pi]\}$, any unitary representation is given by $e^{\lambda i\theta}$ by using an integer λ. Then, we have $L^2(\hat{G}) = L^2(\mathbb{Z})$. So, the initial state is given as a vector state $\varphi(\lambda)$ on $L^2(\mathbb{Z})$. When the true parameter is 0, the probability distribution of the estimate $\hat{\theta} \in \mathbb{R}$ is given as $|\mathcal{F}^{-1}[\varphi](-\hat{\theta})|^2 d\hat{\theta}$ by using the inverse Fourier transform $\mathcal{F}^{-1}[\varphi]$ of φ. When there is no restriction for the support of $\varphi(\lambda)$, we can decrease the error unboundedly. Hence, we restrict the support of the initial state to $\{k \in \mathbb{Z} | |k| \le n\}$.

Consider the case when the gain function $\tilde{R}(\theta, \hat{\theta})$ is $\cos(\theta - \hat{\theta}) = (e^{i(\theta-\hat{\theta})} + e^{i(\theta-\hat{\theta})})/2 = 1 - 2\sin^2(\frac{\theta-\hat{\theta}}{2})$, i.e., the error function is $2\sin^2(\frac{\theta-\hat{\theta}}{2})$. When the initial state is $(\varphi(\lambda))_{\lambda=-n}^{n}$, the relation (4.36) guarantees that the gain is

$$\sum_{\lambda=-n}^{n-1} \frac{1}{2}\varphi(\lambda)\varphi(\lambda+1) + \sum_{\lambda=-n+1}^{n} \frac{1}{2}\varphi(\lambda-1)\varphi(\lambda) = \sum_{\lambda=-n}^{n-1} \varphi(\lambda)\varphi(\lambda+1).$$

Due to Theorem 4.9, this value is maximized by $\varphi(\lambda) = \sin \frac{\pi(\lambda+n+1)}{2n+2}$, and the maximum is $\cos \frac{\pi}{2n+2}$.

On the other hand, under the limit $n \to \infty$, the initial state $\varphi_n(\lambda)$ is approximately given as $\varphi_n(\lambda) = \frac{1}{\sqrt{n}} \hat{\varphi}(\lambda/n)$ by using a square integrable function $\hat{\varphi} \in L^2([-1, 1])$ on the closed interval $[-1, 1]$. When we input this initial state and the true parameter is 0, the estimate $\hat{\theta}$, which is a random variable, converges to 0. Since the random variable $x := n\hat{\theta}$ satisfies $dx = nd\hat{\theta}$, using $y = -\lambda/n$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} |\mathcal{F}^{-1}[\varphi_n](-nx)|^2 dx = \lim_{n\to\infty} \frac{1}{n} \Big| \sum_{\lambda=-n}^{n} \frac{1}{\sqrt{n}} \hat{\varphi}(\lambda/n) e^{-i\frac{x}{n}\lambda} \Big|^2 dx \\
=& \Big| \lim_{n\to\infty} \frac{1}{n} \sum_{\lambda=-n}^{n} \hat{\varphi}(\lambda/n) e^{-i\frac{x}{n}\lambda} |^2 dx = \Big| \int_{-1}^{1} \hat{\varphi}(y) e^{-ixy} dy \Big|^2 dx \\
=& 2\pi |\mathcal{F}^{-1}[\hat{\varphi}](-x)|^2 dx.
\end{aligned} \tag{4.37}$$

So, the probability distribution of the x is given by the inverse Fourier transform $\mathcal{F}^{-1}[\hat{\varphi}]$ of $\hat{\varphi}$. Since the error function can be approximated to $2\sin^2(\frac{\theta-\hat{\theta}}{2})n^2 \cong \frac{(\theta-\hat{\theta})^2 n^2}{2} = x^2/2$, this problem is reduced to the case when $G = \mathbb{R}$, the support of a function $\varphi \in L^2(\mathbb{R})$ is restricted to the closed interval $[-1, 1]$, and the error function is the quadratic function x^2.

Exercise 4.10 Consider the case when $\varphi(x) = \frac{1}{\sqrt{2}}$ on $[-1, 1]$. Calculate $\lim_{n\to\infty} \frac{1}{n}|\mathcal{F}^{-1}[\varphi_n](-nx)|^2$.

Exercise 4.11 Calculate the asymptotic behavior of the optimal gain with the limit $n \to \infty$ when the gain function $\tilde{R}(\theta, \hat{\theta})$ is $\cos(\theta - \hat{\theta})$.

4.3.3 Role of Inverse Fourier Transform for Non-commutative Case

To extend the discussion in Sect. 4.3.2 to the non-commutative case, we denote the linear space composed of matrices (or Hilbert Schmidt operators) on $\mathcal{U}_\lambda$ by $\mathfrak{gl}(\mathcal{U}_\lambda)$, and focus on the space $L^2(\hat{G}) := \oplus_\lambda \mathfrak{gl}(\mathcal{U}_\lambda)$ (For the details of this notation, see [44, Sect. 3.8.2].) Then, we introduce the inverse Fourier transform $\mathcal{F}^{-1}[X] := \sum_{\lambda\in\hat{G}} \sqrt{d_\lambda}\, \mathrm{Tr}\, \mathsf{f}_\lambda(\hat{g}^{-1})X_\lambda$ for a matrix-valued vector X in $L^2(\hat{G})$.

Using the vector $|I\rangle := \sum_{\lambda\in S} |I\rangle\rangle$, we define the POVM $M_{|I\rangle\langle I|}$. So, we have

$$\begin{aligned}
&\mathcal{D}_R(|\phi\rangle) := \mathcal{D}_R(|\phi\rangle\langle\phi|, M_{|I\rangle\langle I|}) = \int_G R(e, \hat{g})|\langle I|f(\hat{g})^\dagger|\phi\rangle|^2 \mu_G(d\hat{g}) \\
=& \int_G R(e, \hat{g})|\langle I|f(\hat{g})^\dagger|\phi\rangle|^2 \mu_G(d\hat{g}) \\
=& \int_G R(e, \hat{g})\left|\int_\Lambda \mathrm{Tr}\, f(\hat{g})^\dagger \phi_\lambda \mu_{\hat{G}[\mathcal{L}]}(d\lambda)\right|^2 \mu_G(d\hat{g}) \\
=& \int_G R(e, \hat{g})|\mathcal{F}^{-1}[\phi](\hat{g})|^2 \mu_G(d\hat{g}).
\end{aligned} \tag{4.38}$$

Notice that the quantity $\mathcal{D}_R(|\phi\rangle)$ can be regarded as a non-commutative extension of $\mathcal{D}_R(|\phi\rangle)$ defined in Sect. 4.3.2.

Also, we define the subset $L^2(\hat{G}) \wedge \mathcal{H}$ of $L^2(\hat{G})$ as the set of elements $X = (X_\lambda)_{\lambda^\dagger}$ of $L^2(\hat{G})$ satisfying that the rank of the matrix X_λ is not greater than the multiplicity of the irreducible representation $\mathcal{U}_\lambda$ in the representation space $\mathcal{H}$. Note that the set $L^2(\hat{G}) \wedge \mathcal{H}$ is not necessarily a vector space in general. Using the subset $L^2(\hat{G}) \wedge \mathcal{H}$, we can show the following theorem.

Theorem 4.10 ([62]) *Let f be a unitary representation of a unimodular group G to a Hilbert space $\mathcal{H}$.*[1] *Then, we obtain*

$$\begin{aligned}
&\min_{\rho\in\mathcal{S}(\mathcal{H})} \min_{M\in\mathcal{M}_{\mathrm{cov}}(G)} \mathcal{D}_R(\rho, M) = \min_{\{p_i\}} \min_{\rho_i\in\mathcal{S}(\mathcal{H})} \min_{M_i\in\mathcal{M}_{\mathrm{cov}}(G)} \sum_i p_i \mathcal{D}_R(\rho_i, M_i) \\
=& \min_{|\phi\rangle\in L^2(\hat{G})\wedge\mathcal{H}:\|\phi\|=1} \mathcal{D}_R(|\phi\rangle),
\end{aligned}$$

[1] In topological group theory, a locally compact Hausdorff topological group is called *unimodular* when the left invariant measure is equal to the right invariant measure.

In the projective representation case with the factor system $\mathcal{E}$, we have the same statement [62]. To clarify the dependence of the factor system $\mathcal{E}$, we denote the inverse Fourier transform by $\mathcal{F}_{\mathcal{E}}^{-1}$.

Proof To show the theorem, we remember that any covariant POVM is written with as M_T. Here, the matrix $T \geq 0$ is written as $T = \sum_k |\eta_k\rangle\langle\eta_k|$ in general. When the true element is the unit element, since G is a unimodular group, the distribution of the estimate $\hat{g}$ is calculated as

$$\sum_k |\langle\eta_k|\mathsf{f}(\hat{g}^{-1})|x\rangle|^2 \mu_G(d\hat{g}) = \sum_k \Big|\sum_{\lambda\in S} \operatorname{Tr} \eta_{k,\lambda}^\dagger \mathsf{f}_\lambda(\hat{g}^{-1}) X_\lambda\Big|^2 \mu_G(d\hat{g})$$
$$= \sum_k \Big|\sum_{\lambda\in S} \operatorname{Tr} \mathsf{f}_\lambda(\hat{g}^{-1}) X_\lambda \eta_{k,\lambda}^\dagger\Big|^2 \mu_G(d\hat{g}) = \sum_k |\langle I|\mathsf{f}(\hat{g}^{-1})|x_k\rangle\rangle|^2 \mu_G(d\hat{g}),$$

where $|x\rangle = \sum_\lambda |X_\lambda\rangle\rangle$, $|\eta_k\rangle = \sum_\lambda |\eta_{k,\lambda}\rangle\rangle$, $|x_k\rangle := \sum_\lambda \frac{1}{\sqrt{d_\lambda}} |X_\lambda \eta_{k,\lambda}^\dagger\rangle\rangle$, and $|I\rangle := \sum_\lambda \sqrt{d_\lambda}|I_\lambda\rangle\rangle$. The relation (4.15) yields that

$$\sum_k \operatorname{Tr} |x_k\rangle\langle x_k| = \sum_k \sum_\lambda \frac{1}{d_\lambda} \operatorname{Tr} \eta_{k,\lambda} X_\lambda^\dagger X_\lambda \eta_{k,\lambda}^\dagger$$
$$= \sum_\lambda \operatorname{Tr} X_\lambda^\dagger X_\lambda \sum_k \frac{1}{d_\lambda} \eta_{k,\lambda}^\dagger \eta_{k,\lambda} = \sum_\lambda \operatorname{Tr} X_\lambda^\dagger X_\lambda I_{n_\lambda} = \operatorname{Tr} |x\rangle\langle x| = 1.$$

Hence, the error is the error when the initial state is the mixed state $\sum_k |x_k\rangle\langle x_k|$ and the measurement is the covariant POVM generated by the matrix $|I\rangle\langle I|$.

Since $\sum_k \langle x_k|x_k\rangle \frac{1}{\langle x_k|x_k\rangle} |x_k\rangle\langle x_k| = \sum_k |x_k\rangle\langle x_k|$, choosing a suitable k, we have

$$\mathcal{D}_{R,\mu_G}\Big(\frac{1}{\langle x_k|x_k\rangle}|x_k\rangle\langle x_k|, M_{|I\rangle\langle I|}\Big) \leq \mathcal{D}_{R,\mu_G}\Big(\sum_k |x_k\rangle\langle x_k|, M_{|I\rangle\langle I|}\Big)$$
$$= \mathcal{D}_{R,\mu_G}(|x\rangle\langle x|, M_T).$$

As the rank of the matrix $X_\lambda \eta_{k,\lambda}^\dagger$ is at most n_λ, the pure state $\frac{1}{\langle x_k|x_k\rangle}|x_k\rangle\langle x_k|$ can be regarded as a pure state on the system $\mathcal{H}$. Hence, we restrict our measurement to a covariant measurement generated by $|I\rangle\langle I|$, and our initial state to a pure state $|x\rangle$ on $\mathcal{H}$. So, we obtain the desired statement. ■

4.3.4 Finite Group

Now, we investigate the matrix-valued vector $X \in L^2(\hat{G}) \wedge \mathcal{H}$ corresponding to the optimal initial state on $\mathcal{H} = \oplus_{\lambda\in S} \mathcal{U}_\lambda \otimes \mathbb{C}^{n_\lambda}$ when G is a finite group and the error function R is given as

$$R(g, \hat{g}) := \begin{cases} 1 \text{ if } \hat{g} \neq g \\ 0 \text{ if } \hat{g} = g. \end{cases}$$

Then, the probability to correctly identify the unknown operation is

$$\mathcal{D}_{1-R}(X) = \frac{|\mathcal{F}^{-1}[X](e)|^2}{|G|} = \frac{|\sum_{\lambda \in S} \sqrt{d_\lambda} \operatorname{Tr} X_\lambda|^2}{|G|}.$$

Hence, Schwarz inequality with respect to the space $L^2(\hat{G})$ implies that

$$\begin{aligned} &\frac{|\sum_{\lambda \in S} \sqrt{d_\lambda} \operatorname{Tr} X_\lambda|^2}{|G|} = \frac{|\sum_{\lambda \in S} \operatorname{Tr} \sqrt{d_\lambda} I_{\min\{d_\lambda, n_\lambda\}} X_\lambda|^2}{|G|} \\ \leq &\frac{(\sum_{\lambda \in S} \operatorname{Tr} d_\lambda I_{\min\{d_\lambda, n_\lambda\}})(\sum_{\lambda \in S} \operatorname{Tr} X_\lambda^\dagger X_\lambda)}{|G|} = \frac{\sum_{\lambda \in S} d_\lambda \min\{d_\lambda, n_\lambda\}}{|G|}. \end{aligned}$$

The equality holds if and only if X is $(\frac{1}{\sum_{\lambda \in S} d_\lambda \min\{d_\lambda, n_\lambda\}} \sqrt{d_\lambda} I_{\min\{d_\lambda, n_\lambda\}})$. That is, the optimal distinguishing probability is calculated to be [41]

$$\mathcal{D}(S, \{n_\lambda\}_{n \in S}) := \frac{\sum_{\lambda \in S} d_\lambda \min\{d_\lambda, n_\lambda\}}{|G|}. \tag{4.39}$$

In particular, when all of irreducible representations $\lambda \in \hat{G}$ are prepared with at least multiplicity d_λ, we can identify the unknown operation with probability 1. On the other hand, when the multiplicity d_λ is less than n_λ, we have

$$\mathcal{D}(S, \{n_\lambda\}_{n \in S}) = \frac{\dim \mathcal{H}}{|G|}. \tag{4.40}$$

Example 4.11 (*Dense coding [8, 39]*) Consider discrete Heisenberg representation $\mathsf{W}_{\mathbb{F}}^r$ of the group $\mathbb{F}_q^{2r}$. When the multiplicity is 1, the optimal distinguishing probability is calculated to be

$$\frac{q^r}{q^{2r}} = \frac{1}{q^r}. \tag{4.41}$$

When the group G is $\mathbb{Z}_d^{2r}$, this value is $\frac{d^r}{d^{2r}} = \frac{1}{d^r}$.

On the other hand, when the reference system is available, the multiplicity can be increased up to the dimension of the representation space. So, the optimal distinguishing probability for the group $\mathbb{F}_q^{2r}$ is calculated to be

$$\frac{q^{2r}}{q^{2r}} = 1. \tag{4.42}$$

For the group $\mathbb{Z}_d^{2r}$, this value is $\frac{d^{2r}}{d^{2r}} = 1$. This fact shows that the use of reference system dramatically improves the optimal distinguishing probability.

This mathematical fact can be applied to the communication as follows. Firstly, we prepare the maximally entangled state, and the sender sends the reference system to the receiver priorly. Second, according to the message to be transmitted, the sender applies the operation described by the representation of the group $\mathbb{F}_q^{2r}$ or $\mathbb{Z}_d^{2r}$ on the representation space. Finally, the sender transmits the representation space to the receiver. Measuring the composite system with the optimal measurement, the receiver recovers the intended message. In this way, the transmission of information with size q^{2r} (d^{2r}) can be realized by transmission of the physical system with the dimension q^r (d^r).

In the following, for simplicity, when we deal with common topics with respect to the groups $\mathbb{F}_q^{2r}$ and $\mathbb{Z}_d^{2r}$, we denote the group by $\mathbb{X}$. For example, when $\mathbb{X} = \mathbb{F}$, a projective unitary representation $\mathsf{W}_{\mathbb{X}}^r$ is a projective unitary representation $\mathsf{W}_{\mathbb{F}}^r$, and when $\mathbb{X} = \mathbb{Z}$, it is a projective unitary representation $\mathsf{W}_{\mathbb{Z}}^r$. We denote the projective unitary representation whose factor system is the square of the factor system of $\mathsf{W}_{\mathbb{X}}^r$ by $\mathsf{W}_{\mathbb{X}:2}^r$. Then, when q or d is odd, the irreducible decomposition of the projective unitary representation $(\mathsf{W}_{\mathbb{X}}^r)^{\otimes 2}$ is composed of the representation $\mathsf{W}_{\mathbb{X}:2}^r$ with multiplicity $|\mathbb{X}|^r$. Hence, applying the above discussion to the projective unitary representation $\mathsf{W}_{\mathbb{X}:2}^r$, we can identify the action of $\mathbb{X}^r$ with probability 1 when the projective unitary representation $(\mathsf{W}_{\mathbb{X}}^r)^{\otimes 2}$ is given.

Example 4.12 (*Discrete symplectic group*) We focus on Metaplectic representation $\mathsf{S}_{\mathbb{F}_q}^r$ of discrete symplectic group $\mathrm{Sp}(2r, \mathbb{F}_q)$. When the multiplicity is 1, due to (4.40), the optimal distinguishing probability is

$$\frac{q^r}{q^{r^2}\prod_{k=1}^r (q^{2k}-1)} = \frac{1}{q^{r^2-r}\prod_{k=1}^r (q^{2k}-1)} \tag{4.43}$$

because the order of $\mathrm{Sp}(2r, \mathbb{F}_q)$ is $q^{r^2}\prod_{k=1}^r (q^{2k}-1)$. In particular, when $r = 1$, we have $\frac{1}{q^2-1}$. On the other hand, when the reference system is available, the multiplicity can be increased up to the dimension of the representation space. So, when $p \neq 2$, the optimal distinguishing probability for the group $\mathrm{Sp}(2r, \mathbb{F}_q)$ is calculated to be

$$\frac{(\frac{q^r+1}{2})^2 + (\frac{q^r-1}{2})^2}{q^{r^2}\prod_{k=1}^r (q^{2k}-1)} = \frac{q^{2r}+1}{2q^{r^2}\prod_{k=1}^r (q^{2k}-1)}. \tag{4.44}$$

In particular, when $r = 1$, this value is calculated to be $\frac{q^2+1}{2(q^2-1)}$. When $p = 2$, this value is written as

$$\frac{q^{2r}}{q^{r^2}\prod_{k=1}^r (q^{2k}-1)} = \frac{1}{q^{r^2-2r}\prod_{k=1}^r (q^{2k}-1)}. \tag{4.45}$$

In particular, when $r = 1$, we have $\frac{q}{q^2-1}$. When $q = 2$, the group $\mathrm{Sp}(2r, \mathbb{F}_q)$ is the symmetric group of order 3, i.e., S_3 and this value is $\frac{2}{3}$.

Example 4.13 (*Clifford group on* $\mathbb{F}_q^{2r}$) We consider the representation $\mathsf{V}^r_{\mathbb{F}} \otimes \overline{\mathsf{V}^r_{\mathbb{F}}}$ of Clifford group $\mathbb{F}_q^{2r} \rtimes \mathrm{Sp}(2r, \mathbb{F}_q)$, where $\bar{\mathsf{f}}$ is the complex conjugate representation of f. When the multiplicity is 1, due to (4.40), the optimal distinguishing probability is

$$\frac{q^{2r}}{q^{r^2+2}\prod_{k=1}^{r}(q^{2k}-1)} = \frac{1}{q^{r^2}\prod_{k=1}^{r}(q^{2k}-1)}. \tag{4.46}$$

In particular, when $r = 1$, this value is $\frac{1}{q(q^2-1)}$. On the other hand, when the reference system is available, the multiplicity can be increased up to the dimension of the representation space. So, the optimal distinguishing probability for Clifford group $\mathbb{F}_q^{2r} \rtimes \mathrm{Sp}(2r, \mathbb{F}_q)$ is calculated to be

$$\frac{1+(q^{2r}-1)^2}{q^{r^2+2}\prod_{k=1}^{r}(q^{2k}-1)}. \tag{4.47}$$

In particular, when $r = 1$, this value is $\frac{q^4-2q^2+2}{q^3(q^2-1)}$.

In the case of the representation $\mathsf{V}^r_{\mathbb{F}} \otimes \mathsf{V}^r_{\mathbb{F}}$, the optimal distinguishing probability equals (4.46) when the multiplicity is 1. However, when the reference system is available, i.e., the multiplicity can be increased up to the dimension of the representation space, it is improved to

$$\frac{(\frac{q^r(q^r+1)}{2})^2 + (\frac{q^r(q^r-1)}{2})^2}{q^{r^2+2}\prod_{k=1}^{r}(q^{2k}-1)} = \frac{q^{2r}+1}{2q^{r^2}\prod_{k=1}^{r}(q^{2k}-1)}. \tag{4.48}$$

In particular, when $r = 1$, it is $\frac{q^2+1}{2q(q^2-1)}$.

Example 4.14 (*direct product group* $\mathbb{F}_q^{2r} \times \mathrm{Sp}(\mathbb{F}_q, 2r)$) To compare the representation $\mathsf{V}^r_{\mathbb{F}} \otimes \overline{\mathsf{V}^r_{\mathbb{F}}}$, $\mathsf{V}^r_{\mathbb{F}} \otimes \mathsf{V}^r_{\mathbb{F}}$, we consider the representations $\mathsf{W}^r_{\mathbb{F}} \overline{\otimes} \mathsf{S}^r_{\mathbb{F}}$ $\mathsf{W}^r_{\mathbb{F}} \overline{\otimes} \overline{\mathsf{S}^r_{\mathbb{F}}}$, $\overline{\mathsf{W}^r_{\mathbb{F}}} \overline{\otimes} \mathsf{S}^r_{\mathbb{F}}$, and $\overline{\mathsf{W}^r_{\mathbb{F}}} \overline{\otimes} \overline{\mathsf{S}^r_{\mathbb{F}}}$ of the direct product group $\mathbb{F}_q^{2r} \times \mathrm{Sp}(\mathbb{F}_q, 2r)$. (For the definitions of these representation, see [44, Sect. 8.3].) The optimal distinguishing probability is decided by the order of the group and the dimensions of the irreducible components as shown in (4.39). That is, to discuss this issue, we only need the division of the dimension of the whole representation by the dimensions of the irreducible components. Although these four representations have different irreducible decompositions, they have the same division of the dimension of the whole representation. In these cases, due to (4.40), the optimal distinguishing probability is (4.46). So, there is no difference for this probability among $\mathsf{V}^r_{\mathbb{F}} \otimes \overline{\mathsf{V}^r_{\mathbb{F}}}$, $\mathsf{V}^r_{\mathbb{F}} \otimes \mathsf{V}^r_{\mathbb{F}}$, $\mathsf{W}^r_{\mathbb{F}} \overline{\otimes} \mathsf{S}^r_{\mathbb{F}}$, $\mathsf{W}^r_{\mathbb{F}} \overline{\otimes} \overline{\mathsf{S}^r_{\mathbb{F}}}$, $\overline{\mathsf{W}^r_{\mathbb{F}}} \overline{\otimes} \mathsf{S}^r_{\mathbb{F}}$, and $\overline{\mathsf{W}^r_{\mathbb{F}}} \overline{\otimes} \overline{\mathsf{S}^r_{\mathbb{F}}}$.

Indeed, this distinguishing problem is the same as the problem of separately distinguishing elements of $\mathbb{F}_q^{2r}$ and $\mathrm{Sp}(\mathbb{F}_q, 2r)$ when two groups $\mathbb{F}_q^{2r}$ and $\mathrm{Sp}(\mathbb{F}_q, 2r)$ act on the different system independently, which can be shown as follows. When we distinguish both separately, the distinguishing probability is the product of both distinguishing probabilities, which can be calculated to be the same as (4.46). So, this fact shows that even though we do not utilize the correlation between these two representation, we can attain the optimal distinguishing probability.

On the other hand, when the reference system is available, the multiplicity can be increased up to the dimension of the representation space. Since (4.42) is 1, the optimal distinguishing probability is (4.44) when $p \neq 2$, and it is (4.45) when $p = 2$. When $p \neq 2$, (4.48) equals (4.44). So, the optimal distinguishing probability in $\mathsf{W}^r_{\mathbb{F}} \overline{\otimes} \mathsf{S}^r_{\mathbb{F}}$ equals that in $\mathsf{V}^r_{\mathbb{F}} \otimes \mathsf{V}^r_{\mathbb{F}}$. However, since

$$\frac{1 + (q^{2r} - 1)^2}{q^{r^2+2} \prod_{k=1}^{r} (q^{2k} - 1)} > \frac{q^{2r} + 1}{2q^{r^2} \prod_{k=1}^{r} (q^{2k} - 1)}, \tag{4.49}$$

the representation $\mathsf{V}^r_{\mathbb{F}} \otimes \overline{\mathsf{V}^r_{\mathbb{F}}}$ has a strictly better optimal distinguishing probability.

When $p = 2$, there is the following relation among (4.45), (4.47), and (4.48);

$$\frac{1}{q^{r^2-2r} \prod_{k=1}^{r} (q^{2k} - 1)} > \frac{1 + (q^{2r} - 1)^2}{q^{r^2+2} \prod_{k=1}^{r} (q^{2k} - 1)} > \frac{q^{2r} + 1}{2q^{r^2} \prod_{k=1}^{r} (q^{2k} - 1)}. \tag{4.50}$$

Hence, the representation $\mathsf{W}^r_{\mathbb{F}} \overline{\otimes} \mathsf{S}^r_{\mathbb{F}}$ has the best optimal distinguishing probability, and the representation $\mathsf{V}^r_{\mathbb{F}} \otimes \overline{\mathsf{V}^r_{\mathbb{F}}}$ has the second best optimal distinguishing probability. That is, the representation $\mathsf{V}^r_{\mathbb{F}} \otimes \mathsf{V}^r_{\mathbb{F}}$ has the worst optimal distinguishing probability.

Example 4.15 (*Clifford group on* $\mathbb{Z}_d^2$) We focus on the representation $\mathsf{V}^r_{\mathbb{Z}} \otimes \overline{\mathsf{V}^r_{\mathbb{Z}}}$ of Clifford group $\mathbb{Z}_d^2 \rtimes \mathrm{Sp}(2, \mathbb{Z}_d)$. Now, for simplicity, we assume that d is a power p^k of a prime p. When the multiplicity is 1, the optimal distinguishing probability is $\frac{p^{2k}}{p^{5k}(1-p^{-2})}$.

On the other hand, when the reference system is available, i.e., the multiplicity can be increased up to the dimension of the representation space, the optimal distinguishing probability is

$$\begin{aligned} &\frac{p^{4(k-1)}(p^2-1)^2 + p^{4(k-2)}(p^2-1)^2 + \cdots + (p^2-1)^2 + 1}{p^{5k}(1-p^{-2})} \\ =& \frac{\frac{(p^2-1)(p^{4k}-1)}{p^2+1} + 1}{p^{5k}(1-p^{-2})} = \frac{p^{4k}(p^2-1)+2}{p^{5k-2}(p^4-1)}. \end{aligned}$$

Example 4.16 (*permutation group [39, 41, 86]*) Next, we consider the case when the permutation group S_n acts on the tensor product space $(\mathbb{C}^d)^{\otimes n}$ as the permutation of the order of tensor product. We assume that the reference system is additionally prepared so that the multiplicity can be increased up to the dimension of the irreducible representation spaces. When the initial state is the measurement are optimal, the optimal distinguishing probability is Plancherel measure $\mu_n(d)$, due to the definition of Plancherel measure [44, Sect. 2.9.3]. Hence, when n is sufficiently large, the optimal distinguishing probability is close to 1 even when the dimension d is almost $2\sqrt{n}$ (See [44, (2.75)], which gives more detailed behavior of the optimal distinguishing probability with respect to d).

4.3.5 General Compact Lie Group

In the following, we investigate the matrix-valued vector $X \in L^2(\hat{G}) \wedge \mathcal{H}$ corresponding to the optimal initial state for the case of a general compact Lie group. For this purpose, we focus on the irreducible decomposition $\mathcal{H} = \oplus_{\boldsymbol{\lambda} \in S} \mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathbb{C}^{n_{\boldsymbol{\lambda}}}$ of the representation space $\mathcal{H}$, and assume that $n_{\boldsymbol{\lambda}} \ge d_{\boldsymbol{\lambda}}$ for any $\boldsymbol{\lambda} \in S$, where $S := \{\boldsymbol{\lambda} \in \hat{G} | n_{\boldsymbol{\lambda}} \neq 0\}$.

Then, for an element $g \in G$, we define the matrix-valued vector $X_g := (\mathsf{f}_{\boldsymbol{\lambda}}(g) X_{\boldsymbol{\lambda}} \mathsf{f}_{\boldsymbol{\lambda}}(g^{-1}))_{\boldsymbol{\lambda}} \in L^2(\hat{G})$. The relation (4.35) yields that

$$\begin{aligned}\mathcal{D}_R(X) = &\int_G R(e, \hat{g}) |\mathcal{F}^{-1}[X](\hat{g}^{-1})|^2 \mu_G(d\hat{g}) \\ = &\int_G R(e, g^{-1}\hat{g}g) |\mathcal{F}^{-1}[X](\hat{g}^{-1})|^2 \mu_G(d\hat{g}) \\ = &\int_G R(e, \hat{g}) |\mathcal{F}^{-1}[X](g\hat{g}^{-1}g^{-1})|^2 \mu_G(d\hat{g}) \\ = &\int_G R(e, \hat{g}) |\mathcal{F}^{-1}[X_{g^{-1}}](\hat{g}^{-1})|^2 \mu_G(d\hat{g}) = \mathcal{D}_R(X_{g^{-1}}). \end{aligned} \tag{4.51}$$

In the following, using the characters of non-trivial irreducible representations, we assume that the error function R is written as

$$R(g, \hat{g}) = -\sum_{\boldsymbol{\lambda} \in \hat{G}} a_{\boldsymbol{\lambda}} (\chi_{\boldsymbol{\lambda}}(\hat{g}g^{-1}) + \chi_{\boldsymbol{\lambda}^*}(\hat{g}g^{-1})), \quad a_{\boldsymbol{\lambda}} = a_{\boldsymbol{\lambda}^*} \ge 0. \tag{4.52}$$

This problem is the same as the maximize the average gain function:

$$\tilde{R}(g, \hat{g}) = \sum_{\boldsymbol{\lambda} \in \hat{G}} a_{\boldsymbol{\lambda}} (\chi_{\boldsymbol{\lambda}}(\hat{g}g^{-1}) + \chi_{\boldsymbol{\lambda}^*}(\hat{g}g^{-1})), \quad a_{\boldsymbol{\lambda}} = a_{\boldsymbol{\lambda}^*} \ge 0. \tag{4.53}$$

For example, when we choose the gain function $R(g, \hat{g})$ to be $\delta_{g,\hat{g}}$, the gain average function is the correctly distinguishing probability.

When $G = \mathrm{SU}(d)$, we often choose the gate fidelity $|\operatorname{Tr} g^\dagger \hat{g}|^2$ as the gain function. When $d = 2$, this function is calculated to be $1 + \chi_1(g) = 1 + \chi_{(2,0)}(g) = 1 + \chi_{[2]}(g)$. On the other hand, when $d > 2$, this value is calculated to be $1 + \chi_{[1,0,\ldots,0,1]}(g)$, so the condition (4.53) holds.

Then, we have the following theorem by using the entangled state $|\Psi_{\boldsymbol{\lambda}}\rangle\rangle := \frac{1}{\sqrt{d_{\boldsymbol{\lambda}}}} \sum_{j=1}^{d_{\boldsymbol{\lambda}}} |\boldsymbol{\lambda}; j; j\rangle$ on the space $\mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathbb{C}^{d_{\boldsymbol{\lambda}}}$, where $|\boldsymbol{\lambda}; j; j\rangle = |\boldsymbol{\lambda}; j\rangle \otimes |j\rangle$ and $\{|\boldsymbol{\lambda}; j\rangle\}_j$ and $\{|j\rangle\}_j$ are CONSs of $\mathcal{U}_{\boldsymbol{\lambda}}$ and $\mathbb{C}^{d_{\boldsymbol{\lambda}}}$, respectively.

Theorem 4.11 ([20]) *Given a matrix-valued vector $X \in L^2(\hat{G})$, we choose a matrix $\Phi_{\boldsymbol{\lambda}}$ on the representation space corresponding to $\boldsymbol{\lambda} \in \hat{G}$ and a non-negative real number $c_{\boldsymbol{\lambda}} \ge 0$ such that $X = (c_{\boldsymbol{\lambda}}\Phi_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda}}$ and $\mathrm{Tr}\, \Phi_{\boldsymbol{\lambda}}^{\dagger}\Phi_{\boldsymbol{\lambda}} = 1$. When the error function is given by (4.52), the relation*

$$\begin{aligned}\mathcal{D}_R(X) \ge \mathcal{D}_R(\{c_{\boldsymbol{\lambda}}\}) &:= -\sum_{\boldsymbol{\lambda},\boldsymbol{\lambda}' \in S} c_{\boldsymbol{\lambda}} c_{\boldsymbol{\lambda}'} \sum_{\boldsymbol{\lambda}'' \in \hat{G}} a_{\boldsymbol{\lambda}''} (C^{\boldsymbol{\lambda}''*}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'*} + C^{\boldsymbol{\lambda}''}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'*}) \\ &= -\sum_{\boldsymbol{\lambda}'' \in \hat{G}} a_{\boldsymbol{\lambda}''} \sum_{\boldsymbol{\lambda},\boldsymbol{\lambda}' \in S} (C^{\boldsymbol{\lambda}'}_{\boldsymbol{\lambda}'',\boldsymbol{\lambda}} + C^{\boldsymbol{\lambda}'}_{\boldsymbol{\lambda}''*,\boldsymbol{\lambda}}) c_{\boldsymbol{\lambda}} c_{\boldsymbol{\lambda}'} \end{aligned} \tag{4.54}$$

holds. When $X = (c_{\boldsymbol{\lambda}}\Psi_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda}}$, The equality holds and $\mathcal{F}^{-1}[X] = \sum_{\boldsymbol{\lambda} \in S} c_{\boldsymbol{\lambda}} \chi_{\boldsymbol{\lambda}}$. Here, $C^{\boldsymbol{\lambda}'}_{\boldsymbol{\lambda}'',\boldsymbol{\lambda}}$ is the multiplicity coefficient and its detail definition is given in [44, (2.45)].

Hence, when the representation space is $\oplus_{\boldsymbol{\lambda} \in S} \mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathbb{C}^{d_{\boldsymbol{\lambda}}}$, the optimization of the identification method is reduced to the maximization of $\mathcal{D}_R(\{c_{\boldsymbol{\lambda}}\})$ with respect to $c = (c_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda} \in S}$. Here, when the initial state is $X = (c_{\boldsymbol{\lambda}}\Psi_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda}}$, to attain the bound $\mathcal{D}_R(\{c_{\boldsymbol{\lambda}}\})$ with the Bayesian method, we do not need to employ the covariant measurement. To discuss this issue, we prepare the following theorem.

Theorem 4.12 ([44, Theorem 4.4]) *The following conditions are equivalent for a measure ν on G and a representation $\mathsf{f}_{\boldsymbol{\lambda}}$ of G, where all of irreducible components are assumed to have formal dimension with respect to the measure ν when the representation space $\mathcal{H}$ is infinite-dimensional.*

(1) $\int_G \mathsf{f}_{\boldsymbol{\lambda}}(g) \otimes \mathsf{f}_{\boldsymbol{\lambda}}(g)^{\dagger} \mu_G(dg) = \int_G \mathsf{f}_{\boldsymbol{\lambda}}(g) \otimes \mathsf{f}_{\boldsymbol{\lambda}}(g)^{\dagger} \nu(dg)$.

(2) For any matrix X, the relation $\int_G \mathsf{f}_{\boldsymbol{\lambda}}(g) X \mathsf{f}_{\boldsymbol{\lambda}}(g)^{\dagger} \mu_G(dg) = \int_G \mathsf{f}_{\boldsymbol{\lambda}}(g) X \mathsf{f}_{\boldsymbol{\lambda}}(g)^{\dagger} \nu(dg)$ holds. When the representation space is infinite-dimensional, we need to additionally impose the trace class condition to X.

(3) For any Hermitian matrix X, the relation $\int_G \mathsf{f}_{\boldsymbol{\lambda}}(g) X \mathsf{f}_{\boldsymbol{\lambda}}(g)^{\dagger} \mu_G(dg) = \int_G \mathsf{f}_{\boldsymbol{\lambda}}(g) X \mathsf{f}_{\boldsymbol{\lambda}}(g)^{\dagger} \nu(dg)$ holds. When the representation space is infinite-dimensional, we need to additionally impose the trace class condition to X as well as Hermiteness.

Then, a measure ν on G is called an S-design when the above conditions hold for any element $\lambda \in S$. In particular, a finite subgroup G' is called a S-design when the uniform distribution $\mu_{G'}$ on G' is an S-design. When a probability measure ν on G is an S-design, the integral $\int_G \mathsf{f}(\hat{g})|I\rangle\langle I|\mathsf{f}(\hat{g})^{-1}\nu(d\hat{g}) = I$ forms a POVM and the POVM is denoted by $M(\nu)$. So, we can consider the identification method with the initial state $X = (c_{\boldsymbol{\lambda}}\Psi_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda}}$ and the measurement $M(\nu)$.

Now, we define $S'' := \{\boldsymbol{\lambda}'' \in \hat{G} | \boldsymbol{\lambda}'' \neq 0, a_{\boldsymbol{\lambda}''} > 0\}$, and $S' := \{\boldsymbol{\lambda}' \in S | \exists \boldsymbol{\lambda} \in S, \exists \boldsymbol{\lambda}'' \in S'', C^{\boldsymbol{\lambda}'}_{\boldsymbol{\lambda},\boldsymbol{\lambda}''} \neq 0\}$. When a probability measure ν is an $S \cup S'$-design, Lemma 4.2 and (3.78) in [44] yield that

$$
\begin{aligned}
&\mathcal{D}_{R,g}(M(\nu), |X\rangle) \\
&= \int_{G'} -\sum_{\lambda'' \in S''} a_{\lambda''}(\chi_{\lambda''}(\hat{g}' g^{-1}) + \chi_{\lambda''^*}(\hat{g}' g^{-1}))\Big|\sum_{\lambda \in S} c_{\lambda}\chi_{\lambda}(\hat{g}' g^{-1})\Big|^2 \mu_{G'}(d\hat{g}') \\
&= -\sum_{\lambda,\lambda' \in S} c_{\lambda} c_{\lambda'} \sum_{\lambda'' \in S''} a_{\lambda''}(C^{\lambda'^*}_{\lambda,\lambda''} + C^{\lambda'}_{\lambda,\lambda''}) \\
&= -\sum_{\lambda,\lambda' \in S} c_{\lambda} c_{\lambda'} \sum_{\lambda'' \in S''} a_{\lambda''}(C^{\lambda'}_{\lambda,\lambda''^*} + C^{\lambda'}_{\lambda,\lambda''}).
\end{aligned}
$$

Hence, when a finite subgroup G' of G is an $S \cup S'$-design, and the measurement $M(\mu_{G'})$ attains the optimal precision in the Bayesian method, even though the measurement $M(\mu_{G'})$ takes the discrete value. When we focus on a projective representation of G with the factor system $\mathcal{E}$, S is a subset of $\hat{G}[\mathcal{E}]$. Since the following proof is still valid even when we replace $\hat{G}$ by $\hat{G}[\mathcal{E}]$, Theorem 4.11 holds for a projective representation.

Proof of Theorem 4.11 $\mathcal{D}_R(X)$ can be calculated as follows.

$$
\begin{aligned}
&\mathcal{D}_R(X) = \int_G R(e,\hat{g})|\mathcal{F}^{-1}[X](\hat{g}^{-1})|^2 \mu_G(d\hat{g}) \\
&= \int_G R(e,\hat{g})\Big|\sum_{\lambda \in S} \operatorname{Tr} \mathsf{f}_{\lambda}(\hat{g}^{-1}) c_{\lambda}\sqrt{d_{\lambda}}\varPhi_{\lambda}\Big|^2 \mu_G(d\hat{g}) \\
&= \int_G R(e,\hat{g}) \sum_{\lambda,\lambda' \in S} c_{\lambda} c_{\lambda'} \operatorname{Tr}(\mathsf{f}_{\lambda^*}(\hat{g}) \otimes \mathsf{f}_{\lambda'}(\hat{g}))(\sqrt{d_{\lambda}}\varPhi_{\lambda} \otimes \sqrt{d_{\lambda'}}\varPhi^{\dagger}_{\lambda'})\mu_G(d\hat{g}) \\
&= \sum_{\lambda,\lambda' \in S} c_{\lambda} c_{\lambda'} \operatorname{Tr}\Big[\int_G R(e,\hat{g})(\mathsf{f}_{\lambda^*}(\hat{g}) \otimes \mathsf{f}_{\lambda'}(\hat{g}))\mu_G(d\hat{g})\Big]\big(\sqrt{d_{\lambda}}(\varPhi_{\lambda} \otimes \sqrt{d_{\lambda'}}\varPhi^{\dagger}_{\lambda'})\big).
\end{aligned}
\tag{4.55}
$$

Using [44, (2.45) and (3.77)], we have

$$
\begin{aligned}
\varXi_{\lambda^*,\lambda'} &:= \int_G -R(e,\hat{g})\mathsf{f}_{\lambda^*}(\hat{g}) \otimes \mathsf{f}_{\lambda'}(\hat{g})\mu_G(d\hat{g}) \\
&= \int_G \sum_{\lambda'' \in \hat{G}} a_{\lambda''}(\chi_{\lambda''}(\hat{g}) + \chi_{\lambda''^*}(\hat{g}))\mathsf{f}_{\lambda^*}(\hat{g}) \otimes \mathsf{f}_{\lambda'}(\hat{g})\mu_G(d\hat{g}) \\
&= \sum_{\lambda'' \in \hat{G}} a_{\lambda''} \int_G (\chi_{\lambda''}(\hat{g}) + \chi_{\lambda''^*}(\hat{g}))\mathsf{f}_{\lambda^*}(\hat{g}) \otimes \mathsf{f}_{\lambda'}(\hat{g})\mu_G(d\hat{g}) \\
&= \sum_{\lambda'' \in \hat{G}} \frac{a_{\lambda''}}{d_{\lambda''}}(C^{\lambda''^*}_{\lambda^*,\lambda'} I_{\lambda''^*} + C^{\lambda''}_{\lambda^*,\lambda'} I_{\lambda''}) \geq 0.
\end{aligned}
$$

Since Ξ_{λ,λ'^*} is invariant for the representation, we have

$$\begin{aligned}
&\operatorname{Tr} \Xi_{\lambda,\lambda'^*} d_\lambda \Phi_\lambda^\dagger \Phi_\lambda \otimes I_{\lambda'} \\
=&\operatorname{Tr} \Xi_{\lambda,\lambda'^*} \int_G (\mathsf{f}_\lambda(g) \otimes \mathsf{f}_{\lambda'^*}(g)) d_\lambda \Phi_\lambda^\dagger \Phi_\lambda \otimes I_{\lambda'} (\mathsf{f}_\lambda(g) \otimes \mathsf{f}_{\lambda'^*}(g))^\dagger \mu_G(dg) \\
=&\operatorname{Tr} \Xi_{\lambda,\lambda'^*} I_\lambda \otimes I_{\lambda'^*} = \sum_{\lambda'' \in \hat{G}} a_{\lambda''} (C^{\lambda''^*}_{\lambda,\lambda'^*} + C^{\lambda''}_{\lambda,\lambda'^*}).
\end{aligned}$$

Similarly, we have

$$\operatorname{Tr} \Xi_{\lambda^*,\lambda'} I_\lambda \otimes d_{\lambda'} \Phi_{\lambda'}^\dagger \Phi_{\lambda'} = \sum_{\lambda'' \in \hat{G}} a_{\lambda''} (C^{\lambda''^*}_{\lambda,\lambda'^*} + C^{\lambda''}_{\lambda,\lambda'^*}).$$

Hence, applying Schwarz inequality with respect to the inner product $\langle A, B \rangle := \operatorname{Tr} \Xi_{\lambda^*,\lambda'} A^\dagger B$ to the case with $A := I_\lambda \otimes \sqrt{d_{\lambda'}} \Phi_{\lambda'}$ and $B := \sqrt{d_\lambda} \Phi_\lambda \otimes I_{\lambda'}$, we have

$$\begin{aligned}
&\operatorname{Tr} \Xi_{\lambda^*,\lambda'} \sqrt{d_\lambda} \Phi_\lambda \otimes \sqrt{d_{\lambda'}} \Phi_{\lambda'}^\dagger \\
\le& \sqrt{\operatorname{Tr} \Xi_{\lambda^*,\lambda'} d_\lambda \Phi_\lambda^\dagger \Phi_\lambda \otimes I_{\lambda'}} \sqrt{\operatorname{Tr} \Xi_{\lambda^*,\lambda'} I_\lambda \otimes d_{\lambda'} \Phi_{\lambda'}^\dagger \Phi_{\lambda'}} \\
=&, \sum_{\lambda'' \in \hat{G}} a_{\lambda''} (C^{\lambda''^*}_{\lambda,\lambda'^*} + C^{\lambda''}_{\lambda,\lambda'^*}).
\end{aligned} \tag{4.56}$$

Combining the 4.56 with the (4.51) and (4.55), we obtain (4.54). As the equality condition for the above Schwarz inequality, we obtain the equality condition $X = (c_\lambda \Psi_\lambda)_\lambda$ for the inequality (4.54). ■

Example 4.17 (SU(2) *[1, 21, 52]*) We consider the case when the group $G = \mathrm{SU}(2)$ naturally acts on $(\mathbb{C}^2)^{\otimes n}$ and the reference system is available. Hence, when n is an even number $2m$, the representation f has the irreducible decomposition $\oplus_{\lambda=0}^m d_\lambda \mathsf{f}_\lambda$, i.e., contains the irreducible representation f_λ with multiplicity d_λ. When n is an odd number $2m+1$, it has the irreducible decomposition $\oplus_{\lambda=0}^m d_{\lambda+\frac{1}{2}} \mathsf{f}_{\lambda+\frac{1}{2}}$.

When the gate fidelity $\mathcal{R}_{gate}(g) = 1 + \chi_1(g^\dagger \hat{g})$ is the gain function, we have

$$C^1_{\lambda,\lambda'} = C^{\lambda'}_{\lambda,1} = \begin{cases} 1 \text{ if } \lambda' = \lambda - 1, \lambda + 1 \\ 0 \text{ otherwise.} \end{cases} \tag{4.57}$$

So, we obtain

$$\mathcal{D}_{\mathcal{R}_{gate}}(\{c_\lambda\}) = \begin{cases} 1 + 2\sum_{\lambda=1}^m c_\lambda c_{\lambda-1} & \text{if } n = 2m \\ 1 + 2\sum_{\lambda=1}^m c_{\lambda+\frac{1}{2}} c_{\lambda-\frac{1}{2}} & \text{if } n = 2m+1. \end{cases} \tag{4.58}$$

Theorem 4.9 guarantees that this value is maximized when $\{c_\lambda\}$ satisfies

$$c_\lambda = \begin{cases} \sin \frac{\lambda+1\pi}{m+2} & \text{if } n = 2m \\ \sin \frac{\lambda+\frac{1}{2}\pi}{m+2} & \text{if } n = 2m+1, \end{cases}$$

and the maximum is $1 + 2\cos\frac{\pi}{m+2}$.

4.3.6 Energy Constraint*

When G is non-compact group, the representation space $\mathcal{U}_\lambda$ is often infinite-dimensional. Now, we consider the case when G is unimodular and $\mathcal{H} = \oplus_{\lambda\in S}\mathcal{U}_\lambda \otimes \mathcal{V}_\lambda$. Usually, it is physically impossible to generate any initial state. Hence, it is natural to impose an energy constraint to the initial state [62]. That is, we introduce a positive semi definite operator H_λ on each representation space $\mathcal{U}_\lambda$, and impose the following energy condition for the initial state

$$\operatorname{Tr}\Big(\bigoplus_{\lambda\in S} H_\lambda \otimes I\Big)\rho \le E \tag{4.59}$$

with a fixed constant. This condition is rewritten as

$$\sum_{\lambda\in S} \operatorname{Tr} H_\lambda X_\lambda X_\lambda^\dagger \le E \tag{4.60}$$

for the initial state $X \in L^2(\hat{G})$.

To state the main theorem, we prepare the function

$$C(E) := \min_{X\in L^2(\hat{G})} \left\{ \mathcal{D}_R(X) \middle| \begin{array}{l} \sum_{\lambda\in S} \operatorname{Tr} H_\lambda X_\lambda X_\lambda^\dagger \le E, \\ \|X\|_{L^2(\hat{G})} = 1 \end{array} \right\}. \tag{4.61}$$

Then, we have the following theorem.

Theorem 4.13 ([62]) *The relations*

$$\begin{aligned} C(E) &\ge \min_{\rho\in\mathcal{S}(\mathcal{H})} \min_{M\in\mathcal{M}_{\mathrm{cov}}(G)} \{\mathcal{D}_R(\rho, M) | \operatorname{Tr} H\rho \le E\} \\ &\ge \min_{\{p_i\}} \min_{\rho_i\in\mathcal{S}(\mathcal{H})} \min_{M_i\in\mathcal{M}_{\mathrm{cov}}(G)} \{\sum_i p_i \mathcal{D}_R(\rho_i, M_i) | \sum_i p_i \operatorname{Tr} H\rho_i \le E\} \\ &= \min_{\{p_i, E_i\}} \{\sum_i p_i C(E_i) | \sum_i p_i E_i = E\} \end{aligned} \tag{4.62}$$

hold. In particular, when the function $C(E)$ is convex, the relations

$$\min_{\rho\in\mathcal{S}(\mathcal{H})}\min_{M\in\mathcal{M}_{\mathrm{cov}}(G)}\{\mathcal{D}_R(\rho,M)|\operatorname{Tr}H\rho\le E\}$$
$$=\min_{\{p_i\}}\min_{\rho_i\in\mathcal{S}(\mathcal{H})}\min_{M_i\in\mathcal{M}_{\mathrm{cov}}(G)}\{\sum_i p_i\mathcal{D}_R(\rho_i,M_i)|\sum_i p_i\operatorname{Tr}H\rho_i\le E\}=C(E) \quad (4.63)$$

hold.

Further, we have the following lemma.

Lemma 4.2 ([62]) *When* $\dim\mathcal{V}_\lambda\ge\dim\mathcal{U}_\lambda$*, the function* $C(E)$ *is convex.*

Therefore, when the above condition holds, it is sufficient to minimize $\mathcal{D}_R(|X\rangle)$ among pure input states $|X\rangle$ under the condition $\langle X|H|X\rangle\le E$. We have the same discussion even in the projective representation case.

Example 4.18 ($\mathbb{R}$) As a typical example, we focus on the group $G=\mathbb{R}$. In this case, $\hat{G}$ is also $\mathbb{R}$. For a basis $|\lambda\rangle$, we have the unitary representation as $\mathsf{f}(x)|\lambda\rangle=e^{-i\lambda x}|\lambda\rangle$. For $\varphi\in L^2(\mathbb{R})$, we impose the energy constraint as $\langle\varphi|\mathsf{Q}^2|\varphi\rangle\le E$. So, the average error is given as

$$\int_{\mathbb{R}}x^2|\mathcal{F}^{-1}[\varphi](x)|\frac{dx}{\sqrt{2\pi}}=\langle\varphi|\mathsf{P}^2|\varphi\rangle. \quad (4.64)$$

So, using famous uncertainty relation, we have $C(E)=\frac{1}{E}$.

Example 4.19 (U(1) *[62]*) As another typical example, we focus on the group $G=\mathrm{U}(1)$. In this case, $\hat{G}$ is $\mathbb{Z}$. For a basis $|n\rangle$ with $n\in\mathbb{Z}$, we have the unitary representation as $\mathsf{f}(x)|n\rangle=e^{-inx}|n\rangle$. For $\varphi\in L^2(\mathbb{Z})$, we impose the energy constraint as $\sum_n n^2|\varphi(n)|^2\le E$, i.e., $\langle\mathcal{F}^{-1}[\varphi]|P^2|\mathcal{F}^{-1}[\varphi]\rangle\le E$. So, the average error is given as

$$\int_{-\pi}^{\pi}(1-\cos x)|\mathcal{F}^{-1}[\varphi](x)|\frac{dx}{\sqrt{2\pi}}=\langle\mathcal{F}^{-1}[\varphi]|(I-\cos Q)|\mathcal{F}^{-1}[\varphi]\rangle. \quad (4.65)$$

By using Mathieu function $a_0(\frac{2}{s})$ [44, Sect. 5.6.2], the relation [44, Lemma 5.6]

$$\min_{\varphi\in L^2_p((-\pi,\pi]):\|\varphi\|=1}\{\langle\varphi|I-\cos(\mathsf{Q})|\varphi\rangle|\langle\varphi|\mathsf{P}^2|\varphi\rangle\le E\}$$
$$=\max_{s>0}\frac{sa_0(\frac{2}{s})}{4}+1-sE \quad (4.66)$$

holds. So, $C(E)$ is given by the above value.

Example 4.20 (SU(2) *[62]*) Next, we focus on the group $G=\mathrm{SU}(2)$. In this case, $\hat{G}$ is given as the set of highest weights. We employ the error function $1-\frac{1}{2}\chi_{\frac{1}{2}}(g)$ and the Hamiltonian $H=\sum_{k=0}^{\infty}\frac{k}{2}(\frac{k}{2}+1)I_{\frac{k}{2}}$ for the energy constraint. So, for the input state $X\in L^2(\hat{G})$, we need to minimize the average error

$$\int_{\mathrm{SU}(2)}(1-\frac{1}{2}\chi_{\frac{1}{2}}(g))|\mathcal{F}^{-1}[X](-\zeta)|^2\mu_{\mathrm{SU}(2)}(dg) \tag{4.67}$$

under the energy constraint

$$\langle X|H|X\rangle \le E. \tag{4.68}$$

When $X=\oplus_{k=0}^{\infty}\beta_{\frac{k}{2}}|X_{\frac{k}{2}}\rangle\rangle$, the odd function $\varphi(\theta):=\sum_{k=0}^{\infty}\beta_{\frac{k}{2}}\sin\frac{k+1}{2}\theta$ satisfies [44, Lemma 5.8]

$$\int_{\mathrm{SU}(2)}(1-\frac{1}{2}\chi_{\frac{1}{2}}(g))|\mathcal{F}^{-1}[X](-\zeta)|^2\mu_{\mathrm{SU}(2)}(dg) \ge \langle\varphi|\cos\frac{\mathsf{Q}}{2}|\varphi\rangle \tag{4.69}$$

$$\langle X|H|X\rangle = \langle\varphi|\mathsf{P}^2|\varphi\rangle, \tag{4.70}$$

where the equality in (4.69) holds when $X=\oplus_{k=0}^{\infty}\beta_{\frac{k}{2}}|I_{\frac{k}{2}}\rangle\rangle$. So, $C(E)$ is given as the following minimum value

$$\max_{\varphi\in L^2_{p,\mathrm{odd}}((-2\pi,2\pi]):\|\varphi\|=1}\{\langle\varphi|I-\cos(\frac{\mathsf{Q}}{2})|\varphi\rangle|\langle\varphi|\mathsf{P}^2|\varphi\rangle\le E+\frac{1}{4}\}. \tag{4.71}$$

By using another Mathieu function $b_2(\frac{8}{s})$ [44, Sect. 5.6.2], this value is calculated as [44, Lemma 5.9]

$$\max_{s>0}\frac{sb_2(\frac{8}{s})}{16}+1-s(E+\frac{1}{4}). \tag{4.72}$$

Example 4.21 (*Heisenberg representation [62]*) In the Heisenberg representation W of $\mathbb{R}^2$ (see [44, Sect. 7.1.1]), since the representation space is infinite-dimensional, it is natural to impose an energy constraint for the initial state as in Theorem 4.13. We denote the inverse Fourier transform with the factor system of W by $\mathcal{F}_{\mathsf{W}}^{-1}$. When the representation space is $L^2(\mathbb{R})$ and the space describing the multiplicity is also $L^2(\mathbb{R})$, the initial state is given as a pure state $|X\rangle\rangle$ on the system $L^2(\mathbb{R})^{\otimes 2}$. When the error is measured as the square of the difference between the true parameter and the estimated value, the average error is given as

$$\int_{\mathbb{R}^2}(x^2+y^2)|\mathcal{F}_{\mathsf{W}}^{-1}[X](-\zeta)|^2dxdy, \tag{4.73}$$

where $\zeta=\frac{x+iy}{\sqrt{2}}$. Therefore, we discuss this problem under the energy constraint as

$$\langle\langle X|\mathsf{Q}^2+\mathsf{P}^2|X\rangle\rangle \le E. \tag{4.74}$$

Since the minimum eigenvalue of $\mathsf{Q}^2+\mathsf{P}^2$, we consider only the case when $E\ge 1$. Since (4.73) and (4.74) are given in (7.69) and (7.70) of [44], Theorem 7.1 of [44] guarantees that the minimum of (4.73) under the constraint (4.74) is $2(E-\sqrt{E^2-1})$. So, the combination of Theorem 4.13 and Lemma 4.2 yields

$$C(E) = 2(E - \sqrt{E^2 - 1}). \tag{4.75}$$

Exercise 4.12 Give the asymptotic expansion of $2(E - \sqrt{E^2 - 1})$ when E is large.

*4.3.7 Proofs of Theorem 4.13 and Lemma 4.2**

In this subsection, we show Theorem 4.13 and Lemma 4.2. Firstly, we show Theorem 4.13. We have the relations

$$\begin{aligned}
C(E) \ge & \min_{\rho\in\mathcal{S}(\mathcal{H})} \min_{M\in\mathcal{M}_{\mathrm{cov}}(G)} \{\mathcal{D}_R(\rho, M) | \operatorname{Tr} H\rho \le E\} \\
\ge & \min_{\{p_i\}} \min_{\rho_i\in\mathcal{S}(\mathcal{H})} \min_{M_i\in\mathcal{M}_{\mathrm{cov}}(G)} \{\sum_i p_i \mathcal{D}_R(\rho_i, M_i) | \sum_i p_i \operatorname{Tr} H\rho_i \le E\} \\
= & \min_{\{p_i\}} \min_{X_i\in\mathcal{H}} \min_{M_i\in\mathcal{M}_{\mathrm{cov}}(G)} \{\sum_i p_i \mathcal{D}_R(|X_i\rangle\langle X_i|, M_i) | \sum_i p_i \langle X_i|H|X_i\rangle \le E\},
\end{aligned} \tag{4.76}$$

where the first inequality can be shown by (2.55). Other relations in (4.76) are trivial. In fact, the relation (2.55) yields that

$$\begin{aligned}
& \min_{\{p_i\}} \min_{X_i\in\mathcal{H}} \min_{M_i\in\mathcal{M}_{\mathrm{cov}}(G)} \{\sum_i p_i \mathcal{D}_R(|X_i\rangle\langle X_i|, M_i) | \sum_i p_i \langle X_i|H|X_i\rangle \le E\} \\
\le & \min_{\{p_i, E_i\}} \{\sum_i p_i C(E_i) | \sum_i p_i E_i = E\}.
\end{aligned} \tag{4.77}$$

Hence, it is enough to show the inequality opposite to (4.77). For pure states $|X_i\rangle\langle X_i|$ and covariant POVM M_{T_i}, we choose $\eta_{k,\boldsymbol{\lambda},i}$ and $|x_{k,i}\rangle$. Then,

$$\begin{aligned}
& \sum_i p_i \langle X_i|H|X_i\rangle = \sum_i p_i \sum_{\boldsymbol{\lambda}\in S} \langle X_{\boldsymbol{\lambda},i}|H_{\boldsymbol{\lambda}}|X_{\boldsymbol{\lambda},i}\rangle \\
= & \sum_i p_i \sum_{\boldsymbol{\lambda}\in S} \operatorname{Tr} X^\dagger_{\boldsymbol{\lambda},i} H_{\boldsymbol{\lambda}} X_{\boldsymbol{\lambda},i} \sum_k \eta^\dagger_{k,\boldsymbol{\lambda},i}\eta_{k,\boldsymbol{\lambda},i} \\
= & \sum_k \sum_i p_i \sum_{\boldsymbol{\lambda}\in S} \operatorname{Tr} \eta_{k,\boldsymbol{\lambda},i} X^\dagger_{\boldsymbol{\lambda},i} H_{\boldsymbol{\lambda}} X_{\boldsymbol{\lambda},i} \eta^\dagger_{k,\boldsymbol{\lambda},i} = \sum_k \sum_i p_i \langle x_{k,i}|H|x_{k,i}\rangle.
\end{aligned} \tag{4.78}$$

Since

$$\sum_i p_i \mathcal{D}_R(|X_i\rangle\langle X_i|, M_i) = \sum_{i,k} \|x_{k,i}\|^2 p_i \mathcal{D}_R(\frac{1}{\|x_{k,i}\|^2}|x_{k,i}\rangle\langle x_{k,i}|, M_{|I\rangle\langle I|}),$$

we obtain the inequality opposite to (4.77).

Further, when $C(E)$ is convex $\min_{\{p_i,E_i\}}\{\sum_i p_i C(E_i)|\sum_i p_i E_i = E\} = C(E)$, which implies (4.63). So, we obtain Theorem 4.13.

Next, we proceed to our proof of Lemma 4.2. For this purpose, we prepare a technical lemma. For a given Hilbert space $\mathcal{H}$, we consider two self-adjoint operators Y and Z on a two-dimensional subspace $\mathcal{V}$ of $\mathcal{H}$. Then, we have the following lemma.

Lemma 4.3 *For any density matrix $\rho \in \mathcal{S}(\mathcal{V})$, there exists a normalized vector $\phi \in \mathcal{V}$ such that*

$$\operatorname{Tr}\rho Y = \langle\phi|Y|\phi\rangle, \quad \operatorname{Tr}\rho Z = \langle\phi|Z|\phi\rangle. \tag{4.79}$$

Proof We discuss only the case when X and Y are not constant. Otherwise, this statement is trivial. In this case, it is enough to show the case when the eigenvalues of Y and Z are 1 and -1. In this case, we can choose a suitable basis such that Y and Z can be written as sums of σ_1 and σ_3. When ρ is $\frac{1}{2}(I + a\sigma_1 + b\sigma_2 + c\sigma_3)$, we choose the pure state $|\phi\rangle\langle\phi| := \frac{1}{2}(I + a\sigma_1 + \hat{b}\sigma_2 + c\sigma_3)$ such that $a^2 + \hat{b}^2 + c^2 = 1$. Then, we have (4.79). ■

Now, we back to our proof of Lemma 4.2. It is enough to show that

$$\begin{aligned} & p\mathcal{D}_R(|X_1\rangle) + (1-p)\mathcal{D}_R(|X_2\rangle) \\ \geq & \min_{X \in L^2(\hat{G})} \left\{ \mathcal{D}_R(X) \,\middle|\, \begin{array}{l} \sum_{\lambda \in S} \operatorname{Tr} H_\lambda X_\lambda X_\lambda^\dagger \leq E, \\ \|X\|_{L^2(\hat{G})} = 1 \end{array} \right\} \end{aligned} \tag{4.80}$$

when $p \in [0,1]$ and $p\langle X_1|H|X_1\rangle + (1-p)\langle X_2|H|X_2\rangle = E$. The map $\rho \mapsto \mathcal{D}_R(\rho, M_{|I\rangle\langle I|})$ is affine. Thus, due to (4.34), there exists a self-adjoint map Y such that $\operatorname{Tr}\rho Y = \mathcal{D}_R(\rho, M_{|I\rangle\langle I|})$. Applying Lemma 4.3 to two operators Y and H and the density matrix $p|\phi_1\rangle\langle\phi_1| + (1-p)|\phi_2\rangle\langle\phi_2|$ with the two-dimensional subspace spanned by $|\phi_1\rangle$ and $|\phi_2\rangle$, we choose a vector ϕ as a superposition of $|\phi_1\rangle$ and $|\phi_2\rangle$ satisfying the condition (5.163). So, we have

$$\begin{aligned} & p\mathcal{D}_R(|\phi_1\rangle) + (1-p)\mathcal{D}_R(|\phi_2\rangle) = \operatorname{Tr}(p|\phi_1\rangle\langle\phi_1| + (1-p)|\phi_2\rangle\langle\phi_2|)Y \\ = & \langle\phi|Y|\phi\rangle = \mathcal{D}_R(|\phi\rangle) \end{aligned}$$

and

$$E = \operatorname{Tr}(p|\phi_1\rangle\langle\phi_1| + (1-p)|\phi_2\rangle\langle\phi_2|)H = \langle\phi|H|\phi\rangle.$$

Thanks to the condition of Lemma 4.2, $|\phi\rangle$ belongs to $\mathcal{K}_{\mathcal{H}}$. Hence, we obtain (4.80), i.e., Lemma 4.2.

4.4 Channel and Symmetry

Next, we consider quantum channels with symmetry. Assume that two quantum systems $\mathcal{H}_1$ and $\mathcal{H}_2$ are given and two projective unitary representations f_1 and f_2 of the same group G operate on these two systems. A quantum channel Λ from the system $\mathcal{H}_1$ to the other system $\mathcal{H}_2$ is called **covariant** with respect to f_1 and f_2 when

$$\Lambda(\mathsf{f}_1(g)\rho\mathsf{f}_1(g)^\dagger) = \mathsf{f}_2(g)\Lambda(\rho)\mathsf{f}_2(g)^\dagger, \quad \forall g \in G. \tag{4.81}$$

Then, we denote the set of covariant channels by $\mathcal{T}_{\mathrm{cov}}(\mathsf{f}_1, \mathsf{f}_2)$. This condition is equivalent to the following condition for the Choi-Jamiolkowski representation $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ of Λ:

$$\begin{aligned}
&(\mathsf{f}_2(g) \otimes \overline{\mathsf{f}_1}(g))^{-1}(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)(\mathsf{f}_2(g) \otimes \overline{\mathsf{f}_1}(g)) \\
=&(\mathsf{f}_2(g)^\dagger \otimes \mathsf{f}_1(g)^T)(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)(\mathsf{f}_2(g) \otimes \overline{\mathsf{f}_1}(g)) \\
=&(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|), \quad \forall g \in G.
\end{aligned} \tag{4.82}$$

When we denote the reference system of $\mathcal{H}_1$ by $\mathcal{H}_{1,R}$, the matrix $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ is invariant with respect to the representation $\mathsf{f}_2 \otimes \overline{\mathsf{f}_1}$ on the system $\mathcal{H}_2 \otimes \mathcal{H}_{1,R}$. When the representation space $\mathcal{H}_2 \otimes \mathcal{H}_{1,R}$ has the irreducible decomposition $\oplus_{\lambda \in S}\mathcal{U}_\lambda \otimes \mathbb{C}^{m_\lambda}$, Schurs lemma guarantees that $(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)$ is written as $\oplus_{\lambda \in S} I_\lambda \otimes T_\lambda$, where T_λ is a positive semi definite matrix on $\mathbb{C}^{m_\lambda}$.

Since the representation space of the representation $\mathsf{f}_2 \otimes \overline{\mathsf{f}_1}$ can be regarded as a the set of linear maps from $\mathcal{H}_1$ to $\mathcal{H}_2$, we have the following lemma.

Lemma 4.4 *Let f_1 and f_2 be irreducible projective unitary representations of the group G on $\mathcal{H}_1$ and $\mathcal{H}_2$. Assume that a TP-CP map Λ from the system $\mathcal{H}_1$ to the system $\mathcal{H}_2$ is covariant with respect to f_1 and f_2. A TP-CP map Λ is trace-preserving if and only if*

$$\Lambda(I_{\mathcal{H}_1}) = \frac{\dim \mathcal{H}_1}{\dim \mathcal{H}_2} I_{\mathcal{H}_2}. \tag{4.83}$$

When the group G is not compact, the above fact holds when the formal dimensions $\dim \mathcal{H}_1$ *and* $\dim \mathcal{H}_2$ *exist. Indeed, when the formal dimension* $\dim \mathcal{H}_1$ *does not exist and only the formal dimension* $\dim \mathcal{H}_2$ *exist, there is no covariant quantum channel.*

Proof Since $\frac{1}{\dim \mathcal{H}_1} I_{\mathcal{H}_1} = \int_G \mathsf{f}_1(g)\rho\mathsf{f}_1(g)^\dagger \mu_G(dg)$, the covariance of Λ implies that

$$\Lambda\Big(\int_G \mathsf{f}_1(g)\rho\mathsf{f}_1(g)^\dagger \mu_G(dg)\Big) = \int_G \mathsf{f}_2(g)\Lambda(\rho)\mathsf{f}_2(g)^\dagger \mu_G(dg) = \frac{\mathrm{Tr}\,\Lambda(\rho)}{\dim \mathcal{H}_2} I_{\mathcal{H}_2}.$$

Since Λ is trace-preserving, we have $\mathrm{Tr}\,\Lambda(\rho) = 1$. So, we obtain (4.83). Conversely, when the relation (4.83) holds, we have $\mathrm{Tr}\,\Lambda(\rho) = 1$. So, Λ is trace-preserving.

When the formal dimension $\dim \mathcal{H}_1$ does not exist and only the formal dimension $\dim \mathcal{H}_2$ exists, we derive contradiction by assuming the existence of the covariant channel Λ. The dual map Λ^* and the state ρ' on $\mathcal{H}_2$ satisfy that

$$\begin{aligned}&\operatorname{Tr} \rho' \frac{1}{\dim \mathcal{H}_2} I_{\mathcal{H}_2} = \operatorname{Tr} \rho' \Lambda \Big(\int_G \mathsf{f}_1(g) \rho \mathsf{f}_1(g)^\dagger \mu_G(dg) \Big) \\ = &\operatorname{Tr} \Lambda^*(\rho') \int_G \mathsf{f}_1(g) \rho \mathsf{f}_1(g)^\dagger \mu_G(dg).\end{aligned}$$

Since the RHS diverges and the LHS does not diverge, we have a contradiction. So, there is no covariant channel. ■

Next, we consider a projective unitary representation f_1 of the group G on the quantum system $\mathcal{H}_1$. Given a probability distribution ν on G, we define the quantum channel $\Lambda[\mathsf{f}_1, \nu]$ with the input and output system $\mathcal{H}_1$ as $\Lambda[\mathsf{f}_1, \nu](\rho) := \int_G \mathsf{f}_1(g) \rho \mathsf{f}_1(g)^\dagger \nu(dg)$. Then, the Choi-Jamiolkowski representation of the channel $\Lambda[\mathsf{f}_1, \nu]$ is

$$\sum_{g \in G} \nu(g) |\mathsf{f}_1(g)\rangle\!\rangle \langle\!\langle \mathsf{f}_1(g)|. \tag{4.84}$$

Conversely, when a channel has the Choi-Jamiolkowski representation given in (4.84), it is given as $\Lambda[\mathsf{f}_1, \nu]$.

Further, we consider another projective unitary representation f_2 of the same group G on the other system $\mathcal{H}_2$ and a channel from $\mathcal{H}_1$ to $\mathcal{H}_2$, We define the **averaged channel** $\overline{\Lambda}_\nu$ of Λ with respect to the probability distribution ν on G as

$$\overline{\Lambda}_\nu(\rho) := \int_G \mathsf{f}_2(g)^{-1} \Lambda(\mathsf{f}_1(g) \rho \mathsf{f}_1(g)^{-1}) \mathsf{f}_2(g) \nu(dg). \tag{4.85}$$

Especially, when ν is the invariant measure μ_G of the compact Lie group G, $\overline{\Lambda}_\mu$ is called the **twirling** of Λ with respect to the projective unitary representations f_1 and f_2 of G. When f_2 is equivalent to f_1, $\overline{\Lambda}_\mu$ is simply called the twirling of Λ with respect to the projective unitary representation f_1 of G.

Theorem 4.14 *Assume that Λ is a quantum channel from the system $\mathcal{H}_1$ to the system $\mathcal{H}_2$ and that f_1 and f_2 are projective unitary representations of a compact Lie group G on the respective systems. Then, the twirling $\overline{\Lambda}_\mu$ of Λ with respect to the projective unitary representations f_1 and f_2 of G is a covariant channel with respect to f_1 and f_2.*

Corollary 4.1 *Assume that a probability measure ν on G is an $\mathsf{f}_2 \otimes \overline{\mathsf{f}_1}$-design under the same condition as Theorem 4.14. Then, the averaged channel $\overline{\Lambda}_\nu$ of Λ with respect to ν is covariant with respect to f_1 and f_2.*

Example 4.22 (*Pauli channel*) Given the discrete Heisenberg representation $\mathsf{W}^r_{\mathbb{X}}$ of the group $\mathbb{X}^{2r}$ on the input and output system with $\mathbb{X} = \mathbb{Z}_d$ or $\mathbb{X} = \mathbb{F}_q$, the covariant channel is called a **Pauli channel**. Any irreducible component of the representation $\mathsf{W}^r_{\mathbb{X}} \otimes \overline{\mathsf{W}^r_{\mathbb{X}}}$ is one-dimensional. When this representation space is a space of matrices, any one-dimensional irreducible component is spanned by the base $\mathsf{W}^r_{\mathbb{X}}(\vec{s})$ given by an element $\vec{s} \in \mathbb{X}^{2r}$. In this case, the channel has the Choi-Jamiolkowski representation $\sum_{\vec{s}\in\mathbb{X}^{2r}} \mathrm{P}(\vec{s})|\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rangle\rangle\langle\langle\mathsf{W}^r_{\mathbb{X}}(\vec{s})|$ with a probability distribution P on $\mathbb{X}^{2r}$. Hence, the Pauli channel $\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}]$ corresponding to the distribution P is expressed as

$$\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}](\rho) = \sum_{\vec{s}\in\mathbb{X}^{2r}} \mathrm{P}(\vec{s})\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rho\mathsf{W}^r_{\mathbb{X}}(\vec{s})^\dagger. \tag{4.86}$$

Conversely, given a general channel Λ, we define the distribution

$$P[\Lambda](\vec{s}) := \frac{1}{|\mathbb{X}^r|^2}\langle\langle\mathsf{W}^r_{\mathbb{X}}(\vec{s})|\Lambda \otimes id(|I\rangle\rangle\langle\langle I|)|\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rangle\rangle. \tag{4.87}$$

The twirling of the channel Λ with respect to the discrete Heisenberg representation $\mathsf{W}^r_{\mathbb{X}}$ equals $\Lambda[\mathsf{W}^r_{\mathbb{X}}, P[\Lambda]]$ [7, 34]. This fact can be shown by comparing both Choi-Jamiolkowski representations.

Example 4.23 (*depolarizing channel*) A channel on the system $\mathbb{C}^r$ is called a **depolarizing channel** when it is covariant with respect to the fundamental representation $\mathsf{f}_{[1,0,\ldots,0]}$. When the representation space of the representation $\mathsf{f}_{[1,0,\ldots,0]} \otimes \overline{\mathsf{f}_{[1,0,\ldots,0]}}$ is expressed as the space of matrices on $\mathbb{C}^r$, it is irreducibly decomposed to the one-dimensional component composed of constant matrices and its orthogonal complement. Hence, the channel has the Choi-Jamiolkowski representation $t|I\rangle\rangle\langle\langle I| + \frac{(1-t)}{r}I$ with a real number $t \geq 0$. Then, the depolarizing channel $\Lambda_{\mathsf{f}_{[1,0,\ldots,0]},t}$ is written as $\Lambda_{\mathsf{f}_{[1,0,\ldots,0]},t}(\rho) = t\rho + (1-t)\rho_{\mathrm{mix}}$. Especially, when Λ is a channel on $\mathbb{C}^r$ and a probability measure ν on $\mathrm{SU}(r)$ is a $(2,0)$-design, the averaged channel $\overline{\Lambda}_\nu$ with respect to ν is a depolarizing channel.

Example 4.24 (*conjugate depolarizing channel*) Given the fundamental representation $\mathsf{f}_{[1,0,\ldots,0]}$ of the group $\mathrm{SU}(r)$ on the input system $\mathbb{C}^r$ and its comppex conjugate representation $\overline{\mathsf{f}_{[1,0,\ldots,0]}} = \mathsf{f}_{[0,\ldots,0,1]}$ on the output system $\mathbb{C}^r$, a covariant channel is called a **conjugate depolarizing channel**. The representation space of $\overline{\mathsf{f}_{[1,0,\ldots,0]}} \otimes \overline{\mathsf{f}_{[1,0,\ldots,0]}} = \mathsf{f}^{\otimes 2}_{[0,\ldots,0,1]}$ is irreducibly decomposed to the symmetric tensor product space $\mathcal{H}_s$ and the alternative tensor product space $\mathcal{H}_a$. Hence, any conjugate depolarizing channel has the Choi-Jamiolkowski representation $\frac{1+(r-1)t}{r}P_s + \frac{1-(r+1)t}{r}P_a$ with a real number $t \geq 0$ and is denoted by $\Lambda_{[1,0,\ldots,0]\to[0,\ldots,0,1],t}$, where P_s and P_a are the projections to $\mathcal{H}_s$ and $\mathcal{H}_a$, respectively. Here, we notice that the transpose operation $\rho \mapsto \rho^T$ has the Choi-Jamiolkowski representation $P_s - P_a$. Hence, the conjugate depolarizing channel $\Lambda_{[1,0,\ldots,0]\to[0,\ldots,0,1],t}$ is written as

$$\Lambda_{[1,0,\ldots,0]\to[0,\ldots,0,1],t}(\rho) = t\rho^T + (1-t)\rho_{\mathrm{mix}}. \tag{4.88}$$

Thus, due to Lemma 2.1, the conjugate depolarizing channel $\Lambda_{[1,0,\ldots,0]\to[0,\ldots,0,1],t}$ is a TP-CP map if and only if $-\frac{1}{r-1} \leq t \leq \frac{1}{r+1}$.

In the following, for irreducible representations $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3 \in \hat{G}$ of a group G, we assume that the representation space $\mathcal{U}_{\boldsymbol{\lambda}_1} \otimes \mathcal{U}_{\boldsymbol{\lambda}_3}$ contains the irreducible representation space $\mathcal{U}_{\boldsymbol{\lambda}_2}$ with multiplicity 1. Then, by letting $P_{\boldsymbol{\lambda}_2}$ be the projection to $\mathcal{U}_{\boldsymbol{\lambda}_2}$ and $d_{\boldsymbol{\lambda}_1}, d_{\boldsymbol{\lambda}_2}$ be the dimensions of the representation spaces $\mathcal{U}_{\boldsymbol{\lambda}_1}, \mathcal{U}_{\boldsymbol{\lambda}_2}$, respectively, the relation (3.54) of [44] implies that

$$\operatorname{Tr} P_{\boldsymbol{\lambda}_2}(\rho \otimes I_{\mathcal{U}_{\boldsymbol{\lambda}_3}}) P_{\boldsymbol{\lambda}_2} \frac{d_{\boldsymbol{\lambda}_2}}{d_{\boldsymbol{\lambda}_1}}, \tag{4.89}$$

where $d_{\boldsymbol{\lambda}_1}$ and $d_{\boldsymbol{\lambda}_2}$ are the the formal dimensions when G is not compact. Thus, we can define the following channel $\Lambda^{+}_{\boldsymbol{\lambda}_1 \to \boldsymbol{\lambda}_2}$, which is called the $\boldsymbol{\lambda}_1 \to \boldsymbol{\lambda}_2$ **extending covariant channel**.

$$\Lambda^{+}_{\boldsymbol{\lambda}_1 \to \boldsymbol{\lambda}_2}(\rho) := \frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}} P_{\boldsymbol{\lambda}_2}(\rho \otimes I_{\mathcal{U}_{\boldsymbol{\lambda}_3}}) P_{\boldsymbol{\lambda}_2}. \tag{4.90}$$

Since the representation space $\mathcal{U}_{\boldsymbol{\lambda}_1^*} \otimes \mathcal{U}_{\boldsymbol{\lambda}_2}$ contains the irreducible representation space $\mathcal{U}_{\boldsymbol{\lambda}_3}$ with multiplicity 1 (see [44, Lemma 2.12]), the Choi-Jamiolkowski representation of the channel $\Lambda^{+}_{\boldsymbol{\lambda}_1 \to \boldsymbol{\lambda}_2}$ is a constant times of the projection to $\mathcal{U}_{\boldsymbol{\lambda}_3}$ (see [44, **(1)** of Lemma 2.13]).

Further, the channel $\Lambda^{-}_{\boldsymbol{\lambda}_2 \to \boldsymbol{\lambda}_1}(\rho) := \operatorname{Tr}_{\mathcal{U}_{\boldsymbol{\lambda}_3}} \rho$ is called the $\boldsymbol{\lambda}_2 \to \boldsymbol{\lambda}_3$ **contracting covariant channel**, where $\operatorname{Tr}_{\mathcal{U}_{\boldsymbol{\lambda}_3}}$ expresses the partial trace on the tensor product space $\mathcal{U}_{\boldsymbol{\lambda}_1} \otimes \mathcal{U}_{\boldsymbol{\lambda}_3}$. Then, due to Item **(1)** of Lemma 2.13 in [44], the Choi-Jamiolkowski representation of the channel $\Lambda^{-}_{\boldsymbol{\lambda}_2 \to \boldsymbol{\lambda}_1}$ is the constant times of the projection to $\mathcal{U}_{\boldsymbol{\lambda}_3^*}$.

4.5 Approximate State Cloning

We have discussed the optimization of measurement in Sect. 4.1 when a given state family $\{\rho_\theta\}_{\theta \in \Theta}$ has a covariant structure with respect to a given unitary (projective) representation f of a group G on a quantum system $\mathcal{H}$. In this section, we optimize a state operation Λ from a system $\mathcal{H}$ to another system $\mathcal{H}_1$ given a unitary (projective) representation f of a group G on a quantum system $\mathcal{H}$ and a function family $\{R_\theta\}_{\theta \in \Theta}$ $\mathcal{S}(\mathcal{H}_1)$. Although a state operation is a quantum channel mathematically, we employ the term "state operation" to express it in this section because we optimize the operation manipulating a state.

Now, we define the performance $R_\theta(\Lambda(\rho_\theta))$ of a state operation Λ with the true parameter θ and consider the problem to to find the state operation Λ to maximize this value. This problem is a generalized problem to contain the approximate state cloning explained later. Since the parameter is unknown, we optimize the worst value $R(\Lambda) := \max_\Lambda R_\theta(\Lambda(\rho_\theta))$ with respect to θ or the averaged value $R_\nu(\Lambda) := \int_\Theta R_\theta(\Lambda(\rho_\theta))\nu(d\theta)$ under the distribution ν for θ.

Since it is not easy to discuss this problem in a general setting, we impose the **covariance condition** with respect to the representation f_1 to the function family

$\{R_\theta\}_{\theta\in\Theta}$ on $\mathcal{S}(\mathcal{H}_1)$. Then, we can show the following theorem by using the averaged channel defined in (4.85) in the same way to the proof of Theorem 4.1.

Theorem 4.15 *Assume that Θ is compact, the function family $\{R_\theta\}_{\theta\in\Theta}$ satisfies the covariance condition with respect to the representation f_1, and each function R_θ is a convex function. Then, the relations*

$$\min_{\Lambda\in\mathcal{T}(\mathcal{H},\mathcal{H}_1)} R(\Lambda) = \min_{\Lambda\in\mathcal{T}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda) = \min_{\Lambda\in\mathcal{T}_{\mathrm{cov}}(\mathcal{H},\mathcal{H}_1)} R(\Lambda)$$
$$= \min_{\Lambda\in\mathcal{T}_{\mathrm{cov}}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda)$$

hold, where $\mathcal{T}(\mathcal{H},\mathcal{H}_1)$ ($\mathcal{T}_{\mathrm{cov}}(\mathcal{H},\mathcal{H}_1)$) is a set of (covariant) TP-CP maps from the system $\mathcal{H}$ to the system $\mathcal{H}_1$. That is, the solution of mini-max method equals the solution of Bayes method with respect to the invariant measure μ, and an optimum operation is given as a covariant operation. Also, any solution of the mini-max method is a covariant operation.

Proof Any operation Λ satisfies $R_\theta(\mathsf{f}_1(g)^\dagger\Lambda(\mathsf{f}(g)\rho_\theta\mathsf{f}(g)^\dagger)\mathsf{f}_1(g)) = R_{g\theta}\Lambda(\rho_{g\theta})$. Hence, the convexity of R_θ implies that

$$\int_G R_{g\theta}\Lambda(\rho_{g\theta})\mu(dg) = \int_G R_\theta(\mathsf{f}_1(g)^\dagger\Lambda(\mathsf{f}(g)\rho_\theta\mathsf{f}(g)^\dagger)\mathsf{f}_1(g))\mu(dg)$$
$$\geq R_\theta\Big(\int_G \mathsf{f}_1(g)^\dagger\Lambda(\mathsf{f}(g)\rho_\theta\mathsf{f}(g)^\dagger)\mathsf{f}_1(g)\mu(dg)\Big) = R_\theta\Big(\int_G \overline{\Lambda}_\mu(\rho_\theta)\Big).$$

Since the operation $\overline{\Lambda}_\mu$ is a covariant TP-CP map, we have

$$\min_{\Lambda\in\mathcal{T}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda) \geq \min_{\Lambda\in\mathcal{T}_{\mathrm{cov}}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda) = \min_{\Lambda\in\mathcal{T}_{\mathrm{cov}}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda).$$

Since the opposite inequalities

$$\min_{\Lambda\in\mathcal{T}_{\mathrm{cov}}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda) \geq \min_{\Lambda\in\mathcal{T}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda) \geq \min_{\Lambda\in\mathcal{T}(\mathcal{H},\mathcal{H}_1)} R_{\mu_\Theta}(\Lambda)$$

also hold, we obtain the desired argument. ∎

When Θ is not compact, the full integral with respect to invariant measure μ is not finite. So, the above statement does not hold with respect to the Bayes method. However, the following theorem holds for mini-max method in the same way as Theorem 4.1.

Theorem 4.16 *Assume that Θ is locally compact. Any covariant state operation Λ satisfies $R(\Lambda) = R_\theta(\Lambda(\rho_\theta))$. Further, the relation $\min_{\Lambda\in\mathcal{T}(\mathcal{H},\mathcal{H}_1)} R(\Lambda) = \min_{\Lambda\in\mathcal{T}_{\mathrm{cov}}(\mathcal{H},\mathcal{H}_1)} R(\Lambda)$ holds.*

In the following, as a typical example of a state operation, we address the **approximate state cloning**, which is formulated as follows. When n copies of an unknown state ρ are given, we find the optimum operation to generate a state that approximates the state $\rho^{\otimes(n+m)}$, i.e., $n+m$ copies of the state ρ. When the candidates of the unknown state are limited and are orthogonal to each other, the perfect cloning is possible. However, it is impossible otherwise. This problem can be resolved as follows. Given covariant state family $\{\rho'_\theta\}_{\theta\in\Theta}$ with respect to a representation f' of a group G on the system $\mathcal{H}'$, we focus on the tate family $\{{\rho'_\theta}^{\otimes n}\}_{\theta\in\Theta}$ on the system $\mathcal{H}$, and measure the performance of the state operation Λ by $R_\theta(\Lambda({\rho'_\theta}^{\otimes n})) = \mathrm{Tr}\,\Lambda({\rho'_\theta}^{\otimes n}){\rho'_\theta}^{\otimes(n+m)}$. Then, we need to maximize $R_\theta(\Lambda({\rho'_\theta}^{\otimes n}))$. Since an approximate state cloning is a special example of optimization of state operation under the group covariant framework, we can apply Theorems 4.15 and 4.16 to this problem. That is, we can address the approximate state cloning in the framework of the optimization of the state operation with the group covariance in the above way. The following lemma holds for approximate state transformation, which is a generalization of the approximate state cloning with a covariant state family composed of pure states.

Lemma 4.5 *Given two covariant state families, the input state family $\{\rho_{1,\theta}\}_{\theta\in\Theta}$ and the target state family $\{\rho_{2,\theta}\}_{\theta\in\Theta}$ composed of pure states with respect to irreducible unitary representations $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \hat{G}$ of the group G, we consider the approximate state transformation with the error function $R_\theta(\Lambda(\rho_{1,\theta})) := \mathrm{Tr}\,\Lambda(\rho_{1,\theta})\rho_{2,\theta}$ as Fig. 4.5. Then, the inequality*

$$\min_{\Lambda\in\mathcal{T}_{\mathrm{cov}}(\mathsf{f}_{\boldsymbol{\lambda}_1},\mathsf{f}_{\boldsymbol{\lambda}_2})} R(\Lambda) \le \frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}} \tag{4.91}$$

holds.

Proof Assume that Λ is a covariant channel. Lemma 4.4 yields

$$\mathrm{Tr}\,\Lambda(\rho_{1,\theta})\rho_{2,\theta} \le \mathrm{Tr}\,\Lambda(I)\rho_{2,\theta} = \mathrm{Tr}\,\frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}} I\rho_{2,\theta} = \frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}}, \tag{4.92}$$

which implies the desired argument. ■

A covariant state family $\{\rho_{2,\theta}\}_{\theta\in\Theta}$ composed of pure state for an irreducible unitary representation $\boldsymbol{\lambda}_2 \in \hat{G}$ of a group G is called an **extended covariant state family** of a covariant state family $\{\rho_{1,\theta}\}_{\theta\in\Theta}$ composed of pure state for an irreducible unitary

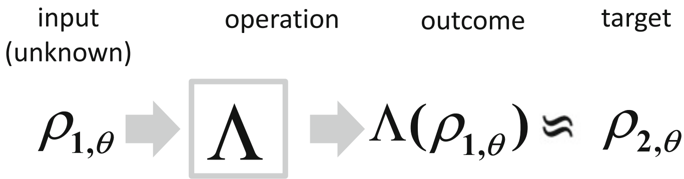


Fig. 4.5 Approximate state transformation

representation $\boldsymbol{\lambda}_1 \in \hat{G}$ of the group G when there exists a covariant state family $\{\rho_{3,\theta}\}_{\theta\in\Theta}$ for an irreducible unitary representation $\boldsymbol{\lambda}_3 \in \hat{G}$ of the group G such that the tensor product space $\mathcal{U}_{\boldsymbol{\lambda}_1} \otimes \mathcal{U}_{\boldsymbol{\lambda}_3}$ contains the irreducible representation $\mathcal{U}_{\boldsymbol{\lambda}_2}$ only with multiplicity 1 and satisfies $\rho_{1,\theta} \otimes \rho_{3,\theta} = \rho_{2,\theta}$.

Lemma 4.6 *When a covariant state family $\{\rho_{2,\theta}\}_{\theta\in\Theta}$ satisfies the condition for an extended covariant state family of $\{\rho_{1,\theta}\}_{\theta\in\Theta}$ as well as the assumption of Lemma 4.5, an extended covariant channel $\Lambda^{+}_{\boldsymbol{\lambda}_1\to\boldsymbol{\lambda}_2}$ for $\boldsymbol{\lambda}_1 \to \boldsymbol{\lambda}_2$ satisfies*

$$R(\Lambda^{+}_{\boldsymbol{\lambda}_1\to\boldsymbol{\lambda}_2}) = \frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}}. \tag{4.93}$$

That is, the equality holds in (4.91). Hence, the state operation $\Lambda_{\boldsymbol{\lambda}_1+\boldsymbol{\lambda}_3=\boldsymbol{\lambda}_2}$ realizes the optimal performance.

Proof We have

$$\begin{aligned}
&\mathrm{Tr}\, \Lambda^{+}_{\boldsymbol{\lambda}_1\to\boldsymbol{\lambda}_2}(\rho_{1,\theta})\rho_{2,\theta} = \mathrm{Tr}\, \frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}} Q_{\boldsymbol{\lambda}_2}(\rho_{1,\theta} \otimes I_{\mathcal{U}_{\boldsymbol{\lambda}_3}}) Q_{\boldsymbol{\lambda}_2}\rho_{2,\theta} \\
\geq\ &\mathrm{Tr}\, \frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}} Q_{\boldsymbol{\lambda}_2}(\rho_{1,\theta} \otimes \rho_{3,\theta}) Q_{\boldsymbol{\lambda}_2}\rho_{2,\theta} = \frac{d_{\boldsymbol{\lambda}_1}}{d_{\boldsymbol{\lambda}_2}}.
\end{aligned}$$

Since the opposite inequality is guaranteed by (4.91), we obtain the desired argument. ■

For example, we consider the when G is a Lie group, the representation $\mathsf{f}_{\boldsymbol{\lambda}_1}$ has the highest weight $\boldsymbol{\lambda}_1$, and the representation $\mathsf{f}_{\boldsymbol{\lambda}_3}$ has the highest weight $\boldsymbol{\lambda}_3$ that is a scalar times of $\boldsymbol{\lambda}_1$. When $\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_3$ and the covariant state families $\{\rho_{2,\theta}\}_{\theta\in\Theta}$ and $\{\rho_{1,\theta}\}_{\theta\in\Theta}$ are coherent state families, $\{\rho_{2,\theta}\}_{\theta\in\Theta}$ is an extended covariant state family of $\{\rho_{1,\theta}\}_{\theta\in\Theta}$. Especially, when the weights $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_3$ are given as integer n and m times of a weight $\boldsymbol{\lambda}$, the problem is approximately cloning from n copies of a coherent state to $n + m$ copies of the same coherent state.

When the problem is given with integer times, we often employ the quantity

$$\mathrm{Tr}\, \rho_{1,\theta} \frac{1}{n+m} \sum_{j=1}^{n+m} (\mathrm{Tr}_{\check{j}}\, \Lambda(\rho_{1,\theta}^{\otimes n})) = \mathrm{Tr}\, \mathcal{E}(\rho_{1,\theta}^{\otimes n}) \frac{1}{n+m} \sum_{j=1}^{n+m} \rho_{1,\theta}^{(j)}$$

as a criterion of the performance of the state operation Λ. This criterion evaluates the average of the closeness for each system. When we employ the state operation $\Lambda^{+}_{n\boldsymbol{\lambda}\to(n+m)\boldsymbol{\lambda}}$ whose optimality is guaranteed by Lemma 4.6, The performance with this criterion is evaluated as follows. In the following derivation, we notice that $P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}}$ is commutative with $\frac{1}{n+m}\sum_{j=1}^{n+m} \rho_{1,\theta}^{(j)}$.

$$
\begin{aligned}
&\mathrm{Tr}\, \Lambda^{+}_{n\boldsymbol{\lambda}\to(n+m)\boldsymbol{\lambda}}(\rho_{1,\theta}^{\otimes n}) \frac{1}{n+m}\sum_{j=1}^{n+m}\rho_{1,\theta}^{(j)} \\
=&\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+m)\boldsymbol{\lambda}}} \mathrm{Tr}\, P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}} \rho_{1,\theta}^{\otimes n} \otimes I_{\boldsymbol{\lambda}}^{\otimes m} P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}} \frac{1}{n+m}\sum_{j=1}^{n+m}\rho_{1,\theta}^{(j)} \\
=&\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+m)\boldsymbol{\lambda}}} \mathrm{Tr}\, P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}} \rho_{1,\theta}^{\otimes n} \otimes I_{\boldsymbol{\lambda}}^{\otimes m} \frac{1}{n+m}\sum_{j=1}^{n+m}\rho_{1,\theta}^{(j)} \\
=&\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+m)\boldsymbol{\lambda}}} \frac{1}{n+m}\sum_{j=1}^{n+m} \mathrm{Tr}\, P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}} \rho_{1,\theta}^{(j)}\rho_{1,\theta}^{\otimes n} \otimes I_{\boldsymbol{\lambda}}^{\otimes m} \\
=&\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+m)\boldsymbol{\lambda}}} \Big(\frac{n}{n+m}\frac{d_{(n+m)\boldsymbol{\lambda}}}{d_{n\boldsymbol{\lambda}}} + \frac{m}{m+n}\frac{d_{(n+m)\boldsymbol{\lambda}}}{d_{(n+1)\boldsymbol{\lambda}}}\Big) \qquad (4.94)\\
=&\frac{n}{n+m} + \frac{m}{n+m}\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+1)\boldsymbol{\lambda}}} = 1 - \frac{m}{n+m}\frac{d_{(n+1)\boldsymbol{\lambda}}-d_{n\boldsymbol{\lambda}}}{d_{(n+1)\boldsymbol{\lambda}}}, \qquad (4.95)
\end{aligned}
$$

where (4.94) is derived by the following fact shown by (4.89);

$$
\begin{aligned}
&\mathrm{Tr}\, P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}} \rho_{1,\theta}^{(j)}\rho_{1,\theta}^{\otimes n} \otimes I_{\boldsymbol{\lambda}}^{\otimes m} \\
=&\begin{cases} \mathrm{Tr}\, P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}} \rho_{1,\theta}^{\otimes n} \otimes I_{\boldsymbol{\lambda}}^{\otimes m} = \frac{d_{(n+m)\boldsymbol{\lambda}}}{d_{n\boldsymbol{\lambda}}} & \text{when } 1 \le j \le n \\ \mathrm{Tr}\, P_{\mathcal{U}_{(n+m)\boldsymbol{\lambda}}} \rho_{1,\theta}^{\otimes n+1} \otimes I_{\boldsymbol{\lambda}}^{\otimes (m-1)} = \frac{d_{(n+m)\boldsymbol{\lambda}}}{d_{(n+1)\boldsymbol{\lambda}}} & \text{when } n+1 \le j \le n+m. \end{cases}
\end{aligned}
$$

That is, the precision (4.95) of this cloning under this criterion is given as a function of the precision $\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+1)\boldsymbol{\lambda}}}$ of cloning with n copies and the ratio $\frac{n}{n+m}$. Especially, when the number $n+m$ of required copies goes to infinity, the precision of cloning goes to the estimation precision $\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+1)\boldsymbol{\lambda}}}$ with n copies.

We often discuss a projective unitary representation. However, when the presentation is finite-dimensional and the group is a Lie group, the projective unitary representation can be usually given as a skew Hermitian representation of the corresponding Lie algebra. In the following examples, we address a skew Hermitian representation of a Lie algebra, which corresponds a unitary representation the universal covering group.

Example 4.25 ($\mathfrak{su}(2)$ *[29, 128]* $\mathfrak{su}(1,1)$) In the case of $\mathfrak{g} = \mathfrak{su}(2), \mathfrak{su}(1,1)$, the highest weight λ is restricted to a negative real number or a positive half integer. Then, the precision of approximation is $\frac{d_{n\lambda}}{d_{(n+m)\lambda}} = \frac{2n\lambda+1}{2(n+m)\lambda+1} = 1 - \frac{m}{n}\frac{1}{1+\frac{1}{2(n+m)\lambda}}$. Here, we assume that $2n\lambda < -1$ for the case of $\mathfrak{g} = \mathfrak{su}(1,1)$. Especially, in the case of one-mode squeezed state, since $\lambda_1 = -\frac{1}{4}$, we have $\frac{d_{n\lambda}}{d_{(n+m)\lambda}} = \frac{n-2}{n+m-2}$ for $n > 2, m > 1$. In the case of two-mode squeezed state, since $\lambda = -\frac{1}{2}$, we have $\frac{d_{n\lambda}}{d_{(n+m)\lambda}} = \frac{n-1}{n+m-1}$ for $m, n > 1$.

Example 4.26 (*[18]*) We consider the case when $\mathfrak{g} = \mathfrak{h}_3 \rtimes \mathfrak{u}(1)$ and $\boldsymbol{\lambda} = (\lambda_Z, \lambda_F)$. The precision is calculated to $\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+m)\boldsymbol{\lambda}}} = \frac{n\lambda_Z}{(n+m)\lambda_Z} = \frac{n}{n+m} = 1 - \frac{m}{n+m}$.

Example 4.27 ($\mathfrak{su}(r)$ *[128]*) Next, we consider the case when $\mathfrak{g} = \mathfrak{su}(r)$ and $\boldsymbol{\lambda} = [1, 0, \ldots, 0]$. Then, the precision is calculated to $\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+m)\boldsymbol{\lambda}}} = \frac{(n+r-1)!(n+m)!}{(n+m+r-1)!n!}$.

Example 4.28 We consider the case when $\mathfrak{g} = \mathfrak{h}(2r, \mathbb{R}) \rtimes \mathfrak{u}(r)$ (i.e., r-mode Bosonic case) and $\boldsymbol{\lambda} = (\lambda_Z, \lambda_F \boldsymbol{v})$. Then, the precision is calculated to $\frac{d_{n\boldsymbol{\lambda}}}{d_{(n+m)\boldsymbol{\lambda}}} = \frac{n^r \lambda_Z^r}{(n+m)^r \lambda_Z^r} = (\frac{n}{n+m})^r$.

In the case of the one-mode or two-mode squeezed state or the r-mode Bosonic system, the above formulas hold even with non-integer real numbers n and m. That is, we can formally consider the approximate state cloning even when the numbers of copies are non-integer real numbers.

Chapter 5
Quantum Error Correction and Its Application

Abstract Quantum error correction is a technology to protect quantum state from noise and decoherence. This technology is essentially based on group symmetry. In particular, a stabilizer code and a CSS (Calderbank-Shor-Smolin) code (typical quantum error corrections) are essentially based on the discrete Heisenberg representation. Since the quantum error correction is complicated, we firstly explain error correction in the classical case in Sect. 5.1.1. Section 5.2 shows the general formulation of quantum error correction. In Sect. 5.3, combining this knowledge and Discrete Heisenberg representation, which is summarized in Sect. 3.3.1, we explain stabilizer code, which is a typical example of quantum error correction. We also discuss CSS (Calderbank-Shor-Smolin) code, which is a special case of a stabilizer code. For simplicity, our analysis in Sect. 5.3 assumes a Pauli channel. In Sect. 5.4, we introduce Clifford Code, which is a generalization of stabilizer code as a general framework of quantum error correction with group symmetry. In Sect. 5.5, we apply quantum error correction to entanglement distillation. Also, we discuss our error correction when the noisy channel is a general channel beyond a Pauli channel. In Sect. 5.6, we apply our discussion of stabilizer code to quantum secure communication. In Sect. 5.7, we discuss quantum cryptography (quantum key distribution) based on the framework of this chapter.

5.1 Error Correction in Classical System Based on Algebraic Structure

5.1.1 Formulation of Problem

We consider a communication via classical channel transmitting an element of a group G. Especially, we assume that the output $g' \in G$ is probabilistic due to the presence of the noise in the communication channel and the output distribution depends on the input signal $g \in G$. Hence, a classical noisy communication channel is written as a probability transition matrix on G. In the following, we do not address a general classical channel, and assume that the multiplication noise $n \in G$ is subject to the probability distribution P^G on G. That is, when the input is $g \in G$ and the

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8_5

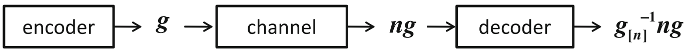


Fig. 5.1 Error correction in classical communication channel

multiplication noise is $n \in G$, the channel output is $ng \in G$. In this setting, we restrict the message to be sent to a subgroup C of G so that we can protect our message from the noise. In this case, when the received signal g' does not belong to C, we can infer that the signal is changed due to the noise. Given an element x of the quotient space G/C, which is often called a coset, we choose the most probable noise in the coset x as a representative g_x. When the received signal $g' \in G$ belongs to an element x of the quotient space G/C, we convert g' to $g_x^{-1}g'$. Then, the recovered message equals the original message with the probability $\sum_{x \in G/C} p(g_x)$. The value $1 - \sum_{x \in G/C} p(g_x)$ is called the decoding error probability. In the following, the subgroup C is called a classical **code space**, and the representative $\{g_x\}_{x \in G/C}$ is called classical **decoding**. Then, when $M = |C|$, the map from a message space $\{1, \ldots, M\}$ to $C \subset G$ is called a classical **encoding**. Especially, the representative $J_{\mathrm{ML}}(x) := \operatorname{argmax}_{g \in x} \mathrm{P}^G(g)$ for each coset $x \in G/C$ is called the **maximum likelihood decoding** of the classical code space C under the probability distribution P^G. When all elements $g \in C$ are transmitted with equal probability, the maximum likelihood decoding realizes the minimum decoding error probability, which is denoted by $\delta[C]$ (Fig. 5.1).

As a practical group, we can consider the vector space $\mathbb{F}_q^r$ over the finite field $\mathbb{F}_q$. In this case, we often construct a code space as a vector subspace over $\mathbb{F}_q$. When the vector subspace C has the dimension k, it is identified with its k basis. Arranging k column vectors corresponding to the basis, we make $k \times r$ matrix A over $\mathbb{F}_q$, which is called the generating matrix of the code space C. Hence, when we identify the vector space $\mathbb{F}_q^k$ with the message space, the generating matrix A can be identified with the encoder.

Exercise 5.1 (*three-digit code*) Consider the case when $G = \mathbb{F}_q^3$ and the classical code space C is the subspace $C_{(3)}$ generated by the generating matrix $(1, 1, 1)^T$. The vector space $\mathbb{F}_q^3$ can be divided to the union of elements of quotient space as $\cup_{b,c \in \mathbb{F}_q} \{(a + b, a + c, a)^T\}_{a \in \mathbb{F}_q}$. Show that we can choose representatives $\{J(x)\}_{x \in \mathbb{F}_q}$ $(J(x) \in x)$ such that $\{(a, 0, 0), (0, a, 0), (0, 0, a)\}_{a \neq 0 \in \mathbb{F}_q}$ is included in the set $\{J(x)\}_{x \in \mathbb{F}_q^3/C_{(3)}}$.

Exercise 5.2 We consider a probability distribution P over $\mathbb{F}_q$ satisfying the following condition. (1) $P(a) > P(b)$ for $a < b$. (2) $P(0)P(q-1) \geq P(1)P(2)$. (3) $P(0)P(q-1)P(q-2) \geq P(1)P(2)P(3)$. Here, all elements of $\mathbb{F}_q$ are labeled by integers $\{0, 1, \ldots, q-1\}$, and 0 and 1 are the zero and the unit elements. We assume that the noise $n \in \mathbb{F}_q^3$ obeys the 3-trial independent and identical distribution of the probability distribution P over $\mathbb{F}_q$. Give the map J_{ML} corresponding to the maximum likelihood decoder in the three-digit code.

Exercise 5.3 We consider a probability distribution P over $\mathbb{F}_q$ satisfying the following condition when $0 < p < 1/2$. (1) $P(0) = 1 - p$. (2) $P(a) = \frac{p}{q-1}$ for $a \neq 0$.

We assume that the noise $n \in \mathbb{F}_q^3$ obeys the 3-trial independent and identical distribution of the probability distribution P over $\mathbb{F}_q$. Calculate the correctly decoding probability under the three-digit code with the decoder given in Exercise 5.2.

Exercise 5.4 (*Hamming code*) We consider the code space $C_{G,1} \subset \mathbb{F}_q^7$ generated by the following generating matrix:

$$G_1 := \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \\ 1\,0\,1\,1 \\ 1\,1\,0\,1 \\ 0\,1\,1\,1 \end{pmatrix}. \tag{5.1}$$

Show that all of elements of $\mathbb{F}_q^7/C_{G,1}$ are given as $[(0, 0, 0, 0, a, b, c)^T]_{a,b,c\in\mathbb{F}_q}$.

Exercise 5.5 Given the above Hamming code $C_{G,1}$ with 7 digits, show that the original message can be recovered whatever error occurs among one-digit errors if the decoder J satisfies that the number of 0 in the sequence $J(x)$ is the minimum among those of elements in x.

Exercise 5.6 We assume that the noise $n \in \mathbb{F}_q^7$ obeys the 7-trial independent and identical distribution of the probability distribution given in Exercise 5.3. Show that the correctly decoding probability is larger than $(1-p)^6(1+6p)$ under the above Hamming code $C_{G,1}$ with 7 digits when the decoder satisfies the condition given in Exercise 5.5.

5.1.2 *Upper Bound of Decoding Error Probability*

In the original definition of Shannon [114], the code space is allowed to be an arbitrary subset of G. However, when we employ an algebraic structure of G, such an arbitrary subset as the code space requires a larger amount of calculation complexity of decoding. So, we do not address such a general code space. That is, this chapter addresses the code space given as a subgroup of G. In the following, we upper bound the probability $\delta[C]$.

Lemma 5.1 (Gallager [28]) *Given a classical code space C, the decoding error probability $\delta[C]$ with the maximum likelihood decoder is upper bounded as*

$$\delta[C] \le \sum_{g\in G} \mathrm{P}^G(g)^{1-st} \left(\sum_{g'\in C\setminus\{e\}} \mathrm{P}^G(gg')^t \right)^s \tag{5.2}$$

with arbitrary real numbers $s, t > 0$.

Proof When we employ the code space C, the decoding error probability with the maximum likelihood decoder is given by $\delta[C] = \sum_{g \in G} \mathrm{P}^G(g)\varDelta_{\mathrm{ML}}(g)$, where $\varDelta_{\mathrm{ML}}(g)$ is the indicator function defined by

$$\varDelta_{\mathrm{ML}}(g) := \begin{cases} 0 \text{ if } \mathrm{P}^G(g) > \mathrm{P}^G(gg'), & \forall g' \in C \setminus \{e\} \\ 1 \text{ otherwise.} \end{cases}$$

Then, using

$$\varDelta_{g'}(g) := \begin{cases} 0 \text{ if } \mathrm{P}^G(g) > \mathrm{P}^G(gg'), \\ 1 \text{ otherwise,} \end{cases}$$

we evaluate the indicator function $\varDelta_{\mathrm{ML}}(g)$ as

$$\varDelta_{\mathrm{ML}}(g) \le \sum_{g' \in C \setminus \{e\}} \varDelta_{g'}(g). \tag{5.3}$$

Here, given $s > 0$, we take the s-th powers of both sides of (5.3) so that we obtain $\varDelta_{\mathrm{ML}}(g) \le \left(\sum_{g' \in C \setminus \{e\}} \varDelta_{g'}(g)\right)^s$. Further, we obtain

$$\varDelta_{g'}(g) \le \frac{\mathrm{P}^G(gg')}{\mathrm{P}^G(g)}. \tag{5.4}$$

Similarly, taking the t-th powers of both sides of (5.4) for $t > 0$, we have $\varDelta_{g'}(g) \le \frac{\mathrm{P}^G(gg')^t}{\mathrm{P}^G(g)^t}$. Summarizing them, we obtain

$$\varDelta_{\mathrm{ML}}(g) \le \left(\sum_{g' \in C \setminus \{e\}} \frac{\mathrm{P}^G(gg')^t}{\mathrm{P}^G(g)^t} \right)^s. \tag{5.5}$$

Then, the decoding error probability can be evaluated as

$$\begin{aligned} \delta[C] \le & \sum_{g \in G} \mathrm{P}^G(g) \left(\sum_{g' \in C \setminus \{e\}} \frac{\mathrm{P}^G(gg')^t}{\mathrm{P}^G(g)^t} \right)^s \\ = & \sum_{g \in G} \mathrm{P}^G(g)^{1-st} \left(\sum_{g' \in C \setminus \{e\}} \mathrm{P}^G(gg')^t \right)^s. \end{aligned} \tag{5.6}$$

■

5.1.3 Evaluation Under Ensemble

Given a probability distribution P^G of the noise n, the above lemma characterize the performance of the code space C by evaluating the decoding error probability. However, it is not easy to evaluate the decoding error probability in general. Hence, we often consider an ensemble of code spaces, which is simplified to a code ensemble, and evaluate the average of the decoding error probability with the maximum likelihood decoder for the ensemble. Evaluation of such an average seems more complicated, however, it enables us to employ a probabilistic method to bring a tighter evaluation.

In the following, we select a subgroup of G as the code space subject to a certain random variable X. That is, we denote the code space decided by the random variable X by C_X. The code ensemble C_X is called ϵ-universal2 [17, 120, 122] when

$$\mathrm{P}_X\{C_X | g \in C_X\} \le \epsilon, \quad \forall g \neq e \in G, \tag{5.7}$$

where the random variable X is subject to the probability distribution P_X. The following theorem holds [28, 122].

Theorem 5.1 *When the code ensemble C_X is ϵ-universal2, the inequality*

$$\mathrm{E}_X \delta[C_X] \le \min_{0\le s\le 1} \epsilon^s e^{\phi(s:\mathrm{P}^G)} \tag{5.8}$$

holds, where $\phi(s : \mathrm{P}^G)$ is defined as $\phi(s : \mathrm{P}^G) := \log(\sum_{g\in G} (\mathrm{P}^G(g))^{\frac{1}{1+s}})^{1+s}$.

Proof Given $s \in [0, 1]$, we substitute $t = \frac{1}{1+s}$ into (5.2). So, we have

$$\delta[C_X] \le \sum_{g\in G} (\mathrm{P}^G(g))^{\frac{1}{1+s}} \left(\sum_{g' \in C_X \setminus \{e\}} (\mathrm{P}^G(gg'))^{\frac{1}{1+s}} \right)^s .$$

By taking the average with respect to the random variable X, the convexity of the function $x \mapsto x^s$ implies that

$$\begin{aligned} \mathrm{E}_X \delta[C_X] \le & \mathrm{E}_X \sum_{g\in G} \mathrm{P}^G(g)^{\frac{1}{1+s}} \left(\sum_{g'\in C_X\setminus\{e\}} \mathrm{P}^G(gg')^{\frac{1}{1+s}} \right)^s \\ \le & \sum_{g\in G} \mathrm{P}^G(g)^{\frac{1}{1+s}} \left(\mathrm{E}_X \sum_{g'\in C_X\setminus\{e\}} \mathrm{P}^G(gg')^{\frac{1}{1+s}} \right)^s \\ \le & \sum_{g\in G} \mathrm{P}^G(g)^{\frac{1}{1+s}} \left(\epsilon \sum_{g'\in G} \mathrm{P}^G(gg')^{\frac{1}{1+s}} \right)^s . \end{aligned} \tag{5.9}$$

Since $\left(\epsilon \sum_{g' \in G} \mathrm{P}^G(gg')^{\frac{1}{1+s}}\right)^s$ does not depend on g, the above value equals $\left(\epsilon \sum_{g' \in G} \mathrm{P}^G(g')^{\frac{1}{1+s}}\right)^s = \epsilon^s \left(\sum_{g' \in G} \mathrm{P}^G(g')^{\frac{1}{1+s}}\right)^s$. Hence, the RHS of (5.9) equals

$$\begin{aligned}\sum_{g \in G} \mathrm{P}^G(g)^{\frac{1}{1+s}} \epsilon^s \left(\sum_{g' \in G} \mathrm{P}^G(g')^{\frac{1}{1+s}}\right)^s &= \epsilon^s \left(\sum_{g \in G} \mathrm{P}^G(g)^{\frac{1}{1+s}}\right)^{1+s} \\ &= \epsilon^s e^{\phi(s:\mathrm{P}^G)}. \end{aligned} \tag{5.10}$$

■

Given two groups G_1 and G_2, their direct product group $G_1 \times G_2$ often does not have suitable ϵ-universal2 code ensemble C_X. When the groups G_1 and G_2 have their respective ϵ-universal2 code ensembles C_{1,X_1} and C_{2,X_2}, from these code ensembles, we can construct the code ensemble $C_{1,X_1} \times C_{2,X_2}$ of the direct product group $G_1 \times G_2$, which satisfies the following theorem.

Theorem 5.2 *Let C_{1,X_1} (C_{2,X_2}) be an ϵ_1(ϵ_2)-universal2 code ensemble of a group G_1 (G_2), respectively. Then, the code ensemble $C_{1,X_1} \times C_{2,X_2}$ of the direct product group $G_1 \times G_2$ satisfies*

$$\begin{aligned}&\mathrm{E}_{X_1,X_2} \delta[C_{1,X_1} \times C_{2,X_2}] \\ &\le \min_{0 \le s \le 1} \epsilon_1^s e^{\phi(s:\mathrm{P}^{G_1})} + \min_{0 \le s \le 1} \epsilon_2^s e^{\phi(s:\mathrm{P}^{G_2|G_1})},\end{aligned} \tag{5.11}$$

where $\phi(s : \mathrm{P}^{G_2|G_1}) := \log(\sum_{g_1 \in G_1} \mathrm{P}^{G_1}(g_1)(\sum_{g \in G}(\mathrm{P}^G(g))^{\frac{1}{1+s}})^{1+s})$.

Proof Given a code space $C_{1,X_1} \times C_{2,X_2}$, we define the following decoder with two steps. Since the maximum likelihood decoder realizes the minimum decoding error probability, it is enough to show the required evaluation for the above specific decoder. In the first step, we apply the maximum likelihood decoder for the code $C_{1,X_1} \subset G_1$ based on the marginal distribution P^{G_1} so that we obtain the first estimate $\hat{n}_1$ for the noise in G_1. In the second step, we apply the maximum likelihood decoder $C_{2,X_2} \subset G_2$ based on the conditional distribution $\mathrm{P}^{G_2|\hat{n}_1}$ so that we obtain the second estimate $\hat{n}_{2:\hat{n}_1}$ for the noise in G_2. Finally, we multiply the inverse of the estimated noise $(\hat{n}_1, \hat{n}_{2:\hat{n}_1})$ to the received signal (g_1, g_2) so that we infer that the original message is $(\hat{n}_1^{-1} g_1, \hat{n}_{2:\hat{n}_1}^{-1} g_2)$.

Hence, the decoding error probability with this decoder is $\mathrm{P}\{\hat{n}_1 \neq n_1\} + \mathrm{P}\{\hat{n}_1 = n_1, \hat{n}_{2:\hat{n}_1} \neq n_2\}$. Theorem 5.1 guarantees that the average of $\mathrm{P}\{\hat{n}_1 \neq n_1\}$ with respect to the code ensemble is upper bounded by $\epsilon_1^s e^{\phi(s:\mathrm{P}^{G_1})}$. The remaining term is upper bounded by the decoding error probability for the maximum likelihood decoder based on the conditional distribution $\mathrm{P}^{G_2|n_1}$ on G_2 as

$$\mathrm{P}\{\hat{n}_1 = n_1, \hat{n}_{2:\hat{n}_1} \neq n_2\} = \mathrm{P}\{\hat{n}_1 = n_1, \hat{n}_{2:n_1} \neq n_2\} \le \mathrm{P}\{\hat{n}_{2:n_1} \neq n_2\}. \tag{5.12}$$

So, Theorem 5.1 guarantees that

$$\begin{aligned}&\mathrm{E}_{X_2}\mathrm{P}\{\hat{n}_{2:n_1} \neq n_2\} = \mathrm{E}_{X_2}\mathrm{E}_{n_1:\mathrm{P}^{G_1}}\mathrm{P}\{\hat{n}_{2:n_1} \neq n_2\}\\ =&\mathrm{E}_{n_1:\mathrm{P}^{G_1}}\mathrm{E}_{X_2}\mathrm{P}\{\hat{n}_{2:n_1} \neq n_2\} \le \mathrm{E}_{n_1:\mathrm{P}^{G_1}} \min_{0\le s\le 1} \epsilon_2^s e^{\phi(s:\mathrm{P}^{G_2|n_1})}.\end{aligned}$$

Combining both evaluations, we obtain (5.11). ■

When the group G has subgroup N and the order of N is coprime to that of another subgroup, the group G often does not have ϵ-universal2 code ensemble. To resolve this problem, we introduce an (ϵ_1, ϵ_2)-double universal2 code ensemble C_X of G as follows. A code ensemble C_X of G is called (ϵ_1, ϵ_2)-**double universal2** for the subgroup $N \subset G$ when

$$\begin{aligned}\mathrm{P}_X\{C_X | g \in C_X\} &\le \epsilon_1, \quad \forall g \in N \setminus \{e\}\\ \mathrm{P}_X\{C_X | g \in C_X\} &\le \epsilon_2, \quad \forall g \in G \setminus N.\end{aligned}$$

Then, the following theorem holds.

Theorem 5.3 *A (ϵ_1, ϵ_2)-double universal2 code ensemble C_X of G for the subgroup $N \subset G$ satisfies*

$$\mathrm{E}_X\delta[C_X] \le \min_{0\le s\le 1} \epsilon_2^s e^{\phi(s:\mathrm{P}^G)} + \min_{0\le s\le 1} \epsilon_1^s e^{\phi(s:\mathrm{P}^{N|G/N})}, \tag{5.13}$$

where $\mathrm{P}^G([g])$, $\mathrm{P}^{[g]}$, and $\phi(s : \mathrm{P}^{N|G/N})$ are defined as

$$\begin{aligned}\mathrm{P}^G([g]) &:= \sum_{g'\in \mathrm{P}^G([g])} \mathrm{P}^G(g')\\ \mathrm{P}^{[g]}(g') &:= \frac{\mathrm{P}^G(g')}{\mathrm{P}^G([g])}, \quad \forall g' \in [g]\\ \phi(s : \mathrm{P}^{N|G/N}) &:= \log(\sum_{[g]\in G/N} \mathrm{P}^G([g])(\sum_{g'\in N}(\mathrm{P}^{[g]}(gg'))^{\frac{1}{1+s}})^{1+s}).\end{aligned} \tag{5.14}$$

Proof Using the relation (5.9) in Theorem 5.1, we have

$$\begin{aligned}&\mathrm{E}_X\delta[C_X]\\ \le&\sum_{g\in G}\mathrm{P}^G(g)^{\frac{1}{1+s}}\left(\epsilon_1\sum_{g'\in N}\mathrm{P}^G(gg')^{\frac{1}{1+s}} + \epsilon_2\sum_{g'\in G}\mathrm{P}^G(gg')^{\frac{1}{1+s}}\right)^s\end{aligned}$$

$$\leq \sum_{g\in G} \mathrm{P}^G(g)^{\frac{1}{1+s}} \left(\epsilon_1 \sum_{g'\in N} \mathrm{P}^G(gg')^{\frac{1}{1+s}} \right)^s$$
$$+ \sum_{g\in G} \mathrm{P}^G(g)^{\frac{1}{1+s}} \left(\epsilon_2 \sum_{g'\in G} \mathrm{P}^G(gg')^{\frac{1}{1+s}} \right)^s ,$$

where we employ the inequality $(x+y)^s \leq x^s + y^s$ for $x, y \geq 0$. The second term can be evaluated in the same way as Theorem 5.1. The first term is calculated as

$$\sum_{g\in G} \mathrm{P}^G(g)^{\frac{1}{1+s}} \left(\epsilon_1 \sum_{g'\in N} \mathrm{P}^G(gg')^{\frac{1}{1+s}} \right)^s$$
$$= \sum_{g\in G/N} \mathrm{P}^G([g]) \sum_{n\in N} \mathrm{P}^{[g]}(gn)^{\frac{1}{1+s}} \left(\epsilon_1 \sum_{g'\in N} \mathrm{P}^{[g]}(gng')^{\frac{1}{1+s}} \right)^s$$
$$= \sum_{g\in G/N} \mathrm{P}^G([g]) \epsilon_1^s \left(\sum_{n\in N} \mathrm{P}^{[g]}(gn)^{\frac{1}{1+s}} \right)^{1+s} .$$

So, we obtain (5.13). ■

5.1.4 Construction of Code Ensemble

However, it is not easy to construct an ϵ-universal2 code ensemble for a general group G with large order. When G is a vector space $\mathbb{F}_q^r$, we have the following lemma [60].

Lemma 5.2 *Let $X = (X_1, \ldots, X_{r-1})$ be $r-1$ independent and uniform random variables taking values in $\mathbb{F}_q$. Then, we define the $\mathbb{F}_q$ valued $r-k \times k$ Toeplitz matrix A_X as $(A_X)_{l,j} := X_{r-k-l+j}$. Then, the code ensemble C_X generated by the generating matrix $\begin{pmatrix} I \\ A_X \end{pmatrix}$ is a q^{k-r}-universal2.*

Proof When all of the first k entries of $\boldsymbol{s} \neq 0 \in \mathbb{F}_q^r$ are 0, the probability that $\boldsymbol{s}$ belongs to the code space C_X is zero. On the other hand, we assume that there is an non-zero entry among the first k entries of $\boldsymbol{s}$, and let the t-th entry be the latest non-zero entry. The element $\boldsymbol{s}$ belongs to the code space C_X if and only if

$$s_{k+l} = X_{r-k-l+t} s_t + \sum_{j=1}^{t-1} X_{r-k-l+j} s_j , \quad l = 1, \ldots, r-k. \tag{5.15}$$

Now, we focus on the condition (5.15) with $l = r - k$. Since X_t is independent of $X_{t-1}, \ldots, X_1$ and the random variable $X_t s_t$ takes all elements of $\mathbb{F}_q$ with equal probability $1/q$, The condition (5.15) with $l = r - k$ holds with probability $1/q$. Next, we focus on the condition (5.15) with $l = u$ under the conditions (5.15) with $l = u + 1, \ldots, r - k$. Since the random variable $X_{r-k-u+t}$ is independent of $X_{r-k-u+t-1}, \ldots, X_1$ and the random variable $X_{r-k-u+t} s_t$ takes all elements of $\mathbb{F}_q$ with equal probability $1/q$, the condition (5.15) with $l = u$ holds with probability $1/q$. Hence, the conditions (5.15) with all l hold probability $1/q^{r-k}$. ■

5.1.5 Asymptotic Theory

In the previous subsections, we have discussed the information transmission via transmitting an element of G whose noise is described by a probability distribution P^G on G. In the following, we discuss the information transmission via transmitting an element of the r-times direct product group $G^r := \overbrace{G \times \cdots G}^{r}$ and the noise subject to the r-fold independent and identical distribution $(\mathrm{P}^G)^r$ of the distribution P^G. Then, we have

$$\phi(s : (\mathrm{P}^G)^r) = r\phi(s : \mathrm{P}^G). \tag{5.16}$$

In particular, the theory for infinitely large r is called asymptotic theory. In this case, we consider the code space C_r that is a subspace of the group G^r for each r, and discuss the sequence $\{C_r\}$ of code spaces. Then, we focus on the limit $\lim_{r\to\infty} \delta[C_r]$ of the decoding error probability of the maximum likelihood decoder and the asymptotic transmission rate $\lim_{r\to\infty} \frac{1}{r} \log |C_r|$. Entropy plays an important role in the asymptotic theory. So, similar to $H(\boldsymbol{p})$, the entropy $H(\mathrm{P}^G)$ of the distribution P^G on G is given as

$$H(\mathrm{P}^G) := -\sum_{g\in G} \mathrm{P}^G(g) \log \mathrm{P}^G(g). \tag{5.17}$$

For a distribution P on the direct product group $G_1 \times G_2$, the conditional entropy $H(\mathrm{P}^{G_2|G_1})$ is defined as

$$H(\mathrm{P}^{G_2|G_1}) := -\sum_{g_1\in G_1} \mathrm{P}^{G_1}(g_1) \sum_{g_2\in G_2} \mathrm{P}^{G_2|g_1}(g_2) \log \mathrm{P}^{G_2|g_1}(g_2). \tag{5.18}$$

The conditional entropy satisfies

$$H(\mathrm{P}^{G_1\times G_2}) = H(\mathrm{P}^{G_2|G_1}) + H(\mathrm{P}^{G_1}). \tag{5.19}$$

When the group G has a subgroup G_1, the conditional entropy $H(\mathrm{P}^{G_1|G/G_1})$ with respect to the subgroup G_1 of the distribution P^G on G is defined as

$$H(\mathrm{P}^{G_1|G/G_1}) := -\sum_{[g]\in G/G_1} \mathrm{P}^G([g]) \sum_{g_1\in G_1} \mathrm{P}^{[g]}(gg_1)\log \mathrm{P}^{[g]}(gg_1). \tag{5.20}$$

Then, we have

$$H(\mathrm{P}^G) = \lim_{s\to 0} \frac{\phi(s:\mathrm{P}^G)}{s} \tag{5.21}$$

$$H(\mathrm{P}^{G_1|G/G_1}) = \lim_{s\to 0} \frac{\phi(s:\mathrm{P}^{G_1|G/G_1})}{s}. \tag{5.22}$$

Although we choose a code space as a subgroup of G. in this chapter, Shannon's channel coding theorem is known in the following way as the optimal performance when employing an arbitrary subset of G as the code space.

Theorem 5.4 ([114]) *When an arbitrary subset of G^r is used as the code space and the limit of the decoding error probability are imposed to zero, the maximum transmission rate is* $\log|G| - H(\mathrm{P}^G)$.

Next, we choose an real number satisfying

$$R > \frac{H(\mathrm{P}^{\mathbb{F}_q})}{\log q} \tag{5.23}$$

and an integer $k := \lfloor (1-R)r \rfloor$. Then, we apply Theorem 5.1 to the q^{k-r}-universal2 code ensemble given in Lemma 5.2. Since the LHS of (5.8) is the average of the decoding error probability of the code ensemble, there exists a code space whose decoding error probability is less than the RHS of (5.8). Now, we address the sequence of code spaces given here. The transmission rate is $\log q - R\log q$, (5.16) guarantees that the RHS of (5.8) is $e^{-r\max_{0\le s\le 1}(sR\log q - \phi(s:\mathrm{P}^{\mathbb{F}_q}))}$. Due to (5.21), the condition (5.23) yields the inequality $R\log q > \frac{\phi(s:\mathrm{P}^{\mathbb{F}_q})}{s}$ with sufficiently small $s > 0$. Hence, $\max_{0\le s\le 1}(sR\log q - \phi(s:\mathrm{P}^{\mathbb{F}_q})) > 0$. So, the RHS of (5.8) approaches to zero exponentially for r. Thus, there exists a sequence of codes such that the decoding error probability goes to zero and the transmission rate is close to $\log q - H(\mathrm{P}^{\mathbb{F}_q})$. Therefore, the sequence of codes given here attains the asymptotically optimal transmission rate given in Theorem 5.4.

5.1.6 Error Correction with Secrecy

When a part of information is leaked to the eavesdropper, we need additional information processing to disable the eavesdropper to recover the original message as

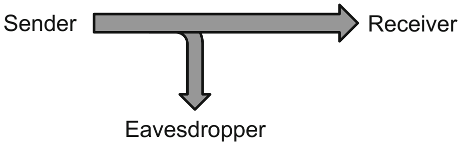


Fig. 5.2 Information transmission with eavesdropper

Fig. 5.2. To resolve this problem, we encode our message to an element of the quotient group C/N for a given normal subgroup N of C instead of an element of C [131]. When the quotient group C/N is used instead of C, the encoding method is called the **privacy amplification**. The correctly decoding probability is calculated to $\mathrm{P}^G(\mathcal{J}N) := \sum_{x\in G/C, n\in N} \mathrm{P}^G(g_x n)$ under the distribution P^G of the noise and the classical decoder $\mathcal{J} := \{g_x\}_{x\in G/C}$. Especially, the message $[g] \in C/N$ is incorrectly decoded to $[g'g] \in C/N$ with probability $\mathrm{P}^G\{\mathcal{J}, N\}^{C/N}([g']) := \mathrm{P}^G(\mathcal{J}g'N)$. The pair of subgroups $N \subset C$ is called a **code pair**. In particular, the choice of the representative $\mathrm{argmax}_{g\in x} \sum_{n\in N} \mathrm{P}^G(gn)$ for each coset $x \in G/C$ is called the **maximum likelihood decoder** of the code pair $N \subset C$ under the probability distribution P^G. This choice realizes the minimum decoding error probability, which is denoted by $\delta_{\mathrm{P}^G}[C/N]$. If there is no possibility for confusion without P^G, we simplify it to $\delta[C/N]$.

We often discuss the case when the normal subgroup N is fixed priorly and the code space C containing N is needed to be chosen dependently of a certain random variable X. In such a case, we need to employ $(1, \epsilon)$-double universal2 code ensemble C_X for a given normal subgroup $N \subset G$. Such a code ensemble can be constructed from a code ensemble in G/N as follows.

Lemma 5.3 *Let N be a normal subgroup of G and π_N be the homomorphism from G to G/N. Given an ϵ-universal2 code ensemble C_X of G/N, $\pi_N^{-1}(C_X)$ is a $(1, \epsilon)$-double universal2 code ensemble for the normal subgroup $N \subset G$.*

This lemma can be shown from the definition of $(1, \epsilon)$-double universal2 code ensemble, and enables us to construct a code ensemble needed in this subsection from a code ensemble given in Sect. 5.1.4 as follows. For such a construction, we prepare the following lemma as a generalization of Lemma 5.1. It can be shown by replacing $G \setminus \{e\}$ by $G \setminus N$.

Lemma 5.4 *Given a code pair $N \subset C$, the decoding error probability $\delta[C/N]$ with maximum likelihood decoder is evaluated as*

$$\delta[C/N] \le \sum_{g\in G} \mathrm{P}^G(g)^{1-st} \left(\sum_{g'\in C\setminus N} \mathrm{P}^G(gg')^t \right)^s$$

with arbitrary $s, t > 0$.

Then, we obtain the following theorem as a generalization of Theorem 5.1. The following theorem can be shown by replacing $G \setminus \{e\}$ by $G \setminus N$ in the proof of Theorem 5.1.

Theorem 5.5 *Given a* $(1, \epsilon)$*-double universal2 code ensemble* C_X *for* $N \subset G$*, the inequality* $\mathrm{E}_X \delta[C_X] \le \min_{0 \le s \le 1} \epsilon^s e^{\phi(s:\mathrm{P}^G)}$ *holds.*

Exercise 5.7 Define the code $C_{G,2}$ by the following generating matrix;

$$G_2 := \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{5.24}$$

where -1 is the inverse element of 1 with respect to addition. Show that the code space $C_{G,1}$ contains the code space $C_{G,2}$. So, we can use $C_{G,1}$ and $C_{G,2}$ as G and N, respectively.

5.2 General Theory for Quantum Error Correction

When a quantum system $\mathcal{H}$ is disturbed by noise during transmission, the state change is described by a TP-CP map on the quantum system $\mathcal{H}$ (See Sect. 2.1). To protect the state from the noise, we can restrict our state into a subspace $\mathcal{H}_0$. In this case, it is possible to remove a special type of noise by converting the state out of $\mathcal{H}_0$ to a state in $\mathcal{H}_0$. Such a subspace $\mathcal{H}_0$ is called an **encoder**, and a TP-CP map from $\mathcal{H}$ to $\mathcal{H}_0$ is called a **decoder**. The pair is called a **quantum error correction** or a **quantum code** simply.

For example, we assume that a quantum system $\mathcal{H}$ is given as a direct sum space $\oplus_{i=0}^{k} \mathcal{H}_i$ and the TP-CP map corresponding to the noisy channel has the Kraus representation:

$$\Lambda(\rho) := \sum_{i=0}^{k} \sqrt{p_i} U_i \rho \sqrt{p_i} U_i^\dagger, \tag{5.25}$$

where U_i is a unitary from $\mathcal{H}_0$ to $\mathcal{H}_i$ and p_i is a probability distribution. Then, we give the decoder by $D(\rho) := \sum_{i=0}^{k} (U_i P_i)^\dagger \rho P_i U_i$, where P_i is the projection to $\mathcal{H}_i$. When ρ is a state on $\mathcal{H}_0$, i.e., the support of ρ is included in $\mathcal{H}_0$, we have $D \circ \Lambda(\rho) = \rho$. When the noise is given as an ideal form like (5.25), the above mentioned combination of the encoder $\mathcal{H}_0$ and the decoder D can remove the noise even in the quantum situation (Fig. 5.3).

However, a general noise in a quantum communication channel cannot be written in an ideal form like (5.25). The performance of the quantum code $(\mathcal{H}_0, D)$ is evaluated by the entanglement fidelity $F_e^2(D \circ \Lambda_{\mathcal{H}_0})$ of the quantum channel $D \circ \Lambda_{\mathcal{H}_0}$. When the channel Λ has the Kraus representation $\Lambda(\rho) := \sum_j F_i \rho F_i^\dagger$, the channel

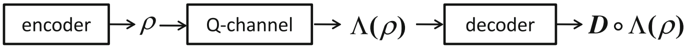


Fig. 5.3 Error correction in quantum communication channel

$D\circ\Lambda_{\mathcal{H}_0}$ has the Kraus representation $(D\circ\Lambda_{\mathcal{H}_0})(\rho)=\sum_{j,i}(U_iP_iF_j^\dagger)^\dagger\rho F_j^\dagger P_iU_i$. Hence, by using (2.47), the entanglement fidelity of the channel $D\circ\Lambda_{\mathcal{H}_0}$ is calculated to $F_e^2(D\circ\Lambda_{\mathcal{H}_0})=\sum_{j,i}\left|\frac{\operatorname{Tr}F_j^\dagger P_iU_i}{\operatorname{Tr}P_0}\right|^2$.

5.3 Code Based on Discrete Heisenberg Representation

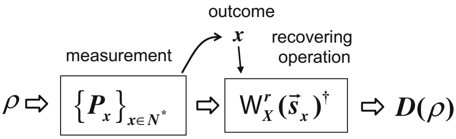
5.3.1 Stabilizer Code

Before starting this section, the author recommends the reader to revisit Sect. 3.3.1, which summarizing discrete Heisenberg representation, because this section constructs quantum error correction by using discrete Heisenberg representation when $\mathbb{X}$ expresses $\mathbb{F}_q$ or $\mathbb{Z}_d$ [14, 15, 30, 35]. Consider the irreducible projective unitary representation $\mathsf{W}_{\mathbb{X}}^r$ of $\mathbb{X}^{2r}$ on the representation space $\mathcal{H}$. Then, given a self orthogonal subgroup N of $\mathbb{X}^{2r}$, i.e., a subgroup N satisfying $N\subset N^\perp$, the representation space has the direct sum decomposition $\mathcal{H}=\oplus_{\vec{s}\in N^*}\mathcal{H}_x$. Here, $\mathsf{W}_{\mathbb{X}}^r(\vec{s})$ moves $\mathcal{H}_x$ to $\mathcal{H}_{x+[\vec{s}]}$. (For the detail of these two facts, see [44, Sect. 8.1.4].) Hence, using the projection P_x to $\mathcal{H}_x$, we have

$$\mathsf{W}_{\mathbb{X}}^r(\vec{s})P_x=P_{x+[\vec{s}]}\mathsf{W}_{\mathbb{X}}^r(\vec{s})P_x=P_{x+[\vec{s}]}\mathsf{W}_{\mathbb{X}}^r(\vec{s}). \tag{5.26}$$

Now, using this structure, we construct quantum error correction as follows. We choose the subspace $\mathcal{H}_0$ as an encoder. Then, we choose a representative $\vec{s}_x$ for each coset $x\in N^*=\mathbb{X}^{2r}/N^\perp$ priorly. At the decoding stage, we apply projective measurement $\{P_x\}_{x\in N^*}$ and obtain the outcome $x\in N^*$. When $x\neq 0$, the error moves the state from the original space $\mathcal{H}_0$ to the space $\mathcal{H}_x$. That is, we infer that an element of $x\subset\mathbb{X}^{2r}$ operates to move the state from $\mathcal{H}_0$ to $\mathcal{H}_x$. Hence, we guess that the representative $\vec{s}_x$ of x acts. So, as decoding process, we correct the error $\mathsf{W}_{\mathbb{X}}^r(\vec{s}_x)$ by applying $\mathsf{W}_{\mathbb{X}}^r(\vec{s}_x)^\dagger$. In summary, this decoding operation is mathematically described as $D(\rho):=\sum_{x\in N^*}\mathsf{W}_{\mathbb{X}}^r(\vec{s}_x)^\dagger P_x\rho P_x\mathsf{W}_{\mathbb{X}}^r(\vec{s}_x)$ as Fig. 5.4.

Fig. 5.4 Decoding operation for quantum error correction



Here, the set of representatives $\mathcal{J} := \{\vec{s}_x\}$ does not necessarily form a subgroup of $\mathbb{X}^{2r}$. Although the set of representatives forming a subgroup has some advantage, for an effective decoding, we need to choose the set $\{\vec{s}_x\}$ of representatives so that it does not form a subgroup. The above constructed quantum error correcting code based on the self orthogonal group N is called the **stabilizer code** with the **stabilizer** N. For example, in the Pauli channel $\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}](\rho) = \sum_{\vec{s}\in\mathbb{X}^{2r}} \mathrm{P}^{\mathbb{X}^{2r}}(\vec{s})\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rho\mathsf{W}^r_{\mathbb{X}}(\vec{s})^\dagger$, we can consider that the error $\mathsf{W}^r_{\mathbb{X}}(\vec{s})$ occurs with probability $\mathrm{P}^{\mathbb{X}^{2r}}(\vec{s})$. In the following, we simplify $\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]$ to $\Lambda[\mathrm{P}^{\mathbb{X}^{2r}}]$.

Now, we write $x = [\vec{s}]$ for $\vec{s} \in \mathbb{X}^{2r}$. Then, from [44, (8.25)], we have

$$\operatorname{Tr} \mathsf{W}^r_{\mathbb{X}}(\vec{s} - \vec{s}_x)|_{\mathcal{H}_0} = \begin{cases} 0 & \text{when } \vec{s} - \vec{s}_x \notin N \\ \operatorname{Tr} P_0 & \text{when } \vec{s} - \vec{s}_x \in N. \end{cases}$$

Due to (2.47), the entanglement fidelity of the quantum channel $D \circ \Lambda_{\mathcal{H}_0}$ is calculated to

$$F_e^2(D \circ \Lambda_{\mathcal{H}_0}) = \sum_{x\in N^*, \vec{s}\in N} \mathrm{P}^{\mathbb{X}^{2r}}(\vec{s}_x + \vec{s}) = \mathrm{P}^{\mathbb{X}^{2r}}(\mathcal{J} + N).$$

When $\mathcal{H}_{x'}$ is used as an encoder instead of $\mathcal{H}_0$, the decoder

$$D_{x'}(\rho) := \sum_{x\in N^*} \mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x-x'})^\dagger P_x \rho P_x \mathsf{W}_{\mathbb{X}}(\vec{s}_{x-x'})$$

has the same performance as the above code.

Since any element of N does not move any vector in $\mathcal{H}_0$ and $\mathcal{H}_x$ and the operation of any element of $N^\perp$ is closed in $\mathcal{H}_0$ and $\mathcal{H}_x$, we can naturally define a representation of the group $N^\perp/N$ on $\mathcal{H}_0$ and $\mathcal{H}_x$. In the following, we denote this representation by $\mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_x}$. In general, when we define the probability distribution $\mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}^{N^\perp/N}([\vec{s}]) := \mathrm{P}^{\mathbb{X}^{2r}}(\mathcal{J} + \vec{s} + N)$ for $[\vec{s}] \in N^\perp/N$, we have

$$\begin{aligned} & P[D_x \circ \Lambda_{\mathcal{H}_x}]([\vec{s}]) \\ = & \frac{1}{\dim \mathcal{H}_x^2} \langle\langle \mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_x}([\vec{s}])|((D_x \circ \Lambda_{\mathcal{H}_x}) \otimes id)(|I_{\mathcal{H}_x}\rangle\rangle\langle\langle I_{\mathcal{H}_x}|)|\mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_x}([\vec{s}])\rangle\rangle \\ = & \mathrm{P}^{\mathbb{X}^{2r}}(\mathcal{J} + \vec{s} + N) = \mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}^{N^\perp/N}([\vec{s}]). \end{aligned}$$

Since $D \circ \Lambda_{\mathcal{H}_x}$ is a Pauli channel, the relation $D \circ \Lambda_{\mathcal{H}_x} = \Lambda[\mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_x}, \mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}^{N^\perp/N}]$ holds. That is, the channel obtained via the above quantum error correction is the Pauli channel with the probability distribution $\mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}$.

Given a encoder $\mathcal{H}_0$ and a Pauli channel, the decoder to maximize the entanglement fidelity is given by the choice of the representative $\operatorname{argmax}_{\vec{s}_x\in x} \sum_{\vec{s}\in N} \mathrm{P}^{\mathbb{X}^{2r}}(\vec{s}_x + \vec{s})$ for any element $x \in N^* = \mathbb{X}^{2r}/N$. This decoder is called the **maximum likelihood decoder** of the stabilizer N under the probability distribution $\mathrm{P}^{\mathbb{X}^{2r}}$. Due to Lemma 8.7 of [44], the dimension of the encoder $\mathcal{H}_0$ is $|\mathbb{X}|^r/|N^*|$. In this way, the construction of decoder and the calculation of entanglement fidelity are treated via the discussion based on the probability distribution $\mathrm{P}^{\mathbb{X}^{2r}}$.

5.3.2 CSS (Calderbank-Shor-Smolin) Code

In this subsection, we construct a stabilizer code by using the structure $\mathbb{X}^{2r} = \mathbb{X}^r \times \mathbb{X}^r$ [16, 36, 118]. In the following, to distinguish the first and second components of $\mathbb{X}^{2r}$, we denote them by $\mathbb{X}_1^r$ and $\mathbb{X}_2^r$, respectively. Then, for a subgroup $C \subset \mathbb{X}^r$, we define the **orthogonal subgroup** $C^\perp := \{\boldsymbol{s} \in \mathbb{X}^r | (\boldsymbol{s}, \boldsymbol{t})_{\mathbb{X}} = 0,\ \forall \boldsymbol{t} \in C\}$ of C, where the inner product $(\boldsymbol{s}, \boldsymbol{t})_{\mathbb{X}}$ is defined in (3.15). So, we have the following lemma.

Lemma 5.5 *The relation $C_2^\perp \times C_1^\perp = (C_1 \times C_2)^\perp$ holds.*

Proof We can easily check the relation $C_2^\perp \times C_1^\perp \subset (C_1 \times C_2)^\perp$. Conversely, an element $(\boldsymbol{s}, \boldsymbol{t}) \in (C_1 \times C_2)^\perp$ satisfies $(\boldsymbol{s}', \boldsymbol{t})_{\mathbb{X}} = (\boldsymbol{s}, \boldsymbol{t}')_{\mathbb{X}}$ for any elements $\boldsymbol{s}' \in C_1, \boldsymbol{t}' \in C_2$. Hence, it is possible to choose $\boldsymbol{s}' = 0$ and $\boldsymbol{t}' = 0$, separately. For $\boldsymbol{s}' = 0$, we have $\boldsymbol{s} \in C_2^\perp$. For $\boldsymbol{t}' = 0$, we have $\boldsymbol{t} \in C_1^\perp$, which implies the desired statement. ■

When two subgroups $C_1, C_2 \subset \mathbb{X}^r$ satisfies the **torsion condition**: $C_1 \subset C_2^\perp$, we have the relation $C_2 \subset C_1^\perp$, which implies that $N := C_1 \times C_2 \subset \mathbb{X}^{2r}$ is a self orthogonal subgroup. The stabilizer code based on this self orthogonal subgroup is called a Calderbank-Shor-Smolin code (**CSS code**). Lemma 5.5 guarantees that $(C_1 \times C_2)^*$ is isomorphic to $\mathbb{X}^r / C_2^\perp \times \mathbb{X}^r / C_1^\perp$. Thus fact derives the dimension of the encoder $\mathcal{H}_0$ of the CSS code as follows.

$$\frac{|\mathbb{X}|^r}{|N^*|} = \frac{|\mathbb{X}|^r}{|\mathbb{X}^r / C_2^\perp| \cdot |\mathbb{X}^r / C_1^\perp|} = \frac{|C_2^\perp| \cdot |C_1^\perp|}{|\mathbb{X}|^r}. \tag{5.27}$$

Then, we can uniquely define the vector $|[\boldsymbol{s}]\rangle := \sum_{\boldsymbol{s}' \in [\boldsymbol{s}]} \frac{1}{\sqrt{|C_1|}} |\boldsymbol{s}'\rangle$ for a coset $[\boldsymbol{s}] \in C_2^\perp / C_1$. So, the encoder $\mathcal{H}_0$ has the CONS $\{|[\boldsymbol{s}]\rangle\}_{[\boldsymbol{s}] \in C_2^\perp / C_1}$. As a generalization of the above vector, for $([\boldsymbol{x}], [\boldsymbol{y}]) \in \mathbb{X}^r / C_2^\perp \times \mathbb{X}^r / C_1^\perp = (C_1 \times C_2)^*$ and $[\boldsymbol{s}] \in C_2^\perp / C_1$, we can uniquely define the vector:

$$|[\boldsymbol{s}], [\boldsymbol{x}], [\boldsymbol{y}]\rangle := \sum_{\boldsymbol{s}' \in [\boldsymbol{s}]} \frac{1}{\sqrt{|C_1|}} \omega_{\mathbb{X}}^{\boldsymbol{s}' \cdot \boldsymbol{y}} |\boldsymbol{s}' + \boldsymbol{x}\rangle.$$

Then, the space $\mathcal{H}_{([\boldsymbol{x}], [\boldsymbol{y}])}$ has the CONS $\{|[\boldsymbol{s}], [\boldsymbol{x}], [\boldsymbol{y}]\rangle\}_{[\boldsymbol{s}] \in C_2^\perp / C_1}$.

Lemma 5.5 guarantees that the pair of the representatives $\boldsymbol{s} \in x \subset \mathbb{X}^r$ and $\boldsymbol{t} \in y \subset \mathbb{X}^r$ of cosets $x \in \mathbb{X}^r / C_2^\perp$ and $y \in \mathbb{X}^r / C_1^\perp$ gives the representative $(\boldsymbol{s}, \boldsymbol{t}) \in (x, y) \subset \mathbb{X}^{2r}$ of the coset $(x, y) \in (C_1 \times C_2)^* = \mathbb{X}^{2r} / (C_1 \times C_2)$. Here, it is also possible to choose a representative $\boldsymbol{s}$ of the coset x dependently of the other coset y. The representative $\boldsymbol{t}$ of the coset y can be chosen in the same way.

When the noise is characterized by a Pauli channel, we can consider that the error $\vec{s} = (\boldsymbol{s}, \boldsymbol{t})$ occurs with probability $\mathrm{P}^{\mathbb{X}^{2r}}(\vec{s}) = \mathrm{P}^{\mathbb{X}^{2r}}(\boldsymbol{s}, \boldsymbol{t})$. The error $\boldsymbol{s}$ in the first component affects on the computational basis and is called the **bit error**. The error $\boldsymbol{t}$ in the second component affects on the phase and is called the **phase error**. Hence, in a CSS code, we can treat the corrections of the bit error and the phase error, separately. When two errors $\boldsymbol{s}$ and $\boldsymbol{t}$ act independently, i.e., the joint distribution $\mathrm{P}^{\mathbb{X}^{2r}}(\vec{s})$ is given as the product distribution $\mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s})\mathrm{P}^{\mathbb{X}_2^r}(\boldsymbol{t})$ of two distributions $\mathrm{P}^{\mathbb{X}_1^r}$ and $\mathrm{P}^{\mathbb{X}_2^r}$ on $\mathbb{X}^r$, we can independently choose the two representatives $\boldsymbol{s}_x$ and $\boldsymbol{t}_y$ of two cosets $x \in C_1^*$ and $y \in C_2^*$. Then, the entanglement fidelity of the channel $D \circ \Lambda_{\mathcal{H}_0}$ is calculated to

$$\begin{aligned}
&F_e^2(D \circ \Lambda_{\mathcal{H}_0}) \\
=& \sum_{x \in \mathbb{X}^r / C_2^\perp} \sum_{y \in \mathbb{X}^r / C_1^\perp} \sum_{\boldsymbol{s}' \in C_1} \sum_{\boldsymbol{t}' \in C_2} \mathrm{P}^{\mathbb{X}^{2r}}((\boldsymbol{s}_x, \boldsymbol{t}_y) + (\boldsymbol{s}', \boldsymbol{t}')) \\
=& \left(\sum_{x \in \mathbb{X}^r / C_2^\perp, \boldsymbol{s}' \in C_1} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}_x + \boldsymbol{s}') \right) \left(\sum_{y \in \mathbb{X}^r / C_1^\perp, \boldsymbol{t}' \in C_2} \mathrm{P}^{\mathbb{X}_2^r}(\boldsymbol{t}_y + \boldsymbol{t}') \right).
\end{aligned}$$

Then, the probability $\sum_{x \in \mathbb{X}^r / C_2^\perp, \boldsymbol{s}' \in C_1} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}_x + \boldsymbol{s}')$ is maximized when we apply the maximum likelihood decoder for the code pair $C_1 \subset C_2^\perp$. The other probability $\sum_{y \in \mathbb{X}^r / C_1^\perp, \boldsymbol{t}' \in C_2} \mathrm{P}^{\mathbb{X}_2^r}(\boldsymbol{t}_y + \boldsymbol{t}')$ is maximized in the same way. Hence, the maximum likelihood of the CSS code can be constructed by the simple combination of the maximum likelihood decoders of the two classical code pairs, which are the dual of the other code pair.

On the other hand, when two errors $\boldsymbol{s}$ and $\boldsymbol{t}$ are not independent of each other, it is better to choose their representatives dependently of each other than to choose them independently. In this case, the joint distribution is given as $\mathrm{P}^{\mathbb{X}^{2r}}(\boldsymbol{s}, \boldsymbol{t}) = \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s})\mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\boldsymbol{t}|\boldsymbol{s})$, where $\mathrm{P}^{\mathbb{X}_1^r}$ is the marginal distribution on $\mathbb{X}_1^r$ and $\mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\cdot|\boldsymbol{s})$ is the conditional distribution of $\mathbb{X}_2^r$ conditioned with $\mathbb{X}_1^r$. In the decoding stage, firstly, we choose the representative $\boldsymbol{s}_x \in x$ of the coset $x \in \mathbb{X}^r / C_2^\perp$. Then, we choose the representative $\boldsymbol{t}_{x,y} \in y$ of the coset $y \in \mathbb{X}^r / C_1^\perp$ dependently of the other coset x.

The entanglement fidelity of the quantum channel $D \circ \Lambda_{\mathcal{H}_0}$ is calculated to

$$\begin{aligned}
&F_e^2(D \circ \Lambda_{\mathcal{H}_0}) \\
=& \sum_{x \in \mathbb{X}^r / C_2^\perp} \sum_{y \in \mathbb{X}^r / C_1^\perp} \sum_{\boldsymbol{s}' \in C_1} \sum_{\boldsymbol{t}' \in C_2} \mathrm{P}^{\mathbb{X}^{2r}}((\boldsymbol{s}_x, \boldsymbol{t}_{x,y}) + (\boldsymbol{s}', \boldsymbol{t}')) \\
=& \sum_{x \in \mathbb{X}^r / C_2^\perp, \boldsymbol{s}' \in C_1} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}_x + \boldsymbol{s}') \left(\sum_{y \in \mathbb{X}^r / C_1^\perp, \boldsymbol{t}' \in C_2} \mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\boldsymbol{t}_{x,y} + \boldsymbol{t}' | \boldsymbol{s}_x + \boldsymbol{s}') \right).
\end{aligned}$$

To evaluate the entanglement fidelity $F_e^2(D \circ \Lambda_{\mathcal{H}_0})$, it is easier to evaluate $1 - F_e^2(D \circ \Lambda_{\mathcal{H}_0})$ because this probability corresponds to the error probability.

In the following, we focus on the following choice of the representative $\boldsymbol{t}_{x,y} \in y$ dependently of the other coset x. Considering that the error $\boldsymbol{t}$ is subject to the distribution $\mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\cdot|\boldsymbol{s}_x)$, we infer the that true error is $\boldsymbol{t}_{y:\boldsymbol{s}_x}$, where $\boldsymbol{t}_{y:\boldsymbol{s}} := \mathrm{argmax}_{\boldsymbol{t}\in y} \sum_{\boldsymbol{t}'\in C_2} \mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\boldsymbol{t}+\boldsymbol{t}'|\boldsymbol{s})$. Hence,

$$\begin{aligned}
&1 - F_e^2(D \circ \Lambda_{\mathcal{H}_0}) \\
\le &(1 - \sum_{x\in\mathbb{X}^r/C_2^\perp} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}_x)) \\
&+ \sum_{x\in\mathbb{X}^r/C_2^\perp} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}_x) \left(1 - \sum_{y\in\mathbb{X}^r/C_1^\perp, \boldsymbol{t}'\in C_2} \mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\boldsymbol{t}_{y:\boldsymbol{s}_x} + \boldsymbol{t}'|\boldsymbol{s}_x)\right) \\
\le &(1 - \sum_{x\in\mathbb{X}^r/C_2^\perp} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}_x)) \\
&+ \sum_{\boldsymbol{s}\in\mathbb{X}^r} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}) \left(1 - \sum_{y\in\mathbb{X}^r/C_1^\perp, \boldsymbol{t}'\in C_2} \mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\boldsymbol{t}_{y:\boldsymbol{s}} + \boldsymbol{t}'|\boldsymbol{s})\right).
\end{aligned} \tag{5.28}$$

In this way, the decoding can be reduced to the analyses of the decoding error probabilities of the classical code $C_2^\perp$ and the classical code pair $C_1^\perp/C_2$.

In the case of CSS codes, since $N^\perp/N = (C_2^\perp \times C_1^\perp)/(C_1 \times C_2) = C_2^\perp/C_1 \times C_1^\perp/C_2$, we have $\mathcal{J} = \{(\boldsymbol{s}_x, \boldsymbol{t}_{x,y})\}_{(x,y)\in N^*}$. Hence, we have $\mathrm{P}\{\mathcal{J} + N\}^{N^\perp/N} = \mathrm{P}\{\mathcal{J} + N\}^{C_2^\perp/C_1\times C_1^\perp/C_2}$. So, letting $\mathcal{J}_{2,x} := \{\boldsymbol{t}_{x,y}\}$, we obtain

$$\mathrm{P}\{\mathcal{J} + N\}^{C_1^\perp/C_2}(0) = \sum_{[\boldsymbol{x}]\in\mathbb{X}_1^r/C_2^\perp} \mathrm{P}^{\mathbb{X}_1^r/C_2^\perp}([\boldsymbol{x}])\mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r/C_2^\perp}(\mathcal{J}_{2,[\boldsymbol{x}]} + C_2|[\boldsymbol{x}]).$$

Exercise 5.8 Define the codes $C_{G,3}$ and $C_{G,4}$ by the following generating matrix;

$$G_3 := \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad G_4 := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 \\ 0 & -1 & -1 & 1 \end{pmatrix}. \tag{5.29}$$

Show that $C_{G,1} = C_{G,3}^\perp$ and $C_{G,2} = C_{G,4}^\perp$.

Exercise 5.9 We assume that the noise $n \in \mathbb{F}_q^7$ obeys the 7-trial independent and identical distribution of the probability distribution given in Exercise 5.3. Show that the correctly decoding probability is larger than $(1-p)^6(1+6p)$ under the above Hamming code $C_{G,4}$ with 7 digits when the decoder satisfies the condition given in Exercise 5.5.

Exercise 5.10 Show that $C_{G,2} \times C_{G,3}$ satisfies the torsion condition.

Exercise 5.11 Assume that the joint distribution $\mathrm{P}^{\mathbb{X}^{2r}}$ is a product distribution, i.e., $\mathrm{P}^{\mathbb{X}^{2r}}(\boldsymbol{s},\boldsymbol{t}) = \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s})\mathrm{P}^{\mathbb{X}_2^r}(\boldsymbol{t})$. We also assume that both distributions $\mathrm{P}^{\mathbb{X}_1^r}$ and $\mathrm{P}^{\mathbb{X}_2^r}$ satisfy conditions given in Exercise 5.3. Show that for the CSS code based on $C_{G,2} \times C_{G,3}$ there exists a decoder $\{\boldsymbol{s}_x\}_{x\in\mathbb{X}^r/C_{G,3}^{\perp}}, \{\boldsymbol{t}_y\}_{y\in\mathbb{X}^r/C_{G,2}^{\perp}}$ such that $1 - F_e^2(D \circ \Lambda_{\mathcal{H}_0}) \le 2(1-(1-p)^6(1+6p))$.

5.3.3 Asymptotic Theory

In the previous subsection, we have addressed the case when a Pauli channel is given by a general distribution $\mathrm{P}^{\mathbb{X}^{2r}}$ on $\mathbb{X}^{2r}$. In the following, we address the case when the Pauli channel is given by the r-fold independent and identical distribution of the distribution $\mathrm{P}^{\mathbb{F}_q^2}$ on $\mathbb{F}_q^2$. Similar to the classical case, the theory with infinitely large r is called the asymptotic theory. In the asymptotic theory, we discuss a sequence of codes. In the following, we deal only with CSS codes. So, we discuss a sequence of $\{(C_{1,r}, C_{2,r})\}$ of code pairs satisfying the torsion condition. The limits of the entanglement fidelity and the transmission rate play important roles in the asymptotic theory because the latter describes the asymptotic behavior of the dimension of transmitted system. Since the dimension of transmitted system is $\frac{|C_{1,r}^{\perp}|\cdot|C_{2,r}^{\perp}|}{|\mathbb{F}_q|^r}$ due to (5.27), the transmission rate is calculated to

$$\lim_{r\to\infty} \frac{1}{r}\log\frac{|C_{1,r}^{\perp}|\cdot|C_{2,r}^{\perp}|}{|\mathbb{F}_q|^r} = \lim_{r\to\infty}\frac{\log|C_{1,r}^{\perp}||C_{2,r}^{\perp}|}{r} - \log|\mathbb{F}_q|. \tag{5.30}$$

In contrast, we impose the former to be 1.

Now, we denote the first and second components of $\mathbb{F}_q^2$ by $\mathbb{F}_{q,1}$ and $\mathbb{F}_{q,2}$. Choosing R_1 and R_2 satisfying

$$R_1 > \frac{H(\mathrm{P}^{\mathbb{F}_{q,1}})}{\log q}, \quad R_2 > \frac{H(\mathrm{P}^{\mathbb{F}_{q,2}|\mathbb{F}_{q,1}})}{\log q}, \tag{5.31}$$

we select the integers $k_1 = \lfloor(1-R_1)r\rfloor$ and $k_2 = \lfloor(1-R_2)r\rfloor$. Then, we choose the code ensemble $C_{1,X,r}^{\perp}$ to be the q^{k_1-r}-universal2 code ensemble given in Lemma 5.2. So, the code ensemble $C_{1,X,r}$ is automatically decided according to the code ensemble $C_{1,X,r}^{\perp}$. Further, combining Lemmas 5.2 and 5.3, we obtain the (1,q^{k_2-r})-double universal2 code ensemble $C_{2,X,r}^{\perp}$ for the subgroup $C_{1,X,r} \subset \mathbb{F}_q^r$ [53, 126]. Then, the dimension of $C_{1,X,r}^{\perp}$ is k_1 and the dimension of $C_{2,X,r}^{\perp}$ is k_2. Hence, since the dimension of the transmitted quantum system is $q^{k_1+k_2-r}$, the transmission rate is $\log q(1-R_1-R_2)$.

On the other hand, by the same discussion as the proofs of Theorems 5.1 and 5.3, the RHS of (5.28) can be evaluated as follows.

$$
\begin{aligned}
&(1 - \sum_{x \in \mathbb{X}^r / C_2^\perp} \mathrm{P}^{\mathbb{F}_{q,1}}(\boldsymbol{s}_x)) \\
&+ \sum_{\boldsymbol{s} \in \mathbb{X}^r} \mathrm{P}^{\mathbb{F}_{q,1}}(\boldsymbol{s}) \left(1 - \sum_{y \in \mathbb{X}^r / C_1^\perp, \boldsymbol{t}' \in C_2} \mathrm{P}^{\mathbb{F}_{q,2}|\mathbb{F}_{q,1}}(\boldsymbol{t}_{y:\boldsymbol{s}} + \boldsymbol{t}' | \boldsymbol{s}) \right) \\
\le & \min_{0 \le s \le 1} q^{s(k_1 - r)} e^{\phi(s:\mathrm{P}^{\mathbb{F}_{q,1}})} + \min_{0 \le s \le 1} q^{s(k_2 - r)} e^{\phi(s:\mathrm{P}^{\mathbb{F}_{q,2}|\mathbb{F}_{q,1}})}.
\end{aligned}
$$

Hence, due to the condition (5.31), this value approaches to zero exponentially for r. Thus, the following transmission rate can be attained [36].

$$
\log q - H(\mathrm{P}^{\mathbb{F}_{q,1}}) - H(\mathrm{P}^{\mathbb{F}_{q,2}|\mathbb{F}_{q,1}}) = \log q - H(\mathrm{P}^{\mathbb{F}_q^2}).
$$

This value equals the coherent information $I_c(\rho_{\mathrm{mix}}, \Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}])$.

5.3.4 Physical Implementation

Firstly, we discuss the physical implementation of the stabilizer code, which is a typical quantum code. Assume that the discrete Heisenberg representation of $\mathbb{F}_q^{2r}$ are given on the quantum system $\mathcal{H}^{\otimes r}$. Then, we consider the subgroup N_k of $\mathbb{F}_q^{2r}$ that is composed of vectors whose final $r - k$ entries are zero. Since the subgroup N_k is a self orthogonal group, we can implement the stabilizer code with the stabilizer N_k easily. Then, the subgroup $N_k^\perp$ is composed of vectors in $\mathbb{F}_q^{2r}$ whose only $r + 1$-th to $r + k$ entries are zero. In this stabilizer code, the encoder is the subspace of $\mathcal{H}^{\otimes r}$ composed of the vector whose first k component is $|0\rangle^{\otimes k}$. In the following, this subspace is described by $|0\rangle^{\otimes k} \otimes \mathcal{H}^{\otimes r-k}$, and is isometric to $\mathcal{H}^{\otimes r-k}$. The decoder is implemented as follows. Firstly, we measure the first k components in the bit basis, and obtain an element x of $\mathbb{F}_q^{2r}/N^\perp = N^*$ as the outcome. Then, we apply the inverse of the unitary $\mathsf{W}_{\mathbb{F}}^r(\vec{s}_x)$ corresponding to the representative $\vec{s}_x$. So, we complete the decoding.

However, the self orthogonal subgroup N does not have such a simple structure in general. When N is given as a vector subspace of $\mathbb{F}_q$, the above construction can be applied as follows. Let k be the dimension of N. The subgroups N and $N^\perp$ are isometric to the subgroups N_k and $N_k^\perp$ as vector subspaces on the finite filed $\mathbb{F}_q$ including the inner products. The symplectic group $\mathrm{Sp}(2r, \mathbb{F}_q)$ is the set of linear maps preserving the inner product on $\mathbb{F}_q^{2r}$. Hence, there is a suitable element $g \in \mathrm{Sp}(2r, \mathbb{F}_q)$ such that $gN_k = N$ and $gN_k^\perp = N^\perp$. That is, firstly, we prepare the state to be sent in the system $\mathcal{H}^{\otimes r-k}$. Then, we apply the unitary $\mathsf{S}_{\mathbb{F}}^r(g)$. So, the encoding is completed. The decoding can be done as follows. We apply the unitary $\mathsf{S}_{\mathbb{F}}^r(g)^\dagger$, measure the first k components with the computational basis, and apply the recovering unitary dependently of the measurement outcome [31, 32].

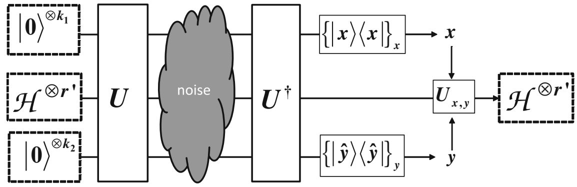

$|\mathbf{0}\rangle^{\otimes k_1}$
$\mathcal{H}^{\otimes r'}$
$|\mathbf{0}\rangle^{\otimes k_2}$
U
noise
$U^\dagger$
$\{|x\rangle\langle x|\}_x$
x
$U_{x,y}$
$\mathcal{H}^{\otimes r'}$
$\{|\hat{y}\rangle\langle\hat{y}|\}_y$
y

Fig. 5.5 Physical implementation of error correction based on CSS code: $\{|x\rangle\langle x|\}_x$ is the measurement based on the computational basis, and $\{|\hat{y}\rangle\langle\hat{y}|\}_y$ is the measurement based on the computational basis. $U_{x,y}$ is the abbreviation of $\mathsf{W}^r_{\mathbb{F}}(\gamma \boldsymbol{s}_{x,y}, 0)\mathsf{W}^r_{\mathbb{F}}(0, (\gamma^{-1})^T \boldsymbol{t}_{x,y})$, and r' is $r - k_1 - k_2$

Further, the encoder can be constructed in a more concrete form for a CSS code. Firstly, we focus on a typical CSS code as follows. Let $C_{1:k_1}$ be a subgroup of $\mathbb{F}_q^r$ containing non-zero entries only in first k_1 components, and $C_{2:k_2}$ be a subgroup of $\mathbb{F}_q^r$ containing non-zero entries only in last k_2 components. The encoding map is the map from $\mathcal{H}^{\otimes r-k_1-k_2}$ to $|0\rangle^{\otimes k_1} \otimes \mathcal{H}^{\otimes r-k_1-k_2} \otimes |0\rangle^{\otimes k_2}$, which is given as $|v\rangle := \sum_x v_x|x\rangle \mapsto |0\rangle^{\otimes k_1} \otimes |v\rangle \otimes |0\rangle^{\otimes k_2}$. The decoder is realized by measuring the first k_1 systems and the last k_2 systems and applying the appropriate unitary in these system dependently of the measurement outcomes.

When the general CSS code is given by the pair of vector subspaces C_1 and C_2 of $\mathbb{F}_q^r$ satisfying the torsion condition, its encoding operation is implemented as follows (See Fig. 5.5). Let k_1 and k_2 be the dimensions of C_1 and C_2, respectively. Also, let a $k_1 \times r$ matrix γ_1 be the generating matrix of the code C_1. We choose a $(r - k_1 - k_2) \times r$ matrix γ_2 such that the $(r - k_2) \times r$ matrix $\gamma_1\gamma_2$ is a generating matrix of $C_2^\perp$. Then, we choose a $k_2 \times r$ matrix γ_3 such that the matrix $\gamma := \gamma_1\gamma_2\gamma_3$ belongs to $\mathrm{GL}(r, \mathbb{F}_q)$. Now, we divide the square matrix $(\gamma^{-1})^T$ into the form $\beta_1\beta_2\beta_3$ in the same way so that β_3 is a generating matrix of the code C_2. This fact can be confirmed by the orthogonal relation between $r - k_2$ column vectors of $\gamma_1\gamma_2$ and k_2 column vectors of β_3. Similarly, the $(r - k_1) \times r$ matrix $\beta_2\beta_3$ is a generating matrix of the code $C_1^\perp$. Hence, the encoding operation of the CSS code is given as the combination of the applications of the matrix γ in the computational basis and the matrix $(\gamma^{-1})^T$ in the dual computational basis. Fortunately, the unitary matrix $U := \mathsf{S}^r_{\mathbb{F}}(M_\gamma)$ defined in [44, (8.36)] is the combination of both operations. (See the sentences after (8.37) in [44].) The decoder can be implemented as follows. Firstly, we apply the unitary $\mathsf{S}^r_{\mathbb{F}}(M_\gamma)^\dagger$. Then, we measure the first k_1 systems with the computational basis and the last k_2 systems with the dual computational basis. Finally, according to the measurement outcomes x and y, we apply the inverse matrices of the unitary corresponding to the representatives $\boldsymbol{s}_{x,y} \in \mathbb{F}^r_{q,1}$ and $\boldsymbol{t}_{x,y} \in \mathbb{F}^r_{q,2}$. That is, we apply the unitaries $\mathsf{W}^r_{\mathbb{F}}(\gamma \boldsymbol{s}_{x,y}, 0)$ and $\mathsf{W}^r_{\mathbb{F}}(0, (\gamma^{-1})^T \boldsymbol{t}_{x,y})$. Therefore, the encoder and the decoder of a CSS code can be realized by the representations of $\mathrm{GL}(r, \mathbb{F}_q)$ and discrete Heisenberg group [33].

On the other hand, we can discuss the case with $\mathbb{X} = \mathbb{Z}_d$ as follows. In a stabilizer code, when N is isometric to $\mathbb{Z}_d^k$ as a commutative group, there exists an element of $\mathrm{Sp}(2r, \mathbb{Z}_d)$ that transforms N_k to N. So, the discussion similar to $\mathbb{F}_q$ is applicable.

In a CSS code, when the subgroups C_1 and C_2 satisfy the torsion condition and are isometric to $\mathbb{Z}_d^{k_1}$ and $\mathbb{Z}_d^{k_2}$, respectively, we can choose the generation matrices of C_1 and $C_2^{\perp}$. So, the discussion similar to $\mathbb{F}_q$ is applicable even for a CSS code.

5.4 Clifford Code*

Quantum error correcting codes with algebraic structure are not limited to stabilizer codes. More generally, such an algebraic code can be constructed from an irreducible unitary representation f on $\mathcal{H}$ of a group G with its normal subgroup N. A code with the above construction is called a Clifford code [81, 82, 84]. In the following, we prepare several definitions for a Clifford code based on the literatures [81, 82, 84].

Let $\hat{N}(\mathsf{f})$ be the set of labels of irreducible unitary representations of N appearing in the irreducible unitary representation f on $\mathcal{H}$ of G as (5.33). For an element $\lambda \in \hat{N}(\mathsf{f})$, we define the **inertia group** $T(N,\lambda)$ as follows. (For the definition of the inertia group, see [44, Sect. 2.5.2].)

$$T(N,\lambda) := \{g \in G | \chi_\lambda(gng^{-1}) = \chi_\lambda(n), \quad \forall n \in N\}. \tag{5.32}$$

Due to the definition [44, Sect. 2.5.2], the inertia group $T(N,\lambda)$ contains N as a normal subgroup. Then, we obtain the following theorem for irreducible decomposition of the normal subgroup N.

Theorem 5.6 *Given an irreducible unitary representation* f *of a group G on $\mathcal{H}$, the representation space $\mathcal{H}$ has the following irreducible decomposition for the repre sentation of a normal subgroup N of G.*

$$\mathcal{H} = \bigoplus_{\lambda \in \hat{N}(\mathsf{f})} \mathcal{U}_\lambda \otimes \mathbb{C}^m. \tag{5.33}$$

That is, the multiplicity m does not depend on $\lambda \in \hat{N}(\mathsf{f})$. Then, the subspace $\mathcal{K}_\lambda := \mathcal{U}_\lambda \otimes \mathbb{C}^m$ of $\mathcal{H}$ is an irreducible representation space for the representation of the inertia group $T(N,\lambda)$. Further, the irreducible representation space $\mathcal{U}_\lambda$ has the same dimension for any $\lambda \in \hat{N}(\mathsf{f})$, and the projection $P(\mathcal{K}_\lambda)$ to $\mathcal{K}_\lambda$ is written as follows.

$$P(\mathcal{K}_\lambda) = \frac{\chi_\lambda(e)}{|N|} \sum_{n \in N} \chi_\lambda(n^{-1}) \mathsf{f}(n). \tag{5.34}$$

Proof First, we make the irreducible decomposition $\mathcal{H} = \bigoplus_{\lambda \in \hat{N}(\mathsf{f})} \mathcal{U}_\lambda \otimes \mathbb{C}^{m_\lambda}$ for the representation of N. Then, for any element $\lambda \in \hat{N}(\mathsf{f})$, we define the subspace $\mathcal{K}_\lambda := \mathcal{U}_\lambda \otimes \mathbb{C}^{m_\lambda}$. Given an element $g \in T(N,\lambda)$, the representation $n (\in N) \mapsto \mathsf{f}_\lambda(gng^{-1})$ of N is the same irreducible representation as f_λ because the character uniquely identifies the representation. Hence, we have $\mathsf{f}_\lambda(n) = \mathsf{f}_\lambda(gng^{-1})$ for $n \in N$. Thus, denoting an element of $\mathsf{f}(g)\mathcal{U}_\lambda$ by $\mathsf{f}(g)|u\rangle$, we have

$$\begin{aligned}\mathsf{f}(n)\mathsf{f}(g)|u\rangle &= \mathsf{f}(g)\mathsf{f}(g^{-1})\mathsf{f}(n)\mathsf{f}(g)|u\rangle = \mathsf{f}(g)\mathsf{f}(g^{-1}ng)|u\rangle \\ =&\mathsf{f}(g)\mathsf{f}_\lambda(g^{-1}ng)|u\rangle = \mathsf{f}(g)\mathsf{f}_\lambda(n)|u\rangle.\end{aligned}$$

Schur's lemma guarantees that the representations of N in $\mathsf{f}(g)\mathcal{U}_\lambda$ is equivalent to f_λ.

On the other hand, given $g \in G \setminus T(N, \lambda)$, the representation $n \mapsto \mathsf{f}_{g(\lambda)}(n) := \mathsf{f}_\lambda(gng^{-1})$ of N is a different representation from f_λ. Since

$$\mathsf{f}(n)\mathsf{f}(g)|u\rangle = \mathsf{f}(g)\mathsf{f}(g^{-1})\mathsf{f}(n)\mathsf{f}(g)|u\rangle = \mathsf{f}(g)\mathsf{f}(g^{-1}ng)|u\rangle = \mathsf{f}(g)\mathsf{f}_{g(\lambda)}(n)|u\rangle,$$

a representation of N in $\mathsf{f}(g)\mathcal{U}_\lambda$ is equivalent to $\mathsf{f}_{g(\lambda)}$. So, the dimension of the representation space of $\mathsf{f}_{g(\lambda)}$ equals the dimension of $\mathcal{U}_\lambda$. Hence, all irreducible representations in $\hat{N}(\mathsf{f})$ have the same dimension.

Since $\mathsf{f}(g)$ moves $\mathcal{K}_\lambda$ to $\mathcal{K}_{g(\lambda)}$ and $\mathsf{f}(g^{-1})$ moves $\mathcal{K}_{g(\lambda)}$ to $\mathcal{K}_\lambda$, the dimension of $\mathcal{K}_\lambda$ does not depend on $\lambda \in \hat{N}(\mathsf{f})$. Hence, m_λ does not depend on $\lambda \in \hat{N}(\mathsf{f})$.

Any irreducible space $\mathcal{U}_\lambda$ for the representation for N is included in an irreducible subspace for the representation of $T(N, \lambda)$ because N is included in $T(N, \lambda)$. Since $\cup_{g\in G}\{\mathsf{f}(g)\mathcal{U}_\lambda$ generates $\mathcal{H}$, we have

$$\oplus_{g\in T(N,\lambda)}\mathsf{f}(g)\mathcal{U}_\lambda = (\oplus_{g\in G}\mathsf{f}(g)\mathcal{U}_\lambda) \cap \mathcal{K}_\lambda = \mathcal{K}_\lambda.$$

So, the space $\mathcal{K}_\lambda$ is irreducible for the representation of $T(N, \lambda)$. Further, (5.34) follows from (2.61) of [44]. ■

For an element $[g] = gT(N, \lambda)$ of the quotient space $G/T(N, \lambda)$, we define $g(\lambda) \in \hat{N}(\mathsf{f})$ as $\mathsf{f}(g)\mathcal{K}_\lambda = \mathcal{K}_{g(\lambda)}$. Since $g(\lambda)$ does not depend on the choice of the representative g of $[g]$, we denote $g(\lambda)$ by $[g](\lambda)$. Conversely, when $g(\lambda) = g'(\lambda)$ for $g, g' \in G$, $[g] = [g'] \in G/T(N, \lambda)$. Hence, we can identify $\hat{N}(\mathsf{f})$ with the quotient space $G/T(N, \lambda)$. Since any element $g' \in T(N, [g]\lambda)$ satisfies

$$\mathsf{f}(gg'g^{-1})\mathsf{f}(g)\mathcal{K}_\lambda = \mathsf{f}(g)\mathsf{f}(g')\mathsf{f}(g^{-1})\mathsf{f}(g)\mathcal{K}_\lambda = \mathsf{f}(g)\mathsf{f}(g')\mathcal{K}_\lambda = \mathsf{f}(g)\mathcal{K}_\lambda,$$

we have $gT(N, \lambda)g^{-1} \subset T(N, [g]\lambda)$. Replacing the roles of $[g]\lambda$ and λ, we have

$$gT(N, \lambda)g^{-1} = T(N, [g]\lambda). \tag{5.35}$$

Schur's lemma guarantees that $\mathsf{f}(g)$ is a constant on $\mathcal{K}_\lambda$ for $g \in C(N)$.

The above given structure can construct quantum error correcting code as follows. Given $\lambda \in \hat{N}(\mathsf{f})$, we fix the encoder $\mathcal{K}_\lambda$. For an element x of the quotient space $G/T(N, \lambda)$, we choose the representative g_x so that the decoder D is given as

$$D(\rho) := \sum_{x\in G/T(N,\lambda)} (P_{x(\lambda)}\mathsf{f}(g_x))^\dagger \rho P_{x(\lambda)}\mathsf{f}(g_x).$$

The pair of the encoder and the decoder given here is called a **Clifford code** [81, 82]. The normal subgroup N is called the **generating group** of the code. When only an element of the subset $\{g_x\}C(N)$ acts on the system $\mathcal{H}$, we can recover the original state perfectly. That is, when the probability distribution P on G has the support as the subset of the subset $\{g_x\}C(N)$ of G and the quantum channel $\Lambda[\mathrm{P}]$ is defined as

$$\Lambda[\mathrm{P}](\rho) := \sum_{g \in \{g_x\}C(N)} \mathrm{P}(g)\mathsf{f}(g)\rho\mathsf{f}(g)^\dagger,$$

the entanglement fidelity $F_e^2(D \circ \Lambda[\mathrm{P}])$ is 1.

Given a state $|x\rangle \in \mathcal{U}_\lambda$, we define the quantum error correcting code with the encoder $|x\rangle \otimes \mathbb{C}^m \subset \mathcal{K}_\lambda$ and the decoder D as

$$D(\rho) := \mathrm{Tr}_{\mathcal{U}_\lambda} \sum_{x \in G/T(N,\lambda)} (P_{x(\lambda)}\mathsf{f}(g_x))^\dagger \rho P_{x(\lambda)}\mathsf{f}(g_x).$$

This code is called a **subsystem code** [83]. This code can perfectly recover the original state when an element of the subset $\{g_x\}N$ acts on the representation space. That is, when the probability distribution P on G has the support included in the subset $\{g_x\}N$ and the quantum channel $\Lambda[\mathrm{P}]$ is defined as

$$\Lambda[\mathrm{P}](\rho) := \sum_{g \in \{g_x\}N} \mathrm{P}(g)\mathsf{f}(g)\rho\mathsf{f}(g)^\dagger,$$

the entanglement fidelity $F_e^2(D \circ \Lambda[\mathrm{P}])$ equals 1. Here, we list several important properties as follows.

Lemma 5.6 *When N is a normal subgroup of G, the following conditions hold.*

(1) *The center $C(N)$ of N is a normal subgroup of G.*
(2) *An irreducible unitary representation f of G and $\lambda \in \hat{N}(\mathsf{f})$ satisfy $C_G(N) \subset T(N, \lambda)$.*

Proof Elements $n \in C(N)$, $n' \in N$ and $g \in G$ satisfy $gn'g^{-1}n = ngn'g^{-1}$. Multiplying g^{-1} from the left and g from the right, we have $n'g^{-1}ng = g^{-1}ngn'$. Hence, we obtain $g^{-1}ng \in C(N)$. So, $C(N)$ is a normal subgroup of G.

Then, elements $g \in C_G(N)$ and $n \in N$ satisfy $\mathsf{f}_\lambda(g^{-1}ng) = \mathsf{f}_\lambda(n)$. So, we have $g \in T(N, \lambda)$. ■

Lemma 5.7 *Assume that $[G, G] \subset C(G)$. An irreducible unitary representation f of a group G satisfies the following conditions.*

(1) *For a label $\lambda \in \hat{N}(\mathsf{f})$, $T(N, \lambda)$ is a normal subgroup of G. Hence, due to (5.35), $T(N, \lambda)$ does not depend on the choice of $\lambda \in \hat{N}(\mathsf{f})$. In the following, we denote $T(N, \lambda)$ by $T(N, \mathsf{f})$.*
(2) *The inertial group for $T(N, \mathsf{f})$ is $T(N, \mathsf{f})$.*

(3) *Given two normal subgroups N_1 and N_2 of G satisfying $N_1 \subset N_2$, the relation $T(N_2, \mathsf{f}) \subset T(N_1, \mathsf{f})$ holds.*

Proof Given elements $g' \in T(N, \lambda), n \in N$, and $g \in G$, we choose $a := gg'g^{-1}g'^{-1} \in C(G)$. Since $gg'g^{-1} = ag'$, we have

$$gg'g^{-1}ngg'^{-1}g^{-1} = ag'ngg'^{-1}g^{-1} = g'ngg'^{-1}g^{-1}a = g'ng'^{-1},$$

which implies that $\chi_\lambda(gg'g^{-1}ngg'^{-1}g^{-1}) = \chi_\lambda(g'ng'^{-1}) = \chi_\lambda(n)$. Hence, we have $gg'g^{-1} \in T(N, \lambda)$. So, $T(N, \lambda)$ is a normal subgroup of G. Thus, we obtain **(1)**.

For distinct two elements $\lambda, \lambda' \in \hat{N}(\mathsf{f})$, the representation of $T(N, \mathsf{f})$ on $\mathcal{K}_\lambda$ is different from the representation on $\mathcal{K}_{\lambda'}$. This fact can be checked by considering the representation of the subgroup N. Due to the one-to-one relation between the character and the equivalent class of irreducible representation, the set $G \setminus T(N, \mathsf{f})$ does not contain any element of the inertial group of $T(N, \mathsf{f})$. Hence, we obtain **(2)**.

Comparing the condition for elements of $T(N_1, \mathsf{f})$ and the condition for elements of $T(N_2, \mathsf{f})$, we obtain **(3)**. ■

Lemma 5.8 *We assume that $[G, G] \subset C(G)$ and that an irreducible unitary representation f of the group G on $\mathcal{H}$ is faithful on $C(G)$. That is, an element $g \in C(G)$ satisfying $\mathsf{f}(g) \neq I$ is limited to the unit element. Then, the following conditions hold.*

(1) $\dim \mathcal{H}^2 = \frac{|G|}{|C(G)|}$.
(2) *A normal subgroup N of G satisfies $T(N, \mathsf{f}) = C_G(C(N))$.*
(3) *A normal subgroup N of G satisfies $T(N, \mathsf{f}) = T(C(N), \mathsf{f})$.*
(4) *When two normal subgroups N_1 and N_2 of G satisfy $C(N_2) \subset N_1 \subset N_2$, we have $T(N_2, \mathsf{f}) = T(N_1, \mathsf{f})$.*
(5) *For a normal subgroup N, any label $\lambda \in \hat{N}(\mathsf{f})$ satisfies the following conditions.*

$$\dim \mathcal{K}_\lambda = \frac{|C(G) \cap N|}{|N|} \dim \mathcal{H}^2_\lambda \dim \mathcal{H} \tag{5.36}$$

$$\dim \mathcal{H}^2_\lambda = \frac{|N|}{|C(N)|}. \tag{5.37}$$

Proof We will show **(1)** finally.
Proof of **(2)**: Due to **(2)** of Lemma 5.6, it is sufficient to show that $T(N, \mathsf{f}) \subset C_G(C(N))$. Given $\lambda \in \hat{N}(\mathsf{f})$ and $g \in T(N, \mathsf{f})$, we choose $n \in C(N)$. So, we have $a := gng^{-1}n^{-1} \in [G, G] \subset C(G)$, which implies that $\mathsf{f}_\lambda(a)$ is a constant ω. Since $g \in T(N, \mathsf{f})$, we have

$$\chi_\lambda(n) = \chi_\lambda(gng^{-1}) = \chi_\lambda(an) = \omega\chi_\lambda(n).$$

Lemma 2.6 of [44] guarantees that $n \in \mathrm{supp}(\chi_\lambda)$. Thus, $\omega = 1$. This fact implies that $gng^{-1}n^{-1} = e$. Therefore, $g \in C_G(C(N))$, which yields **(2)**.
Proof of **(3)**: Substituting $C(N)$ into N in **(2)**, we have $T(C(N), \mathsf{f}) = C_G(C(N))$, which implies **(3)**.

Proof of **(4)**: Lemma 5.7 guarantees $T(N_2, \mathsf{f}) \subset T(N_1, \mathsf{f}) \subset T(C(N_2), \mathsf{f})$. So, **(2)** yields **(4)**.
Proof of **(5)**: Since $\dim \mathcal{K}_\lambda = \mathrm{Tr}\, P(\mathcal{K}_\lambda)$, we calculate the trace of RHS of (5.34) in Theorem 5.6. When an element n of N satisfies that $\mathrm{Tr}\, \mathsf{f}(n)$ is not 0, the element n must be contained in $C(G)$. (For the detail, see [44, Lemma 2.5].) For $n \in N \cap C(G)$, $\mathsf{f}(n)$ and $\mathsf{f}_\lambda(n)$ are the same constant ω times of the identity matrix. Hence, $\chi_\lambda(n^{-1})\mathsf{f}(n) = \dim \mathcal{U}_\lambda I_{\mathcal{H}}$. Thus, (5.34) of Theorem 5.6 yields

$$\mathrm{Tr}\, P(\mathcal{K}_\lambda) = \frac{\chi_\lambda(e)}{|N|} \sum_{n \in N \cap C(G)} \dim \mathcal{U}_\lambda \dim \mathcal{H} = \frac{|C(G) \cap N|}{|N|} \dim \mathcal{H}_\lambda^2 \dim \mathcal{H},$$

which implies (5.36).

Due to **(3)**, when the generating group of the code is replaced by $C(N)$, we have the same inertia group. Then, $\mathcal{K}_\lambda$ is irreducible for the representation of the inertia group $T(C(N), \mathsf{f})$. Hence, applying (5.36) to the case when the generating group of the code is $C(N)$, we have

$$\mathrm{Tr}\, P(\mathcal{K}_\lambda) = \frac{|C(G) \cap C(N)|}{|C(N)|} \dim \mathcal{H}. \tag{5.38}$$

Since $C(G) \cap C(N) = C(G) \cap N$, the comparison between (5.36) and (5.38) yields (5.37).
Proof of **(1)**: Applying (5.37) to the case when the generating group of the code is $C(N)$, we obtain **(1)**. ■

A stabilizer code can be regarded as a special case of a Clifford code as follows. Consider the discrete Heisenberg group $\mathrm{H}(\mathbb{X}, 2r)$ as the group G. We choose a self orthogonal subgroup N of $\mathbb{X}^{2r}$. We choose the extension of $C(\mathrm{H}(\mathbb{X}, 2r))$ by N, which is a normal subgroup of $\mathrm{H}(\mathbb{X}, 2r)$ as the generating group of the code. Then, the stabilizer group with the stabilizer N is the Clifford code with the above generating group.

On the other hand, the subsystem code is useful for the following irreducible unitary representation of the semi direct product $H \rtimes K$. Given an irreducible unitary representation f_H of H, an irreducible unitary representation f'_K of K, and an adjoint representation f_K of f_H, which is a unitary representation of K, we can define the irreducible unitary representation $\mathsf{f}_H \rtimes \mathsf{f}_K \otimes \mathsf{f}'_H$ of the semi direct product $H \rtimes K$. When the normal subgroup H is the generating group of the code under this representation, the representation space of the irreducible unitary representation f'_K of K is the encoder. In this case, the group H has only one kind of representation. So, we have the unique representative. Hence, the decoder is given as the partial trace with respect to the representation space of f_H. When the quantum channel of the noise is given as the TP-CP map $\Lambda[\mathrm{P}]$ by using the probability distribution P on H, the entanglement fidelity of $D \circ \Lambda[\mathrm{P}]$ is 1.

Example 5.1 We consider the unitary representation $(\mathsf{W}^r_{\mathbb{F},\mathrm{H}} \rtimes \mathsf{S}^r_{\mathbb{F}})^{\otimes 2}$ of the group $\mathrm{H}(2r, \mathbb{F}_q) \rtimes \mathrm{Sp}(2r, \mathbb{F}_q)$ and focus on its irreducible representation space $\mathcal{H}_s$, where

q is a power of an odd prime. (For the detail definition of the representation, see [44, Subsection 8.3.3].)

Due to the irreducible decomposition given in [44, (8.41)], the encoder is $\mathbb{C}^{(q^r+1)/2}$. Then, we can perfectly correct the noise caused by any element of $\mathrm{H}(2r, \mathbb{F}_q)$, which is a normal subgroup of $\mathrm{H}(2r, \mathbb{F}_q) \rtimes \mathrm{Sp}(2r, \mathbb{F}_q)$. That is, when we prepare two quantum systems whose dimensions are the same as that of $\mathcal{H}^{\otimes r}$ and the noise operation in $\mathsf{W}^r_{\mathbb{F}}(\vec{s})$ acts as the same unitary on both systems, there is no noise in the space $\mathbb{C}^{(q^r+1)/2}$. Although this quantum error correction code assumes special noises, this code is very useful in such a special noise.

5.5 Application to Entanglement Distillation and General Channel

The previous sections assumes a Pauli channel. However, their discussion can be extended to a general channel. For this purpose, we deal with the application of quantum error correction to entanglement distillation [7, 92, 117]. Given an arbitrary quantum state ρ on the quantum system $\mathcal{H}_B \otimes \mathcal{H}_A := (\mathcal{H}^{\otimes r})^{\otimes 2}$, using a subgroup N of $\mathbb{X}^{2r}$ and the set $\mathcal{J} := \{\vec{s}_x\}_{x \in N^*}$ of representatives of elements of $N^* = \mathbb{X}^{2r}/N^{\perp}$, we give a concrete protocol for **entanglement distillation** by the LOCC operation $\Lambda_{dis}[N, \mathcal{J}]$ defined as

$$\begin{aligned}
&\Lambda_{dis}[N, \mathcal{J}](\rho) \\
:= &\sum_{x,x' \in N^*} (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x-x'}) \otimes I)^\dagger (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'}) \otimes \overline{\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'})})^\dagger (P_x \otimes P^T_{x'}) \\
&\qquad \cdot \rho (P_x \otimes \overline{P_{x'}}) (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'}) \otimes \overline{\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'})}) (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x-x'}) \otimes I) \\
= &\sum_{x,x' \in N^*} (P_0 \otimes P^T_0) (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'}) \otimes \overline{\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'})})^\dagger (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x-x'}) \otimes I)^\dagger \\
&\qquad \cdot \rho (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x-x'}) \otimes I) (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'}) \otimes \overline{\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'})}) (P_0 \otimes \overline{P_0}),
\end{aligned} \tag{5.39}$$

where (5.39) follows from (5.26). Since the TP-CP map $\Lambda_{dis}[N, \mathcal{J}]$ can be realized by the applications of unitaries dependently of the outcomes of the local measurements $\{P_x\}_x$ and $\{P^T_{x'}\}_{x'}$, it is an LOCC operation. This protocol generates a state close to maximally entangled state $\frac{1}{\sqrt{\dim \mathcal{H}_0}} |P_0\rangle\rangle$ on the bipartite system $\mathcal{H}_0^{\otimes 2}$. Then, since the range of P_0 is $\mathcal{H}_0$, the final state is a state on $\mathcal{H}_0^{\otimes 2}$. Since (5.26) yields $\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'}) P_0 \mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'})^\dagger = P_{x'}$, the maximally entangled state $\frac{1}{\sqrt{\dim \mathcal{H}_0}} |P_0\rangle\rangle$ on the bipartite system $\mathcal{H}_0^{\otimes 2}$ satisfies

$$
\begin{aligned}
&\frac{1}{\dim \mathcal{H}_0} \sum_{x' \in N^*} (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'}) \otimes \overline{\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'})})(P_0 \otimes \overline{P_0})|P_0\rangle\rangle\langle\langle P_0| \\
&\qquad\qquad\qquad\qquad \cdot (P_0 \otimes P_0^T)(\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'}) \otimes \overline{\mathsf{W}^r_{\mathbb{X}}(\vec{s}_{x'})})^\dagger \\
=& \frac{1}{\dim \mathcal{H}_0} \sum_{x' \in N^*} |P_{x'}\rangle\rangle\langle\langle P_{x'}| \\
=& \frac{1}{\dim \mathcal{H}^{\otimes r}} \sum_{\vec{s} \in N} |\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rangle\rangle\langle\langle \mathsf{W}^r_{\mathbb{X}}(\vec{s})|,
\end{aligned}
$$

where the final equation follows from the equivalence relation between two conditions $\sum_{\vec{s} \in N} \omega_{\mathbb{X}}^{x'(\vec{s})} = 0$ and $x' \neq 0$, and the relation $|\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rangle\rangle = \sum_{x' \in N^*} \omega_{\mathbb{X}}^{x'(\vec{s})} |P_{x'}\rangle\rangle$ derived by [44, (8.22)].

Hence, by rewriting $x - x'$ x, the above equations yield

$$
\begin{aligned}
&\frac{1}{\dim \mathcal{H}_0} \Lambda_{dis}[N, \mathcal{J}]^*(|P_0\rangle\rangle\langle\langle P_0|) \\
=& \frac{1}{\dim \mathcal{H}^{\otimes r}} \sum_{x \in N^*} (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_x) \otimes I) \sum_{\vec{s} \in N} |\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rangle\rangle\langle\langle \mathsf{W}^r_{\mathbb{X}}(\vec{s})| (\mathsf{W}^r_{\mathbb{X}}(\vec{s}_x) \otimes I)^\dagger \\
=& \frac{1}{\dim \mathcal{H}^{\otimes r}} \sum_{x \in N^*} \sum_{\vec{s} \in N} |\mathsf{W}^r_{\mathbb{X}}(\vec{s} + \vec{s}_x)\rangle\rangle\langle\langle \mathsf{W}^r_{\mathbb{X}}(\vec{s} + \vec{s}_x)|.
\end{aligned}
$$

That is, the RHS is the projection to the space spanned by the pure states $|\mathsf{W}^r_{\mathbb{X}}(\vec{s} + \vec{s}_x)\rangle\rangle$ corresponding to the correctable errors. This projection is denoted by $P_{[N,\mathcal{J}]}$. Hence, the fidelity between $\Lambda_{dis}[N, \mathcal{J}](\rho)$ and the maximally entangled state $\frac{1}{\sqrt{\dim \mathcal{H}_0}} |P_0\rangle\rangle$ is calculated to

$$
\begin{aligned}
&\langle\langle P_0 | \Lambda_{dis}[N, \mathcal{J}](\rho) | P_0 \rangle\rangle = \operatorname{Tr} \rho \frac{1}{\dim \mathcal{H}_0} \Lambda_{dis}[N, \mathcal{J}]^*(|P_0\rangle\rangle\langle\langle P_0|) \\
=& \operatorname{Tr} \rho P_{[N,\mathcal{J}]} = \sum_{x \in N^*} \sum_{\vec{s} \in N} \frac{1}{\dim \mathcal{H}^{\otimes r}} \langle\langle \mathsf{W}^r_{\mathbb{X}}(\vec{s} + \vec{s}_x) | \rho | \mathsf{W}^r_{\mathbb{X}}(\vec{s} + \vec{s}_x) \rangle\rangle.
\end{aligned}
$$

We prepare the initial state on the quantum system $\mathcal{H}_C \otimes \mathcal{H}_A := (\mathcal{H}^{\otimes r})^{\otimes 2}$ to be the maximally entangled state $\frac{1}{\sqrt{\dim \mathcal{H}^{\otimes r}}} |I\rangle\rangle$, and input the system $\mathcal{H}_C$ into the channel Λ whose output system is $\mathcal{H}_B$. In the following, we calculate the fidelity between the maximally entangled state $\frac{1}{\sqrt{\dim \mathcal{H}_0}} |P_0\rangle\rangle$ and the final state of the application of the LOCC protocol $\Lambda_{dis}[N, \mathcal{J}]$ to the above output state.

$$
\begin{aligned}
&\frac{1}{\dim \mathcal{H}^{\otimes r} \dim \mathcal{H}_0} \langle\langle P_0 | \Lambda_{dis}[N, \mathcal{J}] \circ (\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|) | P_0 \rangle\rangle \\
=& \operatorname{Tr} \frac{1}{\dim \mathcal{H}^{\otimes r}} (\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|) P_{[N,\mathcal{J}]}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{x \in N^*} \sum_{\vec{s} \in N} \frac{1}{\dim \mathcal{H}_A^2} \langle\langle \mathsf{W}_{\mathbb{X}}^r(\vec{s} + \vec{s}_x) | (\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|) | \mathsf{W}_{\mathbb{X}}^r(\vec{s} + \vec{s}_x)\rangle\rangle \\
&= P[\Lambda](\mathcal{J} + N) \qquad (5.40) \\
&= \sum_{x \in N^*} \sum_{\vec{s} \in N} \frac{1}{\dim \mathcal{H}_A^2} \langle\langle \mathsf{W}_{\mathbb{X}}^r(\vec{s} + \vec{s}_x) | (\overline{\Lambda}_{\mathbb{X}^{2r}} \otimes id)(|I\rangle\rangle\langle\langle I|) | \mathsf{W}_{\mathbb{X}}^r(\vec{s} + \vec{s}_x)\rangle\rangle.
\end{aligned}
$$

Here, in addition to the above discussion, we employ the definition of $P[\Lambda](\mathcal{J} + N)$ mentioned in Example 4.22 and the properties of the twirled channel $\overline{\Lambda}_{\mathbb{X}^{2r}}$.

Now, we set the initial state to the maximally entangled state $\frac{1}{\sqrt{\dim \mathcal{H}^{\otimes r}}} |I\rangle\rangle$, apply the quantum operation $(\Lambda \otimes id)$, and perform the measurement $\{P_{x'}\}_{x' \in N^*}$ on the system $\mathcal{H}_A$. This operation has the same final state as the same final state when we prepare the initial state $|P_{x'}\rangle\rangle$ with equal probability $\frac{1}{|N|}$ and apply the quantum operation $(\Lambda \otimes id)$. That is, we have

$$
\begin{aligned}
&\frac{1}{\dim \mathcal{H}^{\otimes r}} \sum_{x \in N^*} (I \otimes P_x^T)(\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)(I \otimes P_x^T) \\
&= \frac{1}{\dim \mathcal{H}_0} \frac{1}{|N|} \sum_{x \in N^*} (\Lambda \otimes id)(|P_x\rangle\rangle\langle\langle P_x|). \qquad (5.41)
\end{aligned}
$$

Further,

$$
\Lambda_{dis}[N, \mathcal{J}]\Big(\sum_{x \in N^*} (I \otimes P_x^T) \rho (I \otimes P_x^T)\Big) = \Lambda_{dis}[N, \mathcal{J}](\rho). \qquad (5.42)
$$

Hence, we have

$$
\begin{aligned}
&\sum_{x' \in N} \frac{1}{|N|} \frac{1}{\dim \mathcal{H}_0^2} \langle\langle P_0 | \Lambda_{dis}[N, \mathcal{J}] \circ (\Lambda \otimes id)(|P_{x'}\rangle\rangle\langle\langle P_{x'}|) | P_0\rangle\rangle \\
&= \frac{1}{\dim \mathcal{H}^{\otimes r} \dim \mathcal{H}_0} \langle\langle P_0 | \Lambda_{dis}[N, \mathcal{J}] \circ (\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|) | P_0\rangle\rangle \\
&= P[\Lambda](\mathcal{J} + N), \qquad (5.43)
\end{aligned}
$$

where the first equation follows from (5.41) and (5.42), and the second equation follows from (5.40). The definition of $\Lambda_{dis}[N, \mathcal{J}]$ implies that

$$
\begin{aligned}
&\sum_{x' \in N} \frac{1}{|N|} \frac{1}{\dim \mathcal{H}_0^2} \langle\langle P_0 | \Lambda_{dis}[N, \mathcal{J}] \circ (\Lambda \otimes id)(|P_{x'}\rangle\rangle\langle\langle P_{x'}|) | P_0\rangle\rangle \\
&= \sum_{x' \in N} \frac{1}{|N|} \frac{1}{\dim \mathcal{H}_0^2} \langle\langle P_{x'} | ((D \circ \Lambda) \otimes id)(|P_{x'}\rangle\rangle\langle\langle P_{x'}|) | P_{x'}\rangle\rangle \\
&= \sum_{x' \in N} \frac{1}{|N|} F_e^2(D \circ \Lambda|_{\mathcal{H}_{x'}}). \qquad (5.44)
\end{aligned}
$$

Summarizing the relations (5.43) and (5.44), we have the following theorem.

Theorem 5.7 *For an arbitrary quantum channel Λ, we denote its twirling with respect to the representation $\mathsf{W}^r_{\mathbb{X}}$ by $\Lambda[\mathsf{W}_{\mathbb{X}}, P]$. When we choose the encoder $\mathcal{H}_x$ subject to the distribution of the uniform distribution on N^*, the average of the entanglement fidelity of the application of the quantum error correction $\Lambda[N, \mathcal{J}]$ to the channel Λ is $P(\mathcal{J} + N)$.*

Similar to the above discussion, we can calculate the average of fidelity between the resultant state and the maximally entangled state $\frac{1}{\sqrt{\dim \mathcal{H}_{x'}}} |\mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_{x'}}([\vec{s}])\rangle\rangle$ determined by $[\vec{s}] \in N^{\perp}/N$ as follows.

$$\begin{aligned}
&\sum_{x' \in N} \frac{1}{|N|} \frac{1}{\dim \mathcal{H}_0^2} \langle\langle \mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_{x'}}([\vec{s}])|(D \circ \Lambda) \otimes id)(|P_{x'}\rangle\rangle\langle\langle P_{x'}|)|\mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_{x'}}([\vec{s}])\rangle\rangle \\
=&\sum_{x' \in N} \frac{1}{|N|} \frac{1}{\dim \mathcal{H}_0^2} \langle\langle \mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_0}([\vec{s}])|\Lambda_{dis}[N, \mathcal{J}] \circ (\Lambda \otimes id)(|P_{x'}\rangle\rangle\langle\langle P_{x'}|) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \cdot |\mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_0}([\vec{s}])\rangle\rangle \\
=&\frac{1}{\dim \mathcal{H}^{\otimes r} \dim \mathcal{H}_0} \langle\langle \mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_0}([\vec{s}])|\Lambda_{dis}[N, \mathcal{J}] \circ (\Lambda \otimes id)(|I\rangle\rangle\langle\langle I|)|\mathsf{W}^r_{\mathbb{X}}|_{\mathcal{H}_0}([\vec{s}])\rangle\rangle \\
=&P[\Lambda](\mathcal{J} + N + \vec{s}) = P[\Lambda]\{\mathcal{J} + N\}([\vec{s}]),
\end{aligned} \tag{5.45}$$

where the first, second and third equations follow from the similar discussions as (5.44), the combination of (5.42) and (5.41), and (5.40), respectively.

5.6 Quantum Secure Communication

5.6.1 Case Without Privacy Amplification

In quantum communication, once we succeed in keeping the perfect coherence of the state during the transmission, we can keep the perfect secrecy of the transmitted information because the information is described in only a single particle [111]. In order that an eavesdropper obtain an information, the eavesdropper needs to make an interaction with the transmitted particle, However, once we succeed in keeping the perfect coherence, we succeed the single particle transmission without any interaction with the third party so that the perfect secrecy is realized. In the following, we address the quantitative framework for the secrecy in the respective communication. In general, it is difficult to estimate how much information the eavesdropper obtains from the information leaked to the environment system. Hence, we estimate the secrecy providing the eavesdropper obtains all information leaked to the environment system via the given channel Λ. Thus, we estimate the leaked information to the eavesdropper by evaluating the correlation between the input and out systems

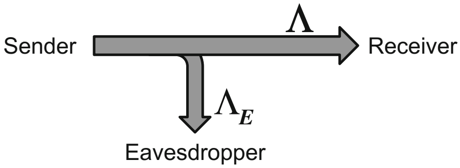


Fig. 5.6 Two quantum channels Λ and Λ_E

of the channel Λ_E to the environment system as Fig. 5.6. The first criterion is the transmission information. When the state $|x\rangle$ is generated on the input system with probability $P(x)$, the transmission information is defined as

$$\begin{aligned} I(P:\Lambda_E) :=& H(\Lambda_E(\sum_x P(x)|x\rangle\langle x|)) - \sum_x P(x)H(\Lambda_E(|x\rangle\langle x|)) \\ =& D(\rho_{A,E}\|\rho_A\otimes\rho_E), \end{aligned}$$

where $\rho_{A,E}$, ρ_A, ρ_E is defined as

$$\rho_{A,E} := \sum_x P(x)|x\rangle\langle x|\otimes\Lambda_E(|x\rangle\langle x|), \quad \rho_A := \mathrm{Tr}_E\,\rho_{A,E}, \quad \rho_E := \mathrm{Tr}_A\,\rho_{A,E}.$$

As another criterion, we often adopt the trace norm between the real state $\rho_{A,E}$ and the ideally non-correlated state $\rho_A\otimes\rho_E$ on the composite system as

$$\begin{aligned} d_1(P:\Lambda_E) :=& \|\rho_{A,E}-\rho_A\otimes\rho_E\|_1 \\ =& \sum_x P(x)\Big\|\Lambda_E(|x\rangle\langle x|) - \Lambda_E(\sum_x P(x)|x\rangle\langle x|)\Big\|_1. \end{aligned} \tag{5.46}$$

Considering the maximum value of the above, we sometimes employ the following criterion

$$d_{1,\max}(P:\Lambda_E) := \max_{x,x':P(x)>0}\|\Lambda_E(|x\rangle\langle x|)-\Lambda_E(|x'\rangle\langle x'|)\|_1.$$

The security based on the criterion $d_{1,\max}(P:\Lambda_E)$ is called the semantic security [3].

Due to (2.8), the channel $\Lambda[\mathsf{W}^r_{\mathbb{X}},\mathrm{P}^{\mathbb{X}^{2r}}]_E$ to the environment system of the Pauli channel $\Lambda[\mathsf{W}^r_{\mathbb{X}},\mathrm{P}^{\mathbb{X}^{2r}}]$ is given by using the basis $|\vec{s}\rangle$ of the environment system as

$$\Lambda[\mathsf{W}^r_{\mathbb{X}},\mathrm{P}^{\mathbb{X}^{2r}}]_E(\rho) = \sum_{\vec{s},\vec{s}'}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\vec{s})}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\vec{s}')}[\mathrm{Tr}\,\mathsf{W}^r_{\mathbb{X}}(\vec{s})\rho\mathsf{W}^r_{\mathbb{X}}(-\vec{s}')]|\vec{s}\rangle\langle\vec{s}'|.$$

When the input state is the computational basis $|\boldsymbol{x}\rangle$, we have $\mathsf{W}^r_{\mathbb{X}}(\vec{s})|\boldsymbol{x}\rangle = \tau_{\mathbb{X}}^{\boldsymbol{s}\cdot\boldsymbol{t}} \omega_{\mathbb{X}}^{\boldsymbol{t}\cdot\boldsymbol{x}}|\boldsymbol{x}+\boldsymbol{s}\rangle$ for $\vec{s}=(\boldsymbol{s},\boldsymbol{t})$. Hence, the output state of the channel $\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]_E$ to the environment system is given as

$$
\begin{aligned}
&\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]_E(|\boldsymbol{x}\rangle\langle\boldsymbol{x}|)\\
=&\sum_{\vec{s},\vec{s}'}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\vec{s})}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\vec{s}')}\langle\boldsymbol{x}|\mathsf{W}^r_{\mathbb{X}}(-\vec{s}')\mathsf{W}^r_{\mathbb{X}}(\vec{s})|\boldsymbol{x}\rangle|\vec{s}\rangle\langle\vec{s}'|\\
=&\sum_{\boldsymbol{s},\boldsymbol{t},\boldsymbol{s}',\boldsymbol{t}'}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\boldsymbol{s},\boldsymbol{t})}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\boldsymbol{s}',\boldsymbol{t}')}\tau_{\mathbb{X}}^{\boldsymbol{s}\cdot\boldsymbol{t}-\boldsymbol{s}'\cdot\boldsymbol{t}'}\omega_{\mathbb{X}}^{(\boldsymbol{t}-\boldsymbol{t}')\cdot\boldsymbol{x}}\\
&\qquad\qquad\qquad\cdot\langle\boldsymbol{x}+\boldsymbol{s}'|\boldsymbol{x}+\boldsymbol{s}\rangle|(\boldsymbol{s},\boldsymbol{t})\rangle\langle(\boldsymbol{s}',\boldsymbol{t}')|\\
=&\sum_{\boldsymbol{s},\boldsymbol{t},\boldsymbol{t}'}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\boldsymbol{s},\boldsymbol{t})}\sqrt{\mathrm{P}^{\mathbb{X}^{2r}}(\boldsymbol{s},\boldsymbol{t}')}\tau_{\mathbb{X}}^{(\boldsymbol{s}+2\boldsymbol{x})\cdot(\boldsymbol{t}-\boldsymbol{t}')}|(\boldsymbol{s},\boldsymbol{t})\rangle\langle(\boldsymbol{s},\boldsymbol{t}')|\\
=&\sum_{\boldsymbol{s}}\mathrm{P}^{\mathbb{X}^r_1}(\boldsymbol{s})\sum_{\boldsymbol{t},\boldsymbol{t}'}\sqrt{\mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1}(\boldsymbol{t}|\boldsymbol{s})}\sqrt{\mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1}(\boldsymbol{t}'|\boldsymbol{s})}\tau_{\mathbb{X}}^{(\boldsymbol{s}+2\boldsymbol{x})\cdot(\boldsymbol{t}-\boldsymbol{t}')}|(\boldsymbol{s},\boldsymbol{t})\rangle\langle(\boldsymbol{s},\boldsymbol{t}')|\\
=&\sum_{\boldsymbol{s}}\mathrm{P}^{\mathbb{X}^r_1}(\boldsymbol{s})\rho_{E,\boldsymbol{s}}(\boldsymbol{x}),
\end{aligned}
$$

where $\rho_{E,\boldsymbol{s}}(\boldsymbol{x}) := \sum_{\boldsymbol{t},\boldsymbol{t}'}\sqrt{\mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1}(\boldsymbol{t}|\boldsymbol{s})}\sqrt{\mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1}(\boldsymbol{t}'|\boldsymbol{s})}\tau_{\mathbb{X}}^{(\boldsymbol{s}+2\boldsymbol{x})\cdot(\boldsymbol{t}-\boldsymbol{t}')}|(\boldsymbol{s},\boldsymbol{t})\rangle\langle(\boldsymbol{s},\boldsymbol{t}')|$.

Lemma 5.9 ([53, 98]) *The inequality*

$$
I(P:\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]_E) \le H(\mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1}) \le H(\mathrm{P}^{\mathbb{X}^r_2}) \tag{5.47}
$$

holds and the equality in the first inequality holds when P is the uniform distribution P_{mix}. Hence, the relation (2.43) implies that

$$
I(P:\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]_E) \le (1-\mathrm{P}^{\mathbb{X}^r_2}(0))\log|\mathbb{X}^r| + h(\mathrm{P}^{\mathbb{X}^r_2}(0)). \tag{5.48}
$$

Proof When P is the uniform distribution P_{mix}, the state $\rho_{E,\boldsymbol{s}}(\boldsymbol{x})$ is a pure state and we have $\sum_{\boldsymbol{x}} P(\boldsymbol{x})\rho_{E,\boldsymbol{s}}(\boldsymbol{x}) = \sum_{\boldsymbol{t}} \mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1}(\boldsymbol{t}|\boldsymbol{s})|(\boldsymbol{s},\boldsymbol{t})\rangle\langle(\boldsymbol{s},\boldsymbol{t})|$, which implies

$$
\begin{aligned}
&I(P:\Lambda_E)\\
=&\sum_{\boldsymbol{s}}\mathrm{P}^{\mathbb{X}^r_1}(\boldsymbol{s})\Big(-\log(\mathrm{P}^{\mathbb{X}^r_1}(\boldsymbol{s}))+H\big(\sum_{\boldsymbol{x}}P(\boldsymbol{x})\rho_{E,\boldsymbol{s}}(\boldsymbol{x})\big)\\
&\qquad\qquad\qquad+\log(\mathrm{P}^{\mathbb{X}^r_1}(\boldsymbol{s}))-H(\rho_{E,\boldsymbol{s}}(\boldsymbol{x}))\Big)\\
=&\sum_{\boldsymbol{s}}\mathrm{P}^{\mathbb{X}^r_1}(\boldsymbol{s})H(\sum_{\boldsymbol{x}}P(\boldsymbol{x})\rho_{E,\boldsymbol{s}}(\boldsymbol{x}))=\sum_{\boldsymbol{s}}\mathrm{P}^{\mathbb{X}^r_1}(\boldsymbol{s})H(\mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1=\boldsymbol{s}})=H(\mathrm{P}^{\mathbb{X}^r_2|\mathbb{X}^r_1}).
\end{aligned}
$$

Even when P is not the uniform distribution, taking the average with respect to the shifted distribution $P_{\boldsymbol{s}}(\boldsymbol{x}) := P(\boldsymbol{x}+\boldsymbol{s})$, we find that the distribution $\sum_{\boldsymbol{s}} \frac{1}{|\mathbb{X}^r|} P_{\boldsymbol{s}}$ is the uniform distribution P_{mix}. Since $I(P : \Lambda_E) = I(P_{\boldsymbol{s}} : \Lambda_E)$, the concavity of mutual information given in (2.39) and the concavity of the entropy yield

$$I(P : \Lambda_E) \le I(P_{\mathrm{mix}} : \Lambda_E) = H(\mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}) \le H(\mathrm{P}^{\mathbb{X}_2^r}),$$

which implies the desired argument. ■

Lemma 5.10 ([56]) *The inequalities*

$$d_1(P_{\mathrm{mix}} : \Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]_E) \le 3\sqrt{1-\mathrm{P}^{\mathbb{X}_2^r}(0)} \tag{5.49}$$

$$d_{1,\max}(P : \Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]_E) \le 4\sqrt{1-\mathrm{P}^{\mathbb{X}_2^r}(0)} \tag{5.50}$$

hold.

Proof Since the matrix composed of the diagonal components of $\rho_{E,\boldsymbol{s}}(\boldsymbol{x})$ is $\sum_{\boldsymbol{t}} \mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\boldsymbol{t}|\boldsymbol{s})|(\boldsymbol{s},\boldsymbol{t})\rangle\langle(\boldsymbol{s},\boldsymbol{t})|$, applying (2.34) to the case with $E_i = \sum_{\boldsymbol{s}} |(\boldsymbol{s},0)\rangle\langle(\boldsymbol{s},0)|$, we obtain the following inequality

$$\|\sum_{\boldsymbol{t}} \mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(\boldsymbol{t}|\boldsymbol{s})|(\boldsymbol{s},\boldsymbol{t})\rangle\langle(\boldsymbol{s},\boldsymbol{t})| - \rho_{E,\boldsymbol{s}}(\boldsymbol{x})\|_1 \le 3\sqrt{1-\mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(0|\boldsymbol{s})}. \tag{5.51}$$

So, the concavity of $x \mapsto \sqrt{x}$ implies (5.49) with $P = P_{\mathrm{mix}}$.

Further, since $E_i\rho_{E,\boldsymbol{s}}(\boldsymbol{x})E_i = E_i\rho_{E,\boldsymbol{s}}(\boldsymbol{x}')E_i$, the application of (2.33) yields

$$\begin{aligned}
&\|\rho_{E,\boldsymbol{s}}(\boldsymbol{x}) - \rho_{E,\boldsymbol{s}}(\boldsymbol{x}')\|_1 \\
\le&\|\rho_{E,\boldsymbol{s}}(\boldsymbol{x}) - E_i\rho_{E,\boldsymbol{s}}(\boldsymbol{x})E_i\|_1 + \|E_i\rho_{E,\boldsymbol{s}}(\boldsymbol{x}')E_i - \rho_{E,\boldsymbol{s}}(\boldsymbol{x}')\|_1 \\
\le&4\sqrt{1-\mathrm{P}^{\mathbb{X}_2^r|\mathbb{X}_1^r}(0|\boldsymbol{s})}.
\end{aligned}$$

Similarly, the concavity of $x \mapsto \sqrt{x}$ implies (5.50). ■

In this way, we can evaluate the amount of information leaked to the environment system by using the error probability $1 - P^{\mathbb{X}_2^r|\mathbb{X}_1^r}(0)$ in the dual computational basis for the case of Pauli channel. This fact can be extended to a general channel as follows.

Theorem 5.8 *Given a general channel Λ, we denote its twirling with respect to the discrete Heisenberg representation by $\Lambda[\mathsf{W}^r_{\mathbb{X}}, \mathrm{P}^{\mathbb{X}^{2r}}]$. Then, we have*

$$I(P_{\mathrm{mix}} : \Lambda_E) \le H(\mathrm{P}^{\mathbb{X}_2^r}) \le (1-\mathrm{P}^{\mathbb{X}_2^r}(0))\log|\mathbb{X}^r| + h(\mathrm{P}^{\mathbb{X}_2^r}(0)) \tag{5.52}$$

$$d_1(P_{\mathrm{mix}} : \Lambda_E) \le 3\sqrt{1 - \mathrm{P}^{\mathbb{X}_2^r}(0)}. \tag{5.53}$$

When the information to be sent is subject to the uniform distribution as the above, once the criterion is limited to $I(P_{\mathrm{mix}} : \Lambda_E)$ or $d_1(P_{\mathrm{mix}} : \Lambda_E)$, the security evaluation similar to Pauli channel is available even for a general channel. However, when the criterion $d_{1,\max}(P_{\mathrm{mix}} : \Lambda_E)$ is employed, the same security evaluation as Pauli channel is not necessarily available for a general channel.

Proof To prove the above inequality, we consider the information processing, in which, the sender generates the secret random number $\boldsymbol{x}$ subject to the uniform distribution P_{mix}, and send it to the receiver via the quantum channel Λ by the following protocol. The sender generates another random number $\boldsymbol{s}$ subject to the uniform distribution P_{mix} independently of $\boldsymbol{x}$, and sends the state $|\boldsymbol{x}+\boldsymbol{s}\rangle$ via the quantum channel Λ. The eavesdropper is assumed to be able to access the classical information $|\boldsymbol{s}\rangle\langle\boldsymbol{s}|$. Then, the receiver operates the unitary $\mathsf{W}_{\mathbb{X}}(-\boldsymbol{s},0)$ on the received system. Hence, the resultant state on the receiver's system is $\mathsf{W}_{\mathbb{X}}(-\boldsymbol{s},0)\Lambda(|\boldsymbol{x}+\boldsymbol{s}\rangle\langle\boldsymbol{x}+\boldsymbol{s}|)\mathsf{W}_{\mathbb{X}}(\boldsymbol{s},0)$. When this information processing is performed, the mutual information between the secret random number x and the information $\Lambda_E(|\boldsymbol{x}+\boldsymbol{s}\rangle\langle\boldsymbol{x}+\boldsymbol{s}|)\otimes|\boldsymbol{s}\rangle\langle\boldsymbol{s}|$ of the eavesdropper is calculated by using $\Lambda_E^{\boldsymbol{s}}(|\boldsymbol{x}\rangle\langle\boldsymbol{x}|) := \Lambda_E(|\boldsymbol{x}+\boldsymbol{s}\rangle\langle\boldsymbol{x}+\boldsymbol{s}|)$ as follows.

$$\begin{aligned}
&\sum_{\boldsymbol{x}} \frac{1}{|\mathbb{X}^r|} D\Bigg(\sum_{\boldsymbol{s}} \frac{1}{|\mathbb{X}^r|}\Lambda_E(|\boldsymbol{x}+\boldsymbol{s}\rangle\langle\boldsymbol{x}+\boldsymbol{s}|)\otimes|\boldsymbol{s}\rangle\langle\boldsymbol{s}| \Bigg\| \\
&\qquad\qquad \sum_{\boldsymbol{x}',\boldsymbol{s}'} \frac{1}{|\mathbb{X}^r|^2}\Lambda_E(|\boldsymbol{x}'+\boldsymbol{s}'\rangle\langle\boldsymbol{x}'+\boldsymbol{s}'|)\otimes|\boldsymbol{s}'\rangle\langle\boldsymbol{s}'|\Bigg) \\
=&\sum_{\boldsymbol{s}} \frac{1}{|\mathbb{X}^r|} I(P_{\mathrm{mix}} : \Lambda_E^{\boldsymbol{s}}) = I(P_{\mathrm{mix}} : \Lambda_E^{\boldsymbol{s}}).
\end{aligned} \tag{5.54}$$

Next, we consider the quantum channel with the input $|\boldsymbol{x}\rangle$ and the output

$$\begin{aligned}
&\Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}](|\boldsymbol{x}\rangle\langle\boldsymbol{x}|) \\
=&\sum_{\boldsymbol{s},\boldsymbol{t}} \frac{1}{|\mathbb{X}^r|^2}\mathsf{W}_{\mathbb{X}}(-\boldsymbol{s},-\boldsymbol{t})\Lambda(\mathsf{W}_{\mathbb{X}}(\boldsymbol{s},\boldsymbol{t})|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\mathsf{W}_{\mathbb{X}}(-\boldsymbol{s},-\boldsymbol{t}))\mathsf{W}_{\mathbb{X}}(\boldsymbol{s},\boldsymbol{t}) \\
=&\sum_{\boldsymbol{s}} \frac{1}{|\mathbb{X}^r|}\mathsf{W}_{\mathbb{X}}(-\boldsymbol{s},0)\Lambda(|\boldsymbol{x}+\boldsymbol{s}\rangle\langle\boldsymbol{x}+\boldsymbol{s}|)\mathsf{W}_{\mathbb{X}}(\boldsymbol{s},0).
\end{aligned}$$

Then, the above quantum channel has a larger environment system than that of $\Lambda_E(|\boldsymbol{x}+\boldsymbol{s}\rangle\langle\boldsymbol{x}+\boldsymbol{s}|)\otimes|\boldsymbol{s}\rangle\langle\boldsymbol{s}|$. Hence, the information processing inequality for the relative entropy yields

$$
\begin{aligned}
& I(P_{\mathrm{mix}} : \Lambda_E^s) \\
=& \sum_{x} \frac{1}{|\mathbb{X}^r|} D\Big(\sum_{s} \frac{1}{|\mathbb{X}^r|} \Lambda_E(|\boldsymbol{x}+\boldsymbol{s}\rangle\langle \boldsymbol{x}+\boldsymbol{s}|) \otimes |\boldsymbol{s}\rangle\langle \boldsymbol{s}| \Big\| \\
& \qquad \sum_{x',s'} \frac{1}{|\mathbb{X}^r|^2} \Lambda_E(|\boldsymbol{x}'+\boldsymbol{s}'\rangle\langle \boldsymbol{x}'+\boldsymbol{s}'|) \otimes |\boldsymbol{s}'\rangle\langle \boldsymbol{s}'| \Big) \\
\le& \sum_{x} \frac{1}{|\mathbb{X}^r|} D(\Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]_E(|\boldsymbol{x}\rangle\langle \boldsymbol{x}|) \| \sum_{x'} \frac{1}{|\mathbb{X}^r|} D(\Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]_E(|\boldsymbol{x}'\rangle\langle \boldsymbol{x}'|)) \\
=& I(P_{\mathrm{mix}} : \Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]_E).
\end{aligned} \tag{5.55}
$$

Hence, (5.47) and (5.48) imply (5.52).

Since a relation similar to (5.54) holds for the trace norm, the combination of the information processing inequality for the trace norm and (5.49) yields (5.53) in a similar way. ■

5.6.2 Case with Privacy Amplification

When a given quantum channel Λ has sufficiently small errors in the computational basis and the dual computational basis, the information transmission has small information leakage. Conversely, when these error probabilities are large, the information transmission via such a channel has less reliability. However, when we apply quantum error correction so that the errors in both bases are corrected, the information transmission has small information leakage. That is, the amount of information leakage can be quantitatively evaluated by using the decoding error probability in the dual computational basis for the quantum channel obtained via quantum error correction.

When the quantum communication channel is given as a Pauli channel $\Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]$, we suppose that the sender applies the encoder based on the CSS code with a code pair C_1 and C_2 satisfying the torsion condition, and the receiver applies the decoder D based on the representative $(\boldsymbol{s}_x, \boldsymbol{t}_{x,y})$ of $(x, y) \in \mathbb{X}^r/C_2^\perp \times \mathbb{X}^r/C_1^\perp = (C_1 \times C_2)^*$. Given $([\boldsymbol{x}], [\boldsymbol{y}]) \in \mathbb{X}^r/C_2^\perp \times \mathbb{X}^r/C_1^\perp = (C_1 \times C_2)^*$ and $[\boldsymbol{s}] \in C_2^\perp/C_1$, we define

$$
\begin{aligned}
\mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}([\boldsymbol{s}]) &:= \Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]_E(|[\boldsymbol{s}], [\boldsymbol{x}], [\boldsymbol{y}]\rangle\langle[\boldsymbol{s}], [\boldsymbol{x}], [\boldsymbol{y}]|) \\
\hat{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}([\boldsymbol{s}]) &:= (D \circ \Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}])_E(|[\boldsymbol{s}], [\boldsymbol{x}], [\boldsymbol{y}]\rangle\langle[\boldsymbol{s}], [\boldsymbol{x}], [\boldsymbol{y}]|).
\end{aligned}
$$

When the sender sends the information by using the space $\mathcal{H}_{[\boldsymbol{x}],[\boldsymbol{y}]}$, the mutual information with the eavesdropper is given as $I(P, \mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]})$. The environment system of the channel $\Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]$ is a part of the environment system of the channel $D \circ \Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]$. That is, the latter environment system is given as the combination of the former environment system and the environment system $\mathcal{E}_D$ of the decoder D. Information processing inequality for the partial trace on $\mathcal{E}_D$ yields

$$I(P, \mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \le I(P, \hat{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \tag{5.56}$$

$$\le H(\mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}^{C_1^{\perp}/C_2}) \tag{5.57}$$

$$\le (1-\mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J}+N\}^{C_1^{\perp}/C_2}(0)) \log \dim \mathcal{H}_{[\boldsymbol{x}],[\boldsymbol{y}]} + h(\mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J}+N\}^{C_1^{\perp}/C_2}(0)), \tag{5.58}$$

with $N = C_1 \times C_2$ and $\mathcal{J} = \{(\boldsymbol{s}_x, \boldsymbol{t}_{x,y})\}_{(x,y)\in N^*}$, where (5.57) follows from Lemma 5.9.

The above information processing inequality and Lemma 5.10 yields

$$d_1(P_{\mathrm{mix}}, \mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \le d_1(P_{\mathrm{mix}}, \hat{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \le 3\sqrt{1 - \mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}^{C_1^{\perp}/C_2}(0)} \tag{5.59}$$

$$d_{1,\max}(P, \mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \le d_{1,\max}(P, \hat{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \le 4\sqrt{1 - \mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}^{C_1^{\perp}/C_2}(0)}. \tag{5.60}$$

Hence, when we choose a quantum error correcting code so that the probability $1 - \mathrm{P}^{\mathbb{X}^{2r}}\{\mathcal{J} + N\}^{C_1^{\perp}/C_2}(0)$ is sufficiently small, our communication channel has sufficient secrecy.

To realize the quantum error correcting code, we need to perform a quantum operation, which requires larger cost. As discussed later, when the communication channel is given as a Pauli channel $\Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]$, applying the error correcting code with privacy amplification based on the code pair $C_1 \subset C_2^{\perp}$ in $\mathbb{X}^r$ given in Subsection 5.1.6, we can realize the same secrecy as we apply the quantum error correcting code based of the above CSS code [117]. Given $[\boldsymbol{x}] \in \mathbb{X}^r/C_2^{\perp}$ and $[\boldsymbol{s}] \in C_2^{\perp}/C_1$, we define

$$\mathsf{W}_{\mathrm{mix},\boldsymbol{x}}(\boldsymbol{s}) := \sum_{\boldsymbol{s}'\in[\boldsymbol{s}]} \frac{1}{|C_1|} \Lambda[\mathsf{W}_{\mathbb{X}}^r, \mathrm{P}^{\mathbb{X}^{2r}}]_E(|\boldsymbol{s}' + \boldsymbol{y}\rangle\langle \boldsymbol{s}' + \boldsymbol{y}|).$$

Since

$$\sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^{\perp}} \frac{1}{|C_2|} \mathsf{W}_{[0],[\boldsymbol{y}]}([\boldsymbol{s}]) = \mathsf{W}_{\mathrm{mix},[0]}(\boldsymbol{s}),$$

the joint convexity of the relative entropy and the trace norm yields [53, 56]

$$I(P, \mathsf{W}_{\mathrm{mix},[0]}) \le \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^{\perp}} \frac{1}{|C_2|} I(P, \mathsf{W}_{[0],[\boldsymbol{y}]})$$

$$d_1(P, \mathsf{W}_{\mathrm{mix},[0]}) \le \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^{\perp}} \frac{1}{|C_2|} d_1(P, \mathsf{W}_{[0],[\boldsymbol{y}]})$$

$$d_{1,\max}(P, \mathsf{W}_{\mathrm{mix},[0]}) \le \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^{\perp}} \frac{1}{|C_2|} d_{1,\max}(P, \mathsf{W}_{[0],[\boldsymbol{y}]}).$$

Hence, due to the relations (5.56), (5.58), (5.59), and (5.60), we can upper bound the RHSs of the above inequalities by the RHSs of (5.58), (5.59), and (5.60).

Similarly, given a general channel Λ, we define $\mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}([\boldsymbol{s}])$, $\hat{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}([\boldsymbol{s}])$, and $\mathsf{W}_{\mathrm{mix},[\boldsymbol{x}]}(\boldsymbol{s})$. Then, we have

$$
\begin{align*}
& \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_2|} I(P, \mathsf{W}_{\mathrm{mix},[\boldsymbol{x}]}) \\
\le & \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_1||C_2|} I(P, \mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \tag{5.61} \\
\le & \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_1||C_2|} I(P, \hat{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \tag{5.62} \\
\le & \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_1||C_2|} H(P[D\circ\Lambda|_{\mathcal{H}_{([\boldsymbol{x}],[\boldsymbol{y}])}}]^{C_1^\perp/C_2}) \tag{5.63} \\
\le & H\Big(\sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_1||C_2|} P[D\circ\Lambda|_{\mathcal{H}_{([\boldsymbol{x}],[\boldsymbol{y}])}}]^{C_1^\perp/C_2}\Big) \tag{5.64} \\
= & H(P[\Lambda]\{\mathcal{J}+N\}^{C_1^\perp/C_2}) \tag{5.65} \\
\le & (1 - P[\Lambda]\{\mathcal{J}+N\}^{C_1^\perp/C_2}(0)) \log\dim\mathcal{H}_0 + h(P[\Lambda]\{\mathcal{J}+N\}^{C_1^\perp/C_2}(0)), \tag{5.66}
\end{align*}
$$

where the inequalities (5.61), (5.62), (5.63), (5.64), (5.65), and (5.66) follow from the joint convexity of relative entropy, an information processing inequality similar to (5.56), (5.52), the concavity of von Neumann entropy, (5.45), and (2.43), respectively.

Similarly, for $d_1(P, \mathsf{W}_{\mathrm{mix},[\boldsymbol{x}]})$, we have

$$
\begin{align*}
& \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_2|} d_1(P_{\mathrm{mix}}, \mathsf{W}_{\mathrm{mix},[\boldsymbol{x}]}) \\
\le & \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_1||C_2|} d_1(P_{\mathrm{mix}}, \mathsf{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \tag{5.67} \\
\le & \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_1||C_2|} d_1(P_{\mathrm{mix}}, \hat{W}_{[\boldsymbol{x}],[\boldsymbol{y}]}) \tag{5.68} \\
\le & \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{3}{|C_1||C_2|} \sqrt{1 - P[D\circ\Lambda|_{\mathcal{H}_{([\boldsymbol{x}],[\boldsymbol{y}])}}]^{C_1^\perp/C_2}(0)} \tag{5.69} \\
\le & 3\sqrt{1 - \sum_{[\boldsymbol{y}]\in\mathbb{X}^r/C_1^\perp} \sum_{[\boldsymbol{x}]\in\mathbb{X}^r/C_2^\perp} \frac{1}{|C_1||C_2|} P[D\circ\Lambda|_{\mathcal{H}_{([\boldsymbol{x}],[\boldsymbol{y}])}}]^{C_1^\perp/C_2}(0)} \tag{5.70} \\
= & 3\sqrt{1 - P[\Lambda]\{\mathcal{J}+N\}^{C_1^\perp/C_2}(0)}, \tag{5.71}
\end{align*}
$$

where (5.67), (5.68), (5.69), (5.70), and (5.71) follow from the joint convexity of trace norm distance, an information processing inequality similar to (5.59), (5.53), concavity of $x \mapsto \sqrt{x}$, and (5.45), respectively.

Summarizing the above discussions, we obtain the following theorem by using

$$\begin{aligned} &\delta_{P[\Lambda]}[[C_1 \times C_2]] \\ :=&1 - \max_{\mathcal{J}_{2,x}=\{\boldsymbol{t}_{x,y}\}} \sum_{[\boldsymbol{x}]\in \mathbb{X}_1^r/C_2^\perp} P[\Lambda]^{\mathbb{X}_1^r/C_2^\perp}([\boldsymbol{x}]) P[\Lambda]^{\mathbb{X}_2^r|\mathbb{X}_1^r/C_2^\perp}(\mathcal{J}_{2,[\boldsymbol{x}]} + C_2|[\boldsymbol{x}]). \end{aligned} \tag{5.72}$$

Theorem 5.9 *When we choose the encoder $[\boldsymbol{x}] \in \mathbb{X}^r/C_2^\perp$ subject to the uniform distribution on $\mathbb{X}^r/C_2^\perp$ and apply the privacy amplification based on the subgroup $C_1 \subset C_2^\perp$, the amount of information leakage via a general quantum channel Λ can be evaluated as*

$$\begin{aligned} &\sum_{[\boldsymbol{x}]\in \mathbb{X}^r/C_2^\perp} \frac{1}{|C_2|} I(P, \mathsf{W}_{\mathrm{mix},[\boldsymbol{x}]}) \\ \le&\delta[[C_1 \times C_2]] \log \dim \mathcal{H}_0 + h(\delta[[C_1 \times C_2]]), \end{aligned} \tag{5.73}$$

$$\sum_{[\boldsymbol{x}]\in \mathbb{X}^r/C_2^\perp} \frac{1}{|C_2|} d_1(P_{\mathrm{mix}}, \mathsf{W}_{\mathrm{mix},[\boldsymbol{x}]}) \le 3\sqrt{\delta[[C_1 \times C_2]]}. \tag{5.74}$$

When the errors of the computational basis and the dual computational basis are independent in the distribution $P[\Lambda]$, the quantity $\delta_{P[\Lambda]}[[C_1 \times C_2]]$ can be simplified as

$$\delta_{P[\Lambda]}[[C_1 \times C_2]] = \delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^\perp/C_2] := 1 - \max_{\mathcal{J}_2=\{\boldsymbol{t}_y\}} P[\Lambda]^{\mathbb{X}_2^r}(\mathcal{J}_2 + C_2).$$

Generally, we have $\delta[[C_1 \times C_2]] \le \delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^\perp/C_2]$. When the marginal distribution $P[\Lambda]^{\mathbb{X}_2^r}$ is known and the correlations between errors of the computational basis and the dual computational basis is unknown, we employ the formulas obtained by the substitution of $\delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^\perp/C_2]$ into $\delta_{P[\Lambda]}[[C_1 \times C_2]]$ in (5.73) and (5.74).

Now, we notice that $(C_1 \times C_2)^*$ is isometric to $\mathbb{X}_2^r/C_2^\perp \times \mathbb{X}_2^r/C_1^\perp$. When we restrict $\vec{s}_{x,y}$ into elements $(\boldsymbol{s}_x, \boldsymbol{t}_y)$ for $(x, y) \in \mathbb{X}_2^r/C_2^\perp \times \mathbb{X}_2^r/C_1^\perp$, the definition (5.72) yields

$$\begin{aligned} &P[\Lambda]\{\{(\boldsymbol{s}_x, \boldsymbol{t}_y)\}_{(x,y)\in \mathbb{X}_2^r/C_2^\perp \times \mathbb{X}_2^r/C_1^\perp} + C_1 \times C_2\}_2(0) \\ =&P[\Lambda]^{\mathbb{X}_2^r}(\{\boldsymbol{t}_y\}_{y\in \mathbb{X}_2^r/C_1^\perp} + C_2), \end{aligned} \tag{5.75}$$

which implies the inequality

$$\delta_{P[\Lambda]}[[C_1 \times C_2]] \le 1 - \max_{\mathcal{J}_2=\{\boldsymbol{t}_y\}_{y\in \mathbb{X}_2^r/C_1^\perp}} P[\Lambda]^{\mathbb{X}_2^r}(\mathcal{J}_2 + C_2) = \delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^\perp/C_2]. \tag{5.76}$$

When the errors of the computational basis and the dual computational basis are independent in the distribution $P[\Lambda]$, even when representatives $\vec{s}_{(x,y)}$ have the general form $(\boldsymbol{s}_{x,y}, \boldsymbol{t}_{x,y})$, we have

$$\begin{aligned} & {}_{P[\Lambda]}\{\{(\boldsymbol{s}_{x,y}, \boldsymbol{t}_{x,y})\}_{(x,y)\in \mathbb{X}_2^r/C_2^\perp \times \mathbb{X}_2^r/C_1^\perp} + C_1 \times C_2\}_2(0) \\ = & \sum_{x\in \mathbb{X}_2^r/C_2^\perp} P[\Lambda]^{\mathbb{X}_1^r}(x) P[\Lambda]^{\mathbb{X}_2^r}(\{\boldsymbol{t}_{x,y}\}_{y\in \mathbb{X}_2^r/C_1^\perp} + C_2), \end{aligned} \tag{5.77}$$

which implies the equality in (5.76).

On the other hand, even when the distribution $P[\Lambda]^{\mathbb{X}_2^r}$ is known, when the errors of the computational basis and the dual computational basis have unknown correlation, we cannot evaluate $\delta_{P[\Lambda]}[[C_1 \times C_2]]$. However, since the relation (5.76) holds, we can employ the formulas obtained from the substitution of $\delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^\perp/C_2]$ into $\delta_{P[\Lambda]}[[C_1 \times C_2]]$ in (5.73) and (5.74).

In this section, we have discussed the secrecy evaluation related to quantum error correction. However, we have another method for the secrecy evaluation based on the universality of hash function [110].

5.7 Application to Quantum Cryptography (Quantum Key Distribution)

Due to the above discussion, when we apply a suitable pair of the classical error correction and the privacy amplification to a given quantum channel with known error distributions with the computational and dual computational bases, we cal realize the secure communication. However, it is difficult to guarantee that the entropy of the error in the dual computational basis is less than a certain level. Quantum key distribution guarantees the above required condition and brings us secure random number shared between distant two parties. The most typical protocol is the combination of the protocol proposed by Bennett and Brassard in 1984 [5] (**BB84 protocol**) with the classical error correction and the classical privacy amplification, which is called the modified BB84 protocol [95, 96, 117]. Although quantum cryptography expresses general cryptographic protocols with quantum system that are not limited to quantum key distribution in general, this book indicates only the quantum key distribution by the quantum cryptography because we do not deal with the quantum cryptography except for quantum key distribution. Firstly, we describe the modified BB84 protocol with classical error correction and the privacy amplification based on discrete Heisenberg representation $\mathsf{W}_{\mathbb{X}}^r$ as Fig. 5.7.

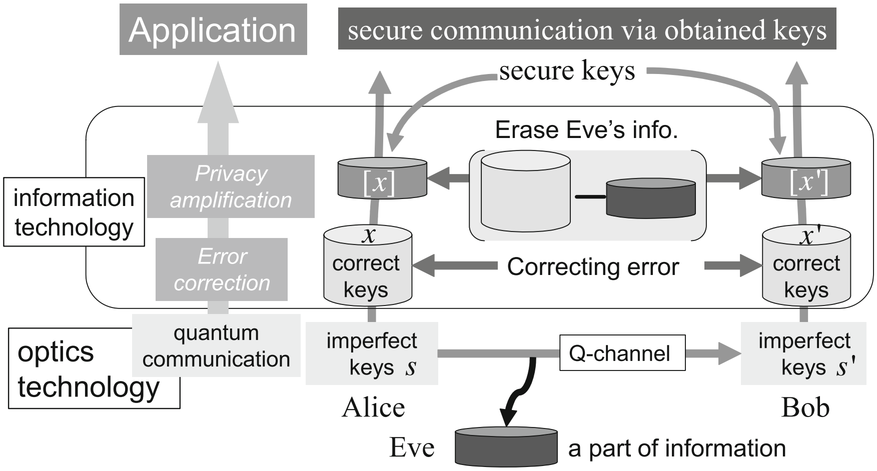


Fig. 5.7 Whole procedure of quantum key distribution

(1): Transmission Alice (Sender) selects the computational basis or the dual computational basis with probability $\frac{1}{2}$, chooses an element of $\mathbb{X}$ subject to the uniform distribution, sends it in the selected basis. She repeats this process so many times.

(2): Detection Bob (Receiver) selects the computational basis or the dual computational basis with probability $\frac{1}{2}$, measures the received system on the selected basis, and obtains the random variable in $\mathbb{X}$ as the outcome. He repeats this process so many times.

(3): Basis verification Alice and Bob share the information of their selected bases via the public channel. Then, they keep their random variables with agreed bases and discard their random variables with disagreed bases.

(4): Error estimation Alice assigns random variables with the ratio α among their random variables with agreed bases. Alice and Bob exchange the assigned random variables via the public channel, and estimate the probabilities $\mathrm{P}^{\mathbb{X}_1}(s)$ and $\mathrm{P}^{\mathbb{X}_2}(s)$ that the difference between their random variables is $s \in \mathbb{X}$ in the computational and dual computational bases among the remaining random variables.

(*) In the following, we apply the protocols for error correction and privacy amplification. Although Steps **(5)** and **(6)** are specified for the computational basis, we perform the same steps for the dual computational basis.

(5): Error correction Alice and Bob select a code $C_2^{\perp}$ as a subgroup of $\mathbb{X}^r$ to correct errors between Alice's random variables $\boldsymbol{s}$ and Bob's random variables $\boldsymbol{s}'$ subject to the distribution $\mathrm{P}^{\mathbb{X}_1}$. They also prepare the set of representatives $\{\boldsymbol{s}^2_{[\boldsymbol{s}]}\}_{[\boldsymbol{s}]\in\mathbb{X}^r/C_2^{\perp}}$ for the decoding of the classical code $C_2^{\perp}$. They prepare another set of representatives $\{\boldsymbol{s}^1_{[\boldsymbol{s}]}\}_{[\boldsymbol{s}]\in\mathbb{X}^r/C_2^{\perp}}$. They exchange the information $[\boldsymbol{s}]$ and $[\boldsymbol{s}']$ in $\mathbb{X}^r/C_2^{\perp}$ via the public channel. Alice obtains the element $\boldsymbol{x} := \boldsymbol{s} - \boldsymbol{s}^1_{[\boldsymbol{s}]}$ of C_2, and Bob obtains the element $\boldsymbol{x}' := \boldsymbol{s}' - \boldsymbol{s}^1_{[\boldsymbol{s}]} - \boldsymbol{s}^2_{[\boldsymbol{s}'-\boldsymbol{s}]}$ of C_2.

(6): Privacy amplification Alice and Bob select a subgroup C_1 of $C_2^{\perp}$ so that the errors of the dual computational basis subject to the probability distribution $\mathrm{P}^{\mathbb{X}_2}$ can be corrected by the code pair $C_1^{\perp}/C_2$. Alice and Bob set their final secure random numbers to be the quotient element $[\boldsymbol{x}]$ and $[\boldsymbol{x}']$ in $C_2^{\perp}/C_1$, respectively.

Here, the set of representatives $\{\boldsymbol{s}^2_{[\boldsymbol{s}]}\}_{[\boldsymbol{s}]\in\mathbb{X}^r/C_2^{\perp}}$ is chosen so that their calculation complexity is not so large and it realizes small decoding error probability because its purpose is error correction. On the other hand, the other set of representatives $\{\boldsymbol{s}^1_{[\boldsymbol{s}]}\}_{[\boldsymbol{s}]\in\mathbb{X}^r/C_2^{\perp}}$ is chosen without considering its decoding error probability because its purpose is not error correction but to adjust $\boldsymbol{s}-\boldsymbol{s}^1_{[\boldsymbol{s}]}$ to be an element of $C_2^{\perp}$. That is, it is enough to chose it only with small calculation complexity. Therefore, we need to select the classical code C_2 so that there exists a set of representatives $\{\boldsymbol{s}^2_{[\boldsymbol{s}]}\}_{[\boldsymbol{s}]\in\mathbb{X}^r/C_2^{\perp}}$ with small calculation complexity for the decoding. Since they do not need to decode for C_1, they can choose it randomly under the condition that $C_1^{\perp}$ contains C_2. To satisfy this condition, we choose the code $C_1^{\perp}$ as a $(1,\epsilon)$-double universal2 code ensemble for $C_2 \subset \mathbb{X}^r$ [56, 122]. When the channel Λ describes the quantum communication channel to transmit these r random variables, the security of the final key of this protocol is the same as the security of the state transmitted via the channel Λ that is given in Theorem 5.9 [117]. Hence, the security can be guaranteed by the formulas (5.73) and (5.74) with the distribution $P[\Lambda]$ defined by the channel Λ. The process to estimate the distribution $P[\Lambda]$ is **(4): Error estimation**. This method cannot estimate the correlation between the errors of the computational and dual computational bases even though they have correlation. However, due to (5.76), the security can be evaluated by the formulas (5.73) and (5.74) with replacement of $\delta_{P[\Lambda]}[[C_1 \times C_2]]$ by $\delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^{\perp}/C_2]$.

When C_1 is chosen to be $(1,\epsilon)$-double universal2, the average security can be evaluated as follows. Let $W^E(\boldsymbol{x})$ be the eavesdropper's state dependently of Alice's final key $\boldsymbol{x}$ and P be the probability distribution subject to Alice's final key $\boldsymbol{x}$. Since the RHSs of (5.73) and (5.74) are convex for $\delta_{P[\Lambda]}[[C_1 \times C_2]]$, the inequality (5.73) guarantees

$$\mathrm{E}_{C_1} I(P, W^E) \le \mathrm{E}_{C_1} \delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^{\perp}/C_2] \log d + h(\mathrm{E}_{C_1} \delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^{\perp}/C_2]), \tag{5.78}$$

where d is the size of the final key. When Alice chooses the initial random number in $\mathbb{X}$ subject to the uniform distribution in **(1): Transmission**. Alice's final key $\boldsymbol{x}$ is also subject to the uniform distribution $P_{\max}$. Then, the inequality (5.74) yields

$$\mathrm{E}_{C_1} d_1(P_{\max}, W^E) \le 3\sqrt{\mathrm{E}_{C_1} \delta_{P[\Lambda]^{\mathbb{X}_2^r}}[C_1^{\perp}/C_2]}. \tag{5.79}$$

On the other hand, the above protocol cannot directly evaluate the $\mathrm{E}_{C_1} d_{1,\max}$ $(P_{\max}, W^E)$, which expresses the best case for the eavesdropper. To evaluate this criterion, we need to modify our protocol as follows. That is, Alice generates the random number subject to the uniform distribution in **(1): Transmission**, and **(5):**

Error correction is modified as follows. The protocol given in this way is called **twirling type-modified BB84 protocol** [53, 56].

(5)': Error correction Based on r random variables $\boldsymbol{s}$(Alice) and $\boldsymbol{s}'$(Bob), they generate corrected random numbers. Alice generates new random numbers $\boldsymbol{x} \in C_2^{\perp}$, and sends $\boldsymbol{y} := \boldsymbol{x} - \boldsymbol{s}$ to Bob via public channel. Bob applies the error correction to the random variable $\boldsymbol{y} + \boldsymbol{s}' = \boldsymbol{x} - \boldsymbol{s} + \boldsymbol{s}'$ and obtain the corrected information $\boldsymbol{x}' \in C_2^{\perp}$.

Now, let Λ be the channel of the communication for the above r random variables. The information obtained by the eavesdropper for the random variable $\boldsymbol{x} \in \mathbb{X}^r$ in the above protocol is the same as that in the following protocol. This fact can be checked from the invariance of the computational basis with respect to $\mathsf{W}_{\mathbb{X}}^r(\boldsymbol{s}, 0)$.

Protocol A Alice generates $\boldsymbol{x} \in C_2^{\perp}$ and $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{X}^r$ according to the respective uniform distributions, and sends $\boldsymbol{y}, \boldsymbol{z}$ to Bob via the public channel. Alice sets the initial state to be $|\boldsymbol{x}\rangle$, and operates the unitary $\mathsf{W}_{\mathbb{X}}^r(-\boldsymbol{y}, -\boldsymbol{z})$ on the state. Then, she transmits the state to Bob via the channel Λ. Bob operates the unitary $\mathsf{W}_{\mathbb{X}}^r(\boldsymbol{y}, \boldsymbol{z})$ on the received state, measure the state with the computational basis, and obtain the random variable $\boldsymbol{x}''$ as the outcome. Finally, Bob obtains the random number $\boldsymbol{x}' \in C_2^{\perp}$ by applying the error correction to the random number $\boldsymbol{x}''$.

In the above **Protocol A**, Alice sends the random variable $\boldsymbol{x} \in C_2^{\perp}$ to Bob via the channel $\Lambda[\mathsf{W}_{\mathbb{X}}^r, P[\Lambda]]$ that can be realized by applying the twirling to Λ [36, 56]. If the eavesdropper has the environment systems of the operation of $\mathsf{W}_{\mathbb{X}}^r(-\boldsymbol{y}, -\boldsymbol{z})$ before the transmission and the operation of $\mathsf{W}_{\mathbb{X}}^r(\boldsymbol{y}, \boldsymbol{z})$ after the transmission as well as the environment system of Λ, the information of the eavesdropper equals the the environment system of the Pauli channel $\Lambda[\mathsf{W}_{\mathbb{X}}^r, P[\Lambda]]$. Since the security analysis of twirling-type modified BB84 protocol can be reduced to the security analysis of **Protocol A**, the security analysis can be reduced to the security analysis with Pauli channel $\Lambda[\mathsf{W}_{\mathbb{X}}^r, P[\Lambda]]$ and the evaluation (5.50) can be applied. Therefore, the concavity of $x \mapsto \sqrt{x}$ in the RHS of (5.50) implies [56]

$$\mathrm{E}_{C_1} d_{1,\max}(P_{\max}, W^E) \le 4\sqrt{\mathrm{E}_{C_1} \delta_{P[\Lambda]^{\mathbb{X}_2^r}} [C_1^{\perp}/C_2]}.$$

The true distributions $P[\Lambda]^{\mathbb{X}_1^r}$ and $P[\Lambda]^{\mathbb{X}_1^r}$ are asymptotically close to r-fold independent and identical distributions of $\mathrm{P}^{\mathbb{X}_1}$ and $\mathrm{P}^{\mathbb{X}_2}$ guessed in Step **(4): Error estimation**, respectively. Hence, when $\mathbb{X} = \mathbb{F}_q$, due to Sect. 5.3.3, in order that $\mathrm{E}_{C_1} \delta_{P[\Lambda]^{\mathbb{X}_2^r}} [C_1^{\perp}/C_2]$ converges to 0, it is sufficient in the asymptotic regime to choose the size of C_1 to be $e^{rH(\mathrm{P}^{\mathbb{X}_2})}$. When we choose the code $C_2^{\perp}$ for the error correction as a sufficiently nice code, we can correct errors with almost probability 1 and can realize the size $e^{rH(\mathrm{P}^{\mathbb{X}_1})}$ of C_2. Hence, it is possible to generate the secure key with the asymptotic key generation rate $\log q - H(\mathrm{P}^{\mathbb{X}_1}) - H(\mathrm{P}^{\mathbb{X}_2})$.

However, in the real system, it is impossible to choose r to be infinity. So, we need to evaluate the RHSs of (5.78) and (5.79) with finite r. Such a security analysis is called finite-length security analysis. In this case, we need to evaluate the difference between the true distributions $P[\Lambda]^{\mathbb{X}_1^r}$ and $P[\Lambda]^{\mathbb{X}_2^r}$ and the distributions $\mathrm{P}^{\mathbb{X}_1}$ and $\mathrm{P}^{\mathbb{X}_2}$ guessed by Step **(4): Error estimation** by using the hypergeometric distribution, which requires a very complicated discussion [53, 61].

Chapter 6
Universal Information Processing

Abstract In a practical setting, it is not easy to perfectly identify the channel or the information source. To avoid this problem in classical information theory, we can employ universal code that works independently of the channel or the information source. In particular, Csiszár and Körner established the method of types for universal code, where the universality is the independence of the protocol from the channel or the information source. Schur duality is the joint representation of the special unitary group and the permutation group and can be regarded as quantum analogue of the method of types. Firstly, this chapter explains the method of types and Schur duality to clarify how Schur duality works as a quantum analogue of the method of types. Then, we proceed to universal codes or protocols in individual topics, estimation of density matrix, hypothesis testing of quantum state, entanglement concentration, quantum data compression, and classical-quantum channel, etc. When we do not care about the universality, we can discuss these topics without use of representation theory. However, to construct protocols to achieve the universality, we need to employ Schur duality theory because they cannot be constructed without use of Schur duality.

6.1 Method of Types

This chapter addresses various types of universal information processings based on Schur duality, whose detail explanation is available in [44, Sect. 4.4]. As its preparation, we deal with type method in classical information theory, which is closely related to the asymptotic behavior of Schur duality. In the following, we describe an probability distribution on the probability space $\mathcal{X} := \{1, \dots, k\}$ by a vector $\boldsymbol{p} = (p_1, \dots, p_k)$, and denote the set of such vectors by $\mathcal{P}(k)$ or $\mathcal{P}(\mathcal{X})$. In the following, for given two probability distributions $\boldsymbol{p}$ and $\boldsymbol{q} = (q_1, \dots, q_k)$, the entropy $H(\boldsymbol{p}) = -\sum_{j=1}^{k} p_j \log p_j$ and the relative entropy $D(\boldsymbol{p}\|\boldsymbol{q}) = \sum_{j=1}^{k} p_j(\log p_j - \log q_j)$ defined in Sect. 2.4 play an important role. In the following, we consider n independent and identical trials under one of the these distributions. Then, the probability space is given as the set $\{1, \dots, k\}^n$, and the probability distribution is denoted by $\boldsymbol{p}^n$, $\boldsymbol{q}^n$ etc. The additivity of entropy and relative entropy imply that $H(\boldsymbol{p}^n) = nH(\boldsymbol{p})$ and $D(\boldsymbol{p}^n\|\boldsymbol{q}^n) = nD(\boldsymbol{p}\|\boldsymbol{q})$. For a sequence $\vec{x} = (x_1, \dots, x_n) \in \{1, \dots, k\}^n$, we define

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8_6

its type $\boldsymbol{n} = (n_1, \dots, n_k)$ with $n_j := \#\{l | x_l = j\}$.[1] That is, the empirical distribution of the sequence is $\frac{\boldsymbol{n}}{n} = (\frac{n_1}{n}, \dots, \frac{n_k}{n})$. Given a sequence $\vec{x}$, we denote its type by $\boldsymbol{n}(\vec{x}) = (n_1(\vec{x}), \dots, n_k(\vec{x}))$, and calculate the probability of its occurrence to be $\prod_{j=1}^{k} p_j^{n_j(\vec{x})}$, which can be expressed with entropy and relative entropy as follows;

$$\prod_{j=1}^{k} p_j^{n_j(\vec{x})} = e^{-n(D(\frac{\boldsymbol{n}(\vec{x})}{n} \| \boldsymbol{p}^n) + H(\frac{\boldsymbol{n}(\vec{x})}{n}))}. \tag{6.1}$$

Now, we denote the set of types of n trials by $\mathcal{T}_n^k$ or $\mathcal{T}_n(\mathcal{X})$. Given $\boldsymbol{n} \in \mathcal{T}_n^k$, we denote the set of sequences $\vec{x} \in \mathcal{X}^n = \{1, \dots, k\}^n$ satisfying $\boldsymbol{n}(\vec{x}) = \boldsymbol{n}$ by $T_{\boldsymbol{n}}$. Then, the cardinality $|T_{\boldsymbol{n}}|$ is $\frac{n!}{\boldsymbol{n}!}$, where $\boldsymbol{n}! := n_1! n_2! \dots n_k!$.

Hence, the n-fold k-nomial distribution is given as a probability distribution $\mathrm{Mul}[\boldsymbol{p}, n]$ on $\mathcal{T}_n^k$, which is defined as

$$\mathrm{Mul}[\boldsymbol{p}, n](\boldsymbol{n}) := \frac{n!}{\boldsymbol{n}!} \prod_{j=1}^{k} p_j^{n_j}, \quad \boldsymbol{n} = (n_1, \dots, n_k) \in \mathcal{T}_n^k, \tag{6.2}$$

where $\mathcal{T}_n^k$ is the set of k integers $n_1, \dots, n_k$ satisfying $\sum_{j=1}^{k} n_j = n$. Since each n_j takes the value in the set $\{0, \dots, n\}$ and the k-th integer n_k is determined by the other integers, the cardinality $|\mathcal{T}_n^k|$ is upper bounded by $(n+1)^{k-1}$. Using **Stirling formula** $n! \cong \sqrt{2\pi n} n^n e^{-n}$, we have

$$\begin{aligned}
\log |\mathcal{T}_n^k| &= \log \binom{n+k}{k-1} + o(1) \\
&= \log \frac{\sqrt{2\pi(n+k)}(n+k)^{n+k} e^{-(n+k)}}{\sqrt{2\pi(n+1)}(n+1)^{n+1} e^{-(n+1)} (k-1)!} + o(1) \\
&= \log \frac{(n+k)^{k-1}}{(k-1)!} \sqrt{\frac{n+k}{n+1}} e^{-k+1} (1 + \frac{k-1}{n+1})^{n+1} + o(1) \\
&= \log \frac{(n+k)^{k-1}}{(k-1)!} + o(1) \\
&= (k-1) \log n - \log(k-1)! + o(1).
\end{aligned} \tag{6.3}$$

Further, $\frac{n!}{\boldsymbol{n}!}$ is bounded as

$$\frac{1}{(n+1)^{k-1}} e^{nH(\frac{\boldsymbol{n}}{n})} \le \frac{n!}{\boldsymbol{n}!} \le e^{nH(\frac{\boldsymbol{n}}{n})}. \tag{6.4}$$

[1] Csiszár-Körner [25] call $\frac{\boldsymbol{n}}{n}$ a type.

Hence, the relations (6.1) and (6.4) yield

$$\frac{1}{(n+1)^{k-1}} e^{-nD(\frac{\boldsymbol{n}}{n}\|\boldsymbol{p})} \le \boldsymbol{p}^n(T_{\boldsymbol{n}}) \le e^{-nD(\frac{\boldsymbol{n}}{n}\|\boldsymbol{p})}. \tag{6.5}$$

The above discussion guarantees that the random variable $\frac{\boldsymbol{n}}{n}$ converges to $\boldsymbol{p}$ in probability. That is, when the set $\mathcal{R}$ has no intersection with a certain neighborhood of $\boldsymbol{p}$, the probability that the random variable $\frac{\boldsymbol{n}}{n}$ belongs to the set $\mathcal{R}$ converges to zero exponentially for n. Hence, when $\boldsymbol{p}$ is unknown, the random variable $\frac{\boldsymbol{n}}{n}$ is can be used as an estimate of $\boldsymbol{p}$. When $\frac{\boldsymbol{n}}{n}$ is used as an estimate, the precision is expressed by the exponential decreasing rate of the probability $\Pr\{\frac{\boldsymbol{n}}{n} \in \mathcal{R}\}$. This type evaluation is called **large deviation**. Therefore, summarizing the above discussion, we obtain the following lemma for large deviation.

Lemma 6.1 *Given a subset $\mathcal{R}$ of $\mathcal{P}(k)$, we define the subset $\mathcal{R}_n := \{\boldsymbol{q} \in \mathcal{R} | n\boldsymbol{q} \in \mathcal{T}_n^k\}$ of $\mathcal{R}$. Then, the probability that the empirical distribution $\frac{\boldsymbol{n}}{n}$ belongs to the subset $\mathcal{R}$ is bounded as*

$$\frac{1}{(n+1)^{k-1}} \max_{\boldsymbol{q}\in\mathcal{R}_n} e^{-nD(\boldsymbol{q}\|\boldsymbol{p})} \le \sum_{\frac{\boldsymbol{n}}{n}\in\mathcal{R}} \boldsymbol{p}^n(T_{\boldsymbol{n}}) \le (n+1)^{k-1} \max_{\boldsymbol{q}\in\mathcal{R}} e^{-nD(\boldsymbol{q}\|\boldsymbol{p})}. \tag{6.6}$$

This fact implies that

$$\lim_{n\to\infty} \frac{-1}{n} \log \sum_{\frac{\boldsymbol{n}}{n}\in\mathcal{R}} \boldsymbol{p}^n(T_{\boldsymbol{n}}) = \min_{\boldsymbol{q}\in\mathcal{R}} D(\boldsymbol{q}\|\boldsymbol{p}). \tag{6.7}$$

Hence, substituting the subset $\{\boldsymbol{p} \in \mathcal{P}(k) | H(\boldsymbol{p}) \le R\}$ or $\{\boldsymbol{p} \in \mathcal{P}(k) | H(\boldsymbol{p}) \ge R\}$ into $\mathcal{R}$, we obtain the following lemma.

Lemma 6.2

$$\lim_{n\to\infty} \frac{-1}{n} \log \sum_{\boldsymbol{n}:H(\frac{\boldsymbol{n}}{n})\le R} \boldsymbol{p}^n(T_{\boldsymbol{n}}) = \min_{\boldsymbol{q}:H(\boldsymbol{q})\le R} D(\boldsymbol{q}\|\boldsymbol{p}). \tag{6.8}$$

$$\lim_{n\to\infty} \frac{-1}{n} \log \sum_{\boldsymbol{n}:H(\frac{\boldsymbol{n}}{n})\ge R} \boldsymbol{p}^n(T_{\boldsymbol{n}}) = \min_{\boldsymbol{q}:H(\boldsymbol{q})\ge R} D(\boldsymbol{q}\|\boldsymbol{p}). \tag{6.9}$$

Notice that the RHSs can be expressed by $\psi(s|\boldsymbol{p})$ due to the relations (2.76) and (2.77).

The following lemma is known [24].

Lemma 6.3 *When the random variable $\boldsymbol{n}$ is subject to the multinomial distribution* Mul$[\boldsymbol{p}, n]$, *the random variable $\frac{\boldsymbol{n}}{n}$ converges to $\boldsymbol{p}$ in probability. Further, the random variable $\frac{\boldsymbol{n}-n\boldsymbol{p}}{\sqrt{n}}$ is asymptotically subject to the normal distribution with the covariance matrix*

$$C_{i,j} = \begin{cases} -p_i p_j & \textit{when } i \neq j \\ p_i(1-p_i) & \textit{when } i = j. \end{cases}$$

Now, we apply Lemmas 6.1 and 6.3 to the quantum system. These lemmas give the limiting distribution and the large deviation analysis as asymptotic evaluation of the error when the eigenvectors of the true density matrix are known and the eigenvector $\boldsymbol{p}$ is to be estimated. For example, when the eigenvectors are the computational basis, this discussion gives the estimation when the true density matrix belongs to the state family $\{\rho(\boldsymbol{p})\}$.

Lemma 6.4 *Assume that the random variable* $\boldsymbol{n} \in \mathcal{T}_n^k$ *is subject to the multinomial distribution* Mul$[\boldsymbol{p}, n]$. *Taking the expectations, we have the following relations.*

$$\lim_{n\to\infty} \mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]} \left[H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p}) \right] = \frac{k-1}{2} \tag{6.10}$$

and

$$\lim_{n\to\infty} \mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]} \left[\log \frac{n!}{\boldsymbol{n}!} - nH(\boldsymbol{p}) + \frac{k-1}{2} \log n \right]$$
$$= -\frac{k-1}{2} \log \frac{2\pi}{e} - \frac{1}{2} \sum_{j=1}^{k} \log p_j. \tag{6.11}$$

Further, the random variables $\sqrt{n}(H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p}))$ *and* $\sqrt{n}(\frac{1}{n} \log \frac{n!}{\boldsymbol{n}!} - H(\boldsymbol{p}))$ *asymptotically obey the normal distribution with average* 0 *and variance* $V(\boldsymbol{p})$ *given in given (2.28).*

Proof Due to Stirling formula $n! \cong \sqrt{2\pi n} n^n e^{-n}$, we have

$$\log \frac{n!}{\boldsymbol{n}!} = nH(\frac{\boldsymbol{n}}{n}) - \frac{k-1}{2} \log n - \frac{k-1}{2} \log 2\pi - \frac{1}{2} \sum_{j=1}^{k} \log \frac{n_j}{n} + o(1). \tag{6.12}$$

Further, when $\frac{n_j}{n}$ belongs to the neighborhood of p_j, the Taylor expansion of $H(\frac{\boldsymbol{n}}{n})$ up to the second derivative yields that

$$nH(\frac{\boldsymbol{n}}{n}) = nH(\boldsymbol{p}) + \sum_{j=1}^{k} \frac{\partial H(\boldsymbol{p})}{\partial p_j}(n_j - p_j) + \frac{1}{2}(n - \sum_{j=1}^{k} \frac{n_j^2}{p_j n}) + o(1). \tag{6.13}$$

Since the expectation of n_j^2 is $np_j(1-p_j) + (np_j)^2$, we have $\frac{1}{2}\mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]}(n - \sum_{j=1}^{k} \frac{n_j^2}{p_j n}) = \frac{k-1}{2}$. Since the expectation of $\frac{n_j}{n}$ is p_j, we have (6.10). Since $\frac{1}{2n}(n - \sum_{j=1}^{k} \frac{n_j^2}{p_j n})$ converges to 0 in probability, the random variable $\sqrt{n}(H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p}))$ asymptotically obeys the normal distribution with average 0 whose variance is

$$V(\boldsymbol{p}) = \sum_{i,j} \frac{\partial H(\boldsymbol{p})}{\partial p_i} C_{i,j} \frac{\partial H(\boldsymbol{p})}{\partial p_j}. \tag{6.14}$$

For derivation of (6.14), see Exercise 6.3. Since (6.12) shows that $\sqrt{n}(\frac{1}{n}\log\frac{n!}{\boldsymbol{n}!} - H(\frac{\boldsymbol{n}}{n}))$ converges to 0 in probability, the other random variable $\sqrt{n}(\frac{1}{n}\log\frac{n!}{\boldsymbol{n}!} - H(\boldsymbol{p}))$ asymptotically obeys the normal distribution with average 0 and variance $V(\boldsymbol{p})$.

Since $\frac{n_j}{n}$ converges to p_j in probability, we have

$$\begin{aligned}
&\mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]}\Big[\log\frac{n!}{\boldsymbol{n}!}\Big] \\
=& nH(\boldsymbol{p}) + \sum_{j=1}^{k} \frac{\partial H(\boldsymbol{p})}{\partial p_j} \mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]}(n_j - p_j) + \frac{1}{2}\mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]}(n - \sum_{j=1}^{k}\frac{n_j^2}{p_j n}) \\
&- \frac{k-1}{2}\log n - \frac{k-1}{2}\log 2\pi - \mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]}\frac{1}{2}\sum_{j=1}^{k}\log\frac{n_j}{n} + o(1) \\
=& nH(\boldsymbol{p}) + \frac{k-1}{2} - \frac{k-1}{2}\log n - \frac{k-1}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{k}\log p_j + o(1),
\end{aligned}$$

which implies the desired statement. ■

Exercise 6.1 List up all of sequences in $\{1, 2\}^6$ corresponding to the type $(2, 6)$.

Exercise 6.2 Show Lemma 6.3.

Exercise 6.3 Show (6.14).

6.2 Asymptotic Theory for Schur Duality

6.2.1 Asymptotic Behavior of Schur Duality

Next, we address the asymptotic behaviors of several quantities appearing in Schur duality. The details of Schur duality are explained in (4.62) of [44] and Schur duality can be regarded as the non-commutative version of type method as Fig. 6.1. For simplicity, given an element $\boldsymbol{n} \in \mathcal{Y}_n^r$, we denote the irreducible representation space $\mathcal{U}_{in^*(\boldsymbol{n})}(\mathrm{SU}(r))$ of the special unitary group $\mathrm{SU}(r)$ and the irreducible representation space $\mathcal{U}_{\boldsymbol{n}}(S_n)$ of the permutation group S_n by $\mathcal{U}_{\boldsymbol{n}}$ and $\mathcal{V}_{\boldsymbol{n}}$, respectively. Hence, Schur duality is simplified as

$$(\mathbb{C}^r)^{\otimes n} = \bigoplus_{\boldsymbol{n}\in\mathcal{Y}_n^r} \mathcal{W}_{\boldsymbol{n}}, \quad \mathcal{W}_{\boldsymbol{n}} := \mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}. \tag{6.15}$$

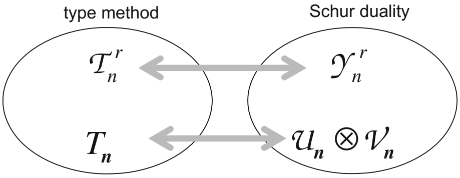


Fig. 6.1 Relation between type method and Schur duality

Then, we denote the projection to $\mathcal{W}_{\boldsymbol{n}}$ by $P_{\boldsymbol{n}}$. In the following, we consider the case when the diagonal elements of the density matrix ρ on $\mathbb{C}^r$ are given as the probability distribution $\boldsymbol{p} = (p_1, \ldots, p_r)$ satisfying $p_1 \geq \cdots \geq p_r$. The set of such probability distributions is denoted by $\mathcal{Y}(r)$. The dimension of $\dim \mathcal{U}_{\boldsymbol{n}}$ is evaluated as

$$\dim \mathcal{U}_{\boldsymbol{n}} \overset{(a)}{=} \prod_{1 \leq j < l \leq r} \frac{l - j + n_l - n_j}{l - j} \overset{(b)}{\leq} (n+1)^{r(r-1)/2}, \tag{6.16}$$

where (a) is shown in [44, (4.46)] and (b) follows from the inequality $\frac{l-j+n_l-n_j}{l-j} \leq n_l - n_j + 1 \leq n + 1$ for respective $j < l$. Thus,

$$\begin{aligned}&\log \dim \mathcal{U}_{\boldsymbol{n}} \\ =&\frac{r(r-1)}{2} \log n + \sum_{1 \leq j < l \leq r} \log(\frac{n_l}{n} - \frac{n_j}{n}) - \sum_{j=1}^{r-1} (r-j) \log j + o(1).\end{aligned} \tag{6.17}$$

Then, the cardinality $|\mathcal{Y}_n^r|$ (the number of Young diagram with depth is not greater than r and size n) is bounded as

$$|\mathcal{Y}_n^r| \leq |\mathcal{T}_n^r| \leq (n+1)^{r-1}. \tag{6.18}$$

Now, we focus on the ratio of elements containing same entries among elements $\boldsymbol{n}$ of $\mathcal{T}_n^r$. The ratio goes to zero as n goes to infinity. Hence, at the limit $n \to \infty$, the cardinality of the set $\mathcal{Y}_n^r$ almost equals the cardinality of the set $\mathcal{T}_n^r$ divided by $r!$. So, the relation (6.3) implies that

$$\log |\mathcal{Y}_n^r| \cong (r-1) \log n - \log (r-1)! r! + o(1). \tag{6.19}$$

Further, the relation [44, (2.72)] yields the following evaluation of the dimension of $\dim \mathcal{V}_{\boldsymbol{n}}$ by noticing that the inequality $n_i + j \geq (n_i - n_j - i + j)$ for $r \geq j > i \geq 1$.

$$\begin{aligned}\dim \mathcal{V}_{\boldsymbol{n}} =&\frac{n!}{(n_1 + r - 1)!(n_2 + r - 2)! \ldots r_d!} \prod_{j>i} (n_i - n_j - i + j) \\ =&\frac{n!}{n_1! n_2! \ldots n_r!} \prod_{j>i} \frac{n_i - n_j - i + j}{n_i + j - 1} \leq \frac{n!}{\boldsymbol{n}!} \leq e^{nH(\frac{\boldsymbol{n}}{n})}.\end{aligned} \tag{6.20}$$

Since $(n_i - n_j - i + j) \geq 1$ for $r \geq j > i \geq 1$, the relation (6.4) yields the opposite inequality;

$$\begin{aligned}
\dim \mathcal{V}_{\boldsymbol{n}} \geq & \frac{n!}{(n_1+r-1)!(n_2+r-2)!\ldots n_d!} \\
\geq & \frac{n!}{n_1!n_2!\ldots n_d!}\left(\frac{1}{n+r}\right)^{r-1}\left(\frac{1}{n+r}\right)^{r-2}\cdots\left(\frac{1}{n+r}\right)^{0} \\
= & \frac{n!}{\boldsymbol{n}!}\left(\frac{1}{n+r}\right)^{\frac{r(r-1)}{2}} \geq e^{nH(\frac{\boldsymbol{n}}{n})}(n+r)^{-\frac{(r+2)(r-1)}{2}}. \qquad (6.21)
\end{aligned}$$

Therefore, we find that the differences among $\log \dim \mathcal{V}_{\boldsymbol{n}}$, $\log \frac{n!}{\boldsymbol{n}!}$, and $nH(\frac{\boldsymbol{n}}{n})$ are bounded by $O(\log n)$. That is, the differences among $\frac{1}{\sqrt{n}} \log \dim \mathcal{V}_{\boldsymbol{n}}$, $\frac{1}{\sqrt{n}} \log \frac{n!}{\boldsymbol{n}!}$, and $\sqrt{n}H(\frac{\boldsymbol{n}}{n})$ converges to zero in probability.

6.2.2 Universal Eigenvalue Estimation

Applying the discussion for the estimation for the multinomial distribution given in Sect. 6.1 to the quantum system, we can estimate the eigenvalues of the density matrix when the eigenvector is known. However, it is usual that the eigenvector is unknown. To estimate the eigenvalue of the density matrix in this case, we need to estimate the eigenvalues without use of the information about the eigenvector. In the following, we deal with the estimation of the eigenvalues of the unknown density matrix ρ by using a measurement on the n-fold tensor product system $\mathcal{H}^{\otimes n}$ when the system $\mathcal{H}^{\otimes n}$ is prepared in the n-fold tensor product state $\rho^{\otimes n}$ as Fig. 6.2.

Then, the density matrix is written as $g\rho(\boldsymbol{p})g^{\dagger}$ by using the sequence of the eigenvalues $\boldsymbol{p}$ and the element $[g]$ in the homogenous space $\mathrm{SU}(r)/H$ of $\mathrm{SU}(r)$ quotiented by the stabilizer H of $\rho(\boldsymbol{p})$. We denote our estimate by $\hat{\boldsymbol{p}}$, and describe the error by using a function of $\hat{\boldsymbol{p}}$ and $\boldsymbol{p}$. This formulation gives a special case of the general model given in Remark 4.1. In this general model, the parametric space Θ is the whole set of density matrices on the system $\mathbb{C}^r$ and the set of estimates is $\mathcal{Y}(r)$. Since $\hat{\boldsymbol{p}}$ and $\boldsymbol{p}$ are invariant with respect to the group action, the function describing the error also satisfies the invariance. Hence, due to a discussion similar to Theorem 4.1, without loss of generality, we can restrict our measurements to measurements satisfying the condition (4.8), i.e.,

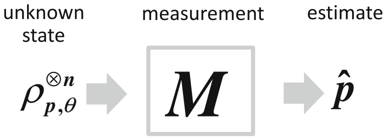


Fig. 6.2 Universal eigenvalue estimation: θ is the parameter for the unitary direction. p is the parameter describing the eigenvalues

$$g^{\otimes n} M_{\hat{\boldsymbol{p}}} g^{\otimes n \dagger} = M_{\hat{\boldsymbol{p}}}. \tag{6.22}$$

Then, the POVM element $M_{\hat{\boldsymbol{p}}}$ is a constant times of projection to an irreducible subspace or a sum of such matrices. In the following, we employ only a POVM $M_{\frac{\boldsymbol{n}}{n}} := P_{\boldsymbol{n}}$ taking the values $\frac{\boldsymbol{n}}{n}$ for our estimation [50, 87, 93], which satisfies the condition (6.22). To investigate the behavior of the estimate, we focus on the probability distribution $Q[\rho^{\otimes n}](\boldsymbol{n}) := \operatorname{Tr} P_{\boldsymbol{n}} \rho^{\otimes n}$. Especially, we denote the probability distribution $Q[\rho^{\otimes n}(\boldsymbol{p})]$ by $Q_{\boldsymbol{p}}$ when the eigenvalues of ρ are given as $\boldsymbol{p}$. When the range of ρ is the whole of $\mathcal{H} = \mathbb{C}^r$, the density matrix ρ can be regarded as a element of $\mathrm{GL}(r)$. Hence, we have $\rho^{\otimes n} = \oplus_{\boldsymbol{n} \in \mathcal{Y}_n^r} \mathsf{f}_{\boldsymbol{n}}(\rho) \otimes I_{\mathcal{V}_{\boldsymbol{n}}}$.

On the other hand, when the range of ρ is a subspace $\mathcal{K} = \mathbb{C}^{r'}$ of $\mathcal{H} = \mathbb{C}^r$, the above discussion becomes more complicated. We give the direct sum decomposition of $\mathcal{K}^{\otimes n}$ as $\mathcal{K}^{\otimes n} = \oplus_{\boldsymbol{n} \in \mathcal{Y}_n^{r'}} \mathcal{U}'_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}$. Then, $\mathsf{f}_{\boldsymbol{n}}(\rho)$ is a matrix on $\mathcal{U}'_{\boldsymbol{n}}$. Since $\mathcal{U}'_{\boldsymbol{n}}$ is a subspace of $\mathcal{U}_{\boldsymbol{n}}$, $\mathsf{f}_{\boldsymbol{n}}(\rho)$ can be regarded as a matrix on $\mathcal{U}_{\boldsymbol{n}}$. So, $\rho^{\otimes n}$ is written as $\rho^{\otimes n} = \oplus_{\boldsymbol{n} \in \mathcal{Y}_n^{r'}} \mathsf{f}_{\boldsymbol{n}}(\rho) \otimes I_{\mathcal{V}_{\boldsymbol{n}}}$. Hence, letting $\mathcal{W}'_{\boldsymbol{n}} := \mathcal{U}'_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}$, we have $\operatorname{Tr} \rho^{\otimes n} P_{\boldsymbol{n}} = \operatorname{Tr} \rho^{\otimes n} P'_{\boldsymbol{n}}$, which shows that the behavior of the measurement outcome is given by the distribution of the measurement outcome of the measurement based on the irreducible decomposition on the tensor product space on the range of ρ even though the whole system is larger than the range of ρ. Also, the above discussion yields that $\|P_{\boldsymbol{n}} \rho^{\otimes n} P_{\boldsymbol{n}}\| = \prod_{i=1}^r p_i^{n_i}$.

In the following, we investigate the behavior of $Q[\rho^{\otimes n}](\boldsymbol{n})$. Notice that the norm of $P_{\boldsymbol{n}} \rho^{\otimes n} P_{\boldsymbol{n}}$ is $\prod_{i=1}^r p_i^{n_i}$. Then, the relations (6.1), (6.16), and (6.20) guarantee that

$$\begin{aligned}
&\operatorname{Tr} P_{\boldsymbol{n}} \rho^{\otimes n} \le \dim \mathcal{V}_{\boldsymbol{n}} \cdot \dim \mathcal{U}_{\boldsymbol{n}} \cdot \prod_{i=1}^r p_i^{n_i} \\
\le & e^{nH(\frac{\boldsymbol{n}}{n})} (n+1)^{r(r-1)/2} e^{-n(D(\frac{\boldsymbol{n}}{n} \| \boldsymbol{p}^n) + H(\frac{\boldsymbol{n}}{n}))} \\
= & (n+1)^{r(r-1)/2} e^{-nD(\frac{\boldsymbol{n}}{n} \| \boldsymbol{p}^n)}.
\end{aligned} \tag{6.23}$$

Conversely, when $\rho^{\otimes n}$ is regarded as a representation of $\mathrm{GL}(r, \mathbb{C})$, the representation space is $\mathcal{U}_{\boldsymbol{n}}$ and the space $\mathcal{V}_{\boldsymbol{n}}$ describes only the multiplicity. So, (6.21) implies that

$$\begin{aligned}
&\operatorname{Tr} P_{\boldsymbol{n}} \rho^{\otimes n} \ge \dim \mathcal{V}_{\boldsymbol{n}} \cdot \prod_{i=1}^r p_i^{n_i} \\
\ge & e^{nH(\frac{\boldsymbol{n}}{n})} (n+r)^{-\frac{(r+2)(r-1)}{2}} e^{-n(D(\frac{\boldsymbol{n}}{n} \| \boldsymbol{p}^n) + H(\frac{\boldsymbol{n}}{n}))} \\
= & (n+r)^{-\frac{(r+2)(r-1)}{2}} e^{-nD(\frac{\boldsymbol{n}}{n} \| \boldsymbol{p}^n)}.
\end{aligned}$$

Hence, similar to Lemma 6.1, we obtain the following lemma.

Lemma 6.5 *Assume that we perform the measurement* $\{P_{\boldsymbol{n}}\}_{\boldsymbol{n} \in \mathcal{Y}_n^r}$ *to the system whose state is* $\rho^{\otimes n}$*. Then, the probability that the outcome* $\frac{\boldsymbol{n}}{n}$ *belongs to the subspace* $\mathcal{R}$ *of* $\mathcal{Y}(r)$ *is evaluated as*

$$\begin{aligned}(n+r)^{-\frac{(r+2)(r-1)}{2}} \max_{\boldsymbol{q}\in\mathcal{R}_n} e^{-nD(\boldsymbol{q}\|\boldsymbol{p})} \le & \sum_{\frac{\boldsymbol{n}}{n}\in\mathcal{R}} \operatorname{Tr} P_{\boldsymbol{n}}\rho^{\otimes n} \\ \le & (n+1)^{\frac{(r+2)(r-1)}{2}} \max_{\boldsymbol{q}\in\mathcal{R}} e^{-nD(\boldsymbol{q}\|\boldsymbol{p})},\end{aligned} \tag{6.24}$$

where $\mathcal{R}_n := \{\boldsymbol{q} \in \mathcal{Y} | n\boldsymbol{q} \in \mathcal{Y}_n^r\}$.

Lemmas 6.1 and 6.5 give the same exponential decreasing rate for the probability that the estimate is out of a certain neighbor hood of the true value. That is, the exponential decreasing rate of the above probability does not depend on the presence or absence of the knowledge of the eigenvector of the density matrix. Due to Lemma 6.5, when the random variable $\boldsymbol{n}$ is subject to the probability distribution $Q[\rho^{\otimes n}]$, the random variable $\frac{\boldsymbol{n}}{n}$ converges to $\boldsymbol{p}$ in probability.

As a more precise evaluation, we have the following lemma [93].

Lemma 6.6 *Assume that* $p_1 > \cdots > p_r$. *When the random variable* $\boldsymbol{n}$ *is subject to the probability distribution* $Q[\rho^{\otimes n}]$, *the expectation of the random variable* $\frac{\boldsymbol{n}-n\boldsymbol{p}}{n}$ *converges to* 0, *and the difference between the expectation and zero has the order* $O(\frac{1}{n})$. *The limiting distribution of the random variable* $\frac{\boldsymbol{n}-n\boldsymbol{p}}{\sqrt{n}}$ *is the normal distribution whose covariance matrix is*

$$C_{i,j} = \begin{cases} -p_i p_j & \text{when } i \neq j \\ p_i(1-p_i) & \text{when } i = j. \end{cases}$$

Hence, when ϵ_n *satisfies* $\sqrt{n}\epsilon_n \to \infty$, *the probability to satisfy* $|\frac{\boldsymbol{n}}{n} - \boldsymbol{p}| \le \epsilon_n$ *converges to* 1.

Lemmas 6.3 and 6.6 give the same limiting distribution for the estimate for the neighbor hood of the true eigenvalues. That is, the asymptotic behavior of the estimate around the true value does not depend on the presence or absence of the knowledge of the eigenvector of the density matrix. The estimation of the eigenvalue does not depend on the eigenvector universally.

Also, combining Lemma 6.6 with (6.13), (6.14), and the discussion in the end of Sect. 6.2.1, we have the following corollary.

Corollary 6.1 *When the random variable* $\boldsymbol{n}$ *is subject to the probability distribution* $Q[\rho^{\otimes n}]$, *the random variables* $\sqrt{n}(H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p}))$, $\sqrt{n}(\frac{1}{n}\log\frac{n!}{\boldsymbol{n}!} - H(\boldsymbol{p}))$, *and* $\sqrt{n}(\frac{1}{n}\log\dim\mathcal{V}_{\boldsymbol{n}} - H(\boldsymbol{p}))$ *asymptotically obey the normal distribution with average* 0 *and variance* $V(\boldsymbol{p})$ *given in (2.28).*

Proof Due to the relation (4.60) in [44], the character $\chi_{\boldsymbol{n}}(\rho)$ of the representation in $\mathcal{U}_{\boldsymbol{n}}$ is calculated as

$$\chi_{\boldsymbol{n}}(\rho) = \frac{\sum_{\sigma'\in S_r} \operatorname{sgn}(\sigma') \prod_{l=1}^r p_l^{n_{\sigma'(l)}+\delta_{\sigma'(l)}}}{\prod_{j<l}(p_l - p_j)}.$$

Lemma 6.5 guarantees that the probability that $\boldsymbol{n}$ is not sufficiently close to $n\boldsymbol{p}$ is almost 0. Hence, we can consider only the case when $\frac{n_j}{n} \in (p_j - \epsilon, p_j + \epsilon)$ for any small real number $\epsilon > 0$ satisfying the condition $p_{j-1} - \epsilon > p_j + 2\epsilon$. Here, for simplicity, we assume that σ' is the transposition between 1 and 2. Then, we have

$$\frac{\prod_{l=1}^r p_l^{n_{\sigma'(l)}+\delta_{\sigma'(l)}}}{\prod_{l=1}^r p_l^{n_l+\delta_l}} \le \frac{\prod_{l=1}^r p_l^{\delta_{\sigma'(l)}}}{\prod_{l=1}^r p_l^{\delta_l}} (\frac{p_2}{p_1})^{2n\epsilon},$$

which goes to zero exponentially. We can show that the above value with any non-identity element $\sigma' \in S_r$ goes to zero by the same way. Hence, we can ignore all of the above terms when σ' is not the identity element.

Due to [44, (2.71)], the above condition shows that all of the components of $\boldsymbol{n}^\sigma := \boldsymbol{n} + \boldsymbol{\delta} - \sigma(\boldsymbol{\delta})$ are positive. Hence,

$$\begin{aligned} Q[\rho^{\otimes n}](\boldsymbol{n}) =& \chi_{\boldsymbol{n}}(\rho) \cdot \dim \mathcal{V}_{\boldsymbol{n}} \cong \frac{\prod_{l=1}^r p_l^{n_l+\delta_l}}{\prod_{j<l}(p_l - p_j)} \sum_{\sigma \in S_r} \operatorname{sgn}(\sigma) \frac{n!}{[\boldsymbol{n} + \boldsymbol{\delta} - \sigma(\boldsymbol{\delta})]!} \\ =& \sum_{\sigma \in S_r} \operatorname{sgn}(\sigma) \frac{\prod_{l=1}^r p_l^{\delta_{\sigma(l)}}}{\prod_{j<l}(p_l - p_j)} \prod_{l=1}^r p_l^{n_l^\sigma} \frac{n!}{\boldsymbol{n}^\sigma !}. \end{aligned}$$

The RHS can be regarded as the weighted sum of the probability distribution $\mathrm{Mul}[\boldsymbol{p}, n, \sigma](\boldsymbol{n}) := \prod_{l=1}^r p_l^{n_l^\sigma} \frac{n!}{\boldsymbol{n}^\sigma !}$ that depends on $\sigma \in S_r$, whose weighted coefficient is $\operatorname{sgn}(\sigma) \frac{\prod_{l=1}^r p_l^{\delta_{\sigma(l)}}}{\prod_{j<l}(p_l - p_j)}$. The expectation of $\frac{\boldsymbol{n}}{n}$ is $\boldsymbol{p} + \frac{\sigma(\boldsymbol{\delta}) - \boldsymbol{\delta}}{n}$ under the probability distribution $\mathrm{Mul}[\boldsymbol{p}, n, \sigma](\boldsymbol{n})$. Hence, $\frac{\boldsymbol{n}}{n} - \boldsymbol{p}$ converges to 0 in probability and its expectation goes to zero with the order $O(\frac{1}{n})$. Further, since the above discussion shows that $\frac{\boldsymbol{n}^\sigma}{\sqrt{n}} - \frac{\boldsymbol{n}}{\sqrt{n}}$ goes to zero, the limiting distribution of $\boldsymbol{n}$ is the normal distribution with covariance matrix $C_{i,j}$ whatever probability distribution among the above distribution gives the stochastic behavior of $\boldsymbol{n}$. Finally, since Vandermonde determinant yields

$$\sum_{\sigma \in S_r} \operatorname{sgn}(\sigma) \frac{\prod_{l=1}^r p_l^{\delta_{\sigma(l)}}}{\prod_{j<l}(p_l - p_j)} = 1, \tag{6.25}$$

we obtain the desired statement. ■

Unfortunately, the above discussion does not treat the case when the true density matrix is close to the completely mixed state. In this case, the evaluation in Lemma 6.6 does not hold. In the qubit case, as shown in [62, Sect. 13.5], the error asymptotically obeys not the normal distribution but the χ^2-distribution with 3 degrees.

Exercise 6.4 Show Corollary 6.1 by combining Lemma 6.6, (6.13), (6.14), and the discussion in the end of Sect. 6.2.1.

Exercise 6.5 Assume that ρ has the eigenvalues p and $1-p$ with $p \in (1/2, 1)$ and that k is an integer between n and $n/2$ such that $n-k$ is an even number. Show the following relation

$$\begin{aligned} & Q[\rho^{\otimes n}]\Big(\frac{n+k}{2}, \frac{n-k}{2}\Big) \\ = & \frac{k+1}{(2p-1)(n+1)}\Big(B\Big(n+1, p, \frac{n+k}{2}+1\Big) - B\Big(n+1, p, \frac{n-k}{2}+1\Big)\Big) \end{aligned}$$

by using (2.74) and (4.59) in [44], where $B(n, p, j) := p^j(1-p)^j\binom{n}{j}$.

6.2.3 *Testing of Density Matrix*

Next, we consider the problem to decide whether the state of the given system is the density matrix ρ_0 or not. In this case, the state ρ_0 is called the null hypothesis, and the set of possible density matrices except for ρ_0 is called the alternative hypothesis and is denoted by S_1. Our decision is "The true density matrix is ρ_0" or "The true density matrix is not ρ_0". In the following, we consider the case when the n-fold tensor product state $\rho^{\otimes n}$ of the unknown state ρ is given.

Hence, we perform a two-valued measurement on the tensor product system $\mathcal{H}^{\otimes n}$. Let M_0 be the POVM element corresponding to the former decision, and M_1 be that corresponding to the latter decision. Since $M_0 + M_1 = I$, we denote our operation by $T_n = M_1$. The probability that we erroneously support the decision M_1 despite the null hypothesis being correct is called the first kind of error probability and is given as $\alpha_n(T_n) := \mathrm{Tr}\, \rho_0^{\otimes n} T_n$. The probability that we erroneously support the decision M_0 despite the null hypothesis being incorrect is called the second kind of error probability and is given as $\beta_n(T_n) := \max_{\rho \in S_1} \mathrm{Tr}\, \rho^{\otimes n}(I - T_n)$. It is known as quantum Stein Lemma [68] that the relative $D(\rho\|\rho_0)$ gives the maximum exponential decreasing rate of the first kind of error probability when the true state is the density matrix $\rho \in S_1$ under the condition that the second kind of error probability goes to zero;

$$\lim_{n\to\infty} \mathrm{Tr}\, \rho^{\otimes n} T_n = 1.$$

In the following, we focus on the minimum relative entropy $\min_{\rho \in S_1} D(\rho\|\rho_0)$ as Fig. 6.3. Given a real number $R \in (0, \min_{\rho \in S_1} D(\rho\|\rho_0))$ and an arbitrary element $\rho \in S_1$, we construct a sequence of tests $\{T_n\}$ such that

$$\lim_{n\to\infty} \frac{-1}{n} \log \mathrm{Tr}\, \rho_0^{\otimes n} T_n = R \tag{6.26}$$

$$\sup_{\rho \in S_1} \lim_{n\to\infty} \mathrm{Tr}\, \rho^{\otimes n} T_n = 1. \tag{6.27}$$

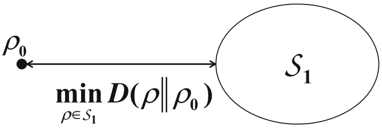


Fig. 6.3 Minimum relative entropy

This sequence of tests has the universality with respect to the independence of the choice of $\rho \in S_1$, which is often called the quantum Sanov Lemma [10].

The above estimation of the eigenvalues deals with the measurement $\{P_{\boldsymbol{n}}\}$. To distinguish the states, we need the information described in the system $\mathcal{U}_{\boldsymbol{n}}$, which is included in the system $\mathcal{W}_{\boldsymbol{n}}$. That is, we need a rank-1 projective measurement commutative to $\mathsf{f}_{\boldsymbol{n}}(\rho)$ on the system $\mathcal{U}_{\boldsymbol{n}}$, which is denoted by $\{P^{\boldsymbol{n}}_{\boldsymbol{l}}\}$. Then, we employ the measurement $\{P_{\boldsymbol{n},\boldsymbol{l}}\}$ on $\mathcal{H}^{\otimes n}$ defined by the projections $P_{\boldsymbol{n},\boldsymbol{l}} := P^{\boldsymbol{n}}_{\boldsymbol{l}} \otimes I_{\mathcal{V}_{\boldsymbol{n}}}$. The probability distribution of the obtained outcome is denoted by $Q[\rho^{\otimes n}, \rho_0^{\otimes n}](\boldsymbol{n}, \boldsymbol{l}) := \operatorname{Tr} P_{\boldsymbol{n},\boldsymbol{l}}\rho^{\otimes n}$. Based on this measurement and the real number R, we define the test T_n as

$$T_n := \sum_{(\boldsymbol{n},\boldsymbol{l})\in\Omega_n} P_{\boldsymbol{n},\boldsymbol{l}} \tag{6.28}$$
$$\Omega_n := \{(\boldsymbol{n},\boldsymbol{l}) | \log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}}\mathsf{f}_{\boldsymbol{n}}(\rho_0) \le -\log\dim \mathcal{V}_{\boldsymbol{n}} - nR\}.$$

Then, we have the following lemma.

Lemma 6.7 *The test T_n defined in (6.28) satisfies the conditions (6.26) and (6.27). Especially, when the set S_1 is compact, the limit* $\lim_{n\to\infty}$ *is commutative with* $\sup_{\rho\in S_1}$ *in (6.27), which implies that*

$$\lim_{n\to\infty} \max_{\rho\in S_1} \operatorname{Tr} \rho^{\otimes n} T_n = 1. \tag{6.29}$$

For the proof of Lemma 6.7, we prepare the following lemma [47].

Lemma 6.8 *Under the probability distribution $Q[\rho^{\otimes n}, \rho_0^{\otimes n}]$, the random variables* $\frac{1}{n}\log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}}\mathsf{f}_{\boldsymbol{n}}(\rho_0)$ *and* $\frac{1}{n}\log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}}\mathsf{f}_{\boldsymbol{n}}(\rho)$ *converge to* $\operatorname{Tr}\rho\log\rho_0$ *and* $\operatorname{Tr}\rho\log\rho$ *in probability, respectively.*

Proof Since $\rho_0^{\otimes n}$ is commutative with $P_{\boldsymbol{n},\boldsymbol{l}}$, we have

$$\begin{aligned}
&\sum_{\boldsymbol{n},\boldsymbol{l}} \operatorname{Tr} P_{\boldsymbol{n},\boldsymbol{l}}\rho^{\otimes n}(\frac{1}{n}\log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}}\mathsf{f}_{\boldsymbol{n}}(\rho_0) - \operatorname{Tr}\rho\log\rho_0)^2 \\
=&\operatorname{Tr}\rho^{\otimes n}(\frac{1}{n}\log\sum_{\boldsymbol{n},\boldsymbol{l}} P_{\boldsymbol{n},\boldsymbol{l}}\rho_0^{\otimes n}P_{\boldsymbol{n},\boldsymbol{l}} - \operatorname{Tr}\rho\log\rho_0)^2
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{Tr} \rho^{\otimes n} (\frac{1}{n} \log \rho_0^{\otimes n} - \operatorname{Tr} \rho \log \rho_0)^2 \\
&= \operatorname{Tr} \rho^{\otimes n} (\frac{1}{n} (\log \rho_0)^{(n)} - \operatorname{Tr} \rho \log \rho_0)^2 = \frac{\Delta_\rho^2 \log \rho_0}{n}. \qquad (6.30)
\end{aligned}
$$

So, the random variable $\frac{1}{n} \log \operatorname{Tr} P_{\boldsymbol{l}}^{\boldsymbol{n}} \mathsf{f}_{\boldsymbol{n}}(\rho_0)$ converges to $\operatorname{Tr} \rho \log \rho_0$ in probability.

On the other hand, similar to the above discussion, Schwarz inequality with respect to the inner product $\operatorname{Tr} \rho^{\otimes n} XY$ on $\mathcal{H}^{\otimes n}$ implies that

$$
\begin{aligned}
&\sum_{\boldsymbol{n},\boldsymbol{l}} \operatorname{Tr} P_{\boldsymbol{n},\boldsymbol{l}} \rho^{\otimes n} (\frac{1}{n} \log \operatorname{Tr} P_{\boldsymbol{l}}^{\boldsymbol{n}} \mathsf{f}_{\boldsymbol{n}}(\rho) - \operatorname{Tr} \rho \log \rho)^2 \\
=& \operatorname{Tr} \rho^{\otimes n} (\frac{1}{n} \log \sum_{\boldsymbol{n},\boldsymbol{l}} P_{\boldsymbol{n},\boldsymbol{l}} \rho^{\otimes n} P_{\boldsymbol{n},\boldsymbol{l}} - \operatorname{Tr} \rho \log \rho)^2 \\
\le& 2 \operatorname{Tr} \rho^{\otimes n} (\frac{1}{n} \log \sum_{\boldsymbol{n},\boldsymbol{l}} P_{\boldsymbol{n},\boldsymbol{l}} \rho^{\otimes n} P_{\boldsymbol{n},\boldsymbol{l}} - \frac{1}{n} \log \rho^{\otimes n})^2 \\
&+ 2 \operatorname{Tr} P_{\boldsymbol{n},\boldsymbol{l}} \rho^{\otimes n} P_{\boldsymbol{n},\boldsymbol{l}} (\frac{1}{n} \log \rho^{\otimes n} - \operatorname{Tr} \rho \log \rho)^2 \\
=& \frac{2}{n^2} \sum_{\boldsymbol{n}} Q[\rho^{\otimes n}](\boldsymbol{n}) \operatorname{Tr} \mathsf{f}_{\boldsymbol{n}}(\rho) (\log \sum_{\boldsymbol{l}} P_{\boldsymbol{l}}^{\boldsymbol{n}} \mathsf{f}_{\boldsymbol{n}}(\rho) P_{\boldsymbol{l}}^{\boldsymbol{n}} - \log \mathsf{f}_{\boldsymbol{n}}(\rho))^2 + \frac{2\Delta_\rho^2 \log \rho}{n} \\
\le& \frac{4}{n^2} \max_{\boldsymbol{n} \in \mathcal{Y}_n^r} (\log \dim \mathcal{U}_{\boldsymbol{n}})^2 + \frac{2\Delta_\rho^2 \log \rho}{n} \qquad (6.31) \\
\le& \frac{r^2 (r-1)^2 (\log(n+1))^2}{n^2} + \frac{2\Delta_\rho^2 \log \rho}{n}, \qquad (6.32)
\end{aligned}
$$

where the inequalities (6.31) and (6.32) follow from (2.29) and (6.16), respectively. Since the RHS of (6.31) converges to zero, under the probability distribution $Q[\rho^{\otimes n}, \rho_0^{\otimes n}]$, the random variable $\frac{1}{n} \log \operatorname{Tr} P_{\boldsymbol{l}}^{\boldsymbol{n}} \mathsf{f}_{\boldsymbol{n}}(\rho)$ converges to $\operatorname{Tr} \rho \log \rho$ in probability. ∎

Proof of Lemma 6.7 The definition of Ω_n yields the inequality $\operatorname{Tr} \rho_0^{\otimes n} P_{\boldsymbol{n},\boldsymbol{l}} \le e^{-nR}$ for $(\boldsymbol{n}, \boldsymbol{l}) \in \Omega_n$. Hence, using (6.16) and (6.18), we have

$$
\begin{aligned}
\operatorname{Tr} \rho^{\otimes n} T_n =& \sum_{(\boldsymbol{n},\boldsymbol{l}) \in \Omega_n} \operatorname{Tr} \rho_0^{\otimes n} P_{\boldsymbol{n},\boldsymbol{l}} \le \sum_{(\boldsymbol{n},\boldsymbol{l}) \in \Omega_n} e^{-nR} \\
\le& |\mathcal{Y}_n^r| \cdot \max_{\boldsymbol{n} \in \mathcal{Y}_n^r} \dim \mathcal{U}_{\boldsymbol{n}} \cdot e^{-nR} \le (n+1)^{r(r-1)/2+r-1} e^{-nR}.
\end{aligned}
$$

Hence, we obtain (6.26).

For $\rho \in S_1$, we define $\Omega[\rho]_n$ as

$$
\Omega[\rho]_n := \{(\boldsymbol{n}, \boldsymbol{l}) | \log \operatorname{Tr} P_{\boldsymbol{l}}^{\boldsymbol{n}} \mathsf{f}_{\boldsymbol{n}}(\rho_0) - \log \operatorname{Tr} P_{\boldsymbol{l}}^{\boldsymbol{n}} \mathsf{f}_{\boldsymbol{n}}(\rho) \le -nR\}.
$$

Since $\dim \mathcal{V}_{\boldsymbol{n}} \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}} \mathsf{f}_{\boldsymbol{n}}(\rho) \le 1$, the condition for Ω_n holds when the condition for $\Omega[\rho]_n$ holds. That is, $\Omega[\rho]_n \subset \Omega_n$. Hence, it is sufficient to show that

$$\sum_{(\boldsymbol{n},\boldsymbol{l}) \in \Omega[\rho]_n} \operatorname{Tr} \rho^{\otimes n} P_{\boldsymbol{n},\boldsymbol{l}} \to 1. \tag{6.33}$$

Since Lemma 6.8 guarantees that the random variable $\log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}} \mathsf{f}_{\boldsymbol{n}}(\rho_0) - \log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}} \mathsf{f}_{\boldsymbol{n}}(\rho)$ converges to $D(\rho\|\rho_0)$ in probability, the relation (6.33) holds. So, we obtain (6.27).

In particular, when the set S_1 is compact, since $R_0 := \max_{\rho \in S_1} D(\rho\|\rho_0) < R$, Markov inequality (Lemma 2.6) guarantees that

$$\begin{aligned} & Q[\rho^{\otimes n}, \rho_0^{\otimes n}](\Omega[\rho]_n{}^c) \\ \le & \frac{\mathrm{E}_{Q[\rho^{\otimes n}, \rho_0^{\otimes n}]}\big[(\frac{1}{n} \log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}} \mathsf{f}_{\boldsymbol{n}}(\rho_0) - \frac{1}{n} \log \operatorname{Tr} P^{\boldsymbol{n}}_{\boldsymbol{l}} \mathsf{f}_{\boldsymbol{n}}(\rho) - D(\rho\|\rho_0))^2\big]}{(R - D(\rho\|\rho_0))^2} \\ \le & \frac{\frac{2\Delta_\rho^2 \log \rho_0}{n} + \frac{2r^2(r-1)^2(\log(n+1))^2}{n^2} + \frac{4\Delta_\rho^2 \log \rho}{n}}{(R - R_0)^2}, \end{aligned} \tag{6.34}$$

where the final inequality follows from Lemma 2.7 and the inequalities (6.30) and (6.32) in the proof of Lemma 6.8. Since the relation (6.34) holds independently of $\rho \in S_1$, we obtain (6.29). ■

6.2.4 Dimension of Irreducible Space for Permutation Group

Next, we focus on the dimension of the irreducible representation space $\mathcal{V}_{\boldsymbol{n}}$ with respect to the permutation group. Due to the relations (6.20), (6.21), and (6.16), the random variables $\frac{1}{n} \log(\dim \mathcal{V}_{\boldsymbol{n}})$ and $\frac{1}{n} \log(\operatorname{rank} P_{\boldsymbol{n}})$ coincide to the entropy $H(\frac{\boldsymbol{n}}{n})$ under the limit $n \to \infty$. Based on this property, we address the behaviors of these random variables under the probability distribution $Q[\rho^{\otimes n}]$. Due to Lemma 6.5, the above random variables, $H(\frac{\boldsymbol{n}}{n})$ etc., converge to $H(\boldsymbol{p})$ in probability. Especially, when ϵ_n satisfies $\sqrt{n}\epsilon_n \to \infty$, Lemma 6.6 guarantees that the probability that $|H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p})| \le \epsilon_n$ goes to 1. Hence, the expectations of these random variables converge to $H(\boldsymbol{p})$. However, the speeds of the convergences of these expectations depend on the random variable. The asymptotic behavior of the expectation of $\log(\dim \mathcal{V}_{\boldsymbol{n}})$ is given as the following lemma [93].

Lemma 6.9 *The asymptotic behavior of the expectation of* $\log(\dim \mathcal{V}_{\boldsymbol{n}})$ *under the probability distribution* $Q[\rho^{\otimes n}]$ *is given as*

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{E}_{Q[\rho^{\otimes n}]}\big[\log(\dim \mathcal{V}_{\boldsymbol{n}})\big] - nH(\boldsymbol{p}) + \frac{r-1}{2}\log n \\
=&-\frac{r-1}{2}\log 2\pi e - \frac{1}{2}\sum_{j=1}^{r}\log p_j + \log \prod_{i,j:i<j}(p_i - p_j) \\
&- \frac{\sum_{\sigma\in S_r}\operatorname{sgn}(\sigma)\prod_i p_i^{\delta_{\sigma(i)}}\log\prod_j p_j^{\delta_{\sigma(k)}}}{\prod_{i,j:i<j}(p_i-p_j)}.
\end{aligned} \tag{6.35}$$

Proof Firstly, we calculate the expectation of the random variable $\log(\dim \mathcal{V}_{\boldsymbol{n}})$ under the probability distribution $\mathrm{Mul}[\boldsymbol{p}, n, \sigma](\boldsymbol{n}) := \prod_{l=1}^{r} p_l^{n_l^\sigma}\frac{n!}{\boldsymbol{n}^\sigma!}$, which is given in the proof of Lemma 6.5 dependently of $\sigma \in S_r$.

$$\begin{aligned}
\log(\dim \mathcal{V}_{\boldsymbol{n}}) =& \log\left[\sum_{\sigma'\in S_r}\operatorname{sgn}(\sigma')\frac{n!}{[\boldsymbol{n}+\boldsymbol{\delta}-\sigma'(\boldsymbol{\delta})]!}\right] \\
=&\log\left[\sum_{\sigma'\in S_r}\operatorname{sgn}(\sigma')\frac{n!}{[\boldsymbol{n}^\sigma+\sigma()-\sigma'(\boldsymbol{\delta})]!}\right] \\
=&\log\frac{n!}{\boldsymbol{n}^\sigma!} + \log\sum_{\sigma'\in S_r}\operatorname{sgn}(\sigma')\frac{\boldsymbol{n}^\sigma!}{[\boldsymbol{n}^\sigma+\sigma()-\sigma'(\boldsymbol{\delta})]!}.
\end{aligned} \tag{6.36}$$

Then, the expectation of the first term of (6.36) has been already given in Lemma 6.4. On the other hand, the second term of (6.36) converges to the following value in probability under the probability distribution $\prod_{l=1}^{r} p_l^{n_l^\sigma}\frac{n!}{\boldsymbol{n}^\sigma!}$.

$$\begin{aligned}
&\log\sum_{\sigma'\in S_r,}\operatorname{sgn}(\sigma')\frac{\boldsymbol{n}^\sigma!}{[\boldsymbol{n}^\sigma+\sigma()-\sigma'(\boldsymbol{\delta})]!} \\
=&\log\sum_{\sigma'\in S_r}\operatorname{sgn}(\sigma')\frac{\prod_{i:\delta_{\sigma(i)}-\delta_{\sigma'(i)}<0}\prod_{j=1}^{\delta_{\sigma(i)}-\delta_{\sigma'(i)}}(n_i^\sigma - j + 1)}{\prod_{i:\delta_{\sigma(i)}-\delta_{\sigma'(i)}>0}\prod_{j=1}^{\delta_{\sigma'(i)}-\delta_{\sigma(i)}}(n_i^\sigma + j)} \\
=&\log\sum_{\sigma'\in S_r}\operatorname{sgn}(\sigma')\frac{\prod_{i:\delta_{\sigma(i)}-\delta_{\sigma'(i)}<0}\prod_{j=1}^{\delta_{\sigma(i)}-\delta_{\sigma'(i)}}\frac{n_i^\sigma-j+1}{n}}{\prod_{i:\delta_{\sigma(i)}-\delta_{\sigma'(i)}>0}\prod_{j=1}^{\delta_{\sigma'(i)}-\delta_{\sigma(i)}}\frac{n_i^\sigma+j}{n}} \\
\to&\log\sum_{\sigma'\in S_r}\operatorname{sgn}(\sigma')\frac{\prod_{i:\delta_{\sigma(i)}-\delta_{\sigma'(i)}<0}p_i^{\delta_{\sigma'(i)}-\delta_{\sigma(i)}}}{\prod_{i:\delta_{\sigma(i)}-\delta_{\sigma'(i)}>0}p_i^{\delta_{\sigma(i)}-\delta_{\sigma'(i)}}} \\
=&\log\sum_{\sigma'\in S_r}\operatorname{sgn}(\sigma')\prod_i p_i^{\delta_{\sigma'(i)}-\delta_{\sigma(i)}} = \log\prod_{i,j:i<j}(p_i-p_j)\prod_j p_j^{-\delta_{\sigma(j)}} \\
=&\log\prod_{i,j:i<j}(p_i-p_j) - \log\prod_j p_j^{\delta_{\sigma(j)}},
\end{aligned}$$

where we employed Vandermonde determinant. Summarizing the above discussions, we have

$$\begin{aligned} & \mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n,\sigma]}\Big[\log(\dim \mathcal{V}_{\boldsymbol{n}}) - nH(\boldsymbol{p}) + \frac{r-1}{2}\log n\Big] \\ \to & -\frac{r-1}{2}\log 2\pi e - \frac{1}{2}\sum_{j=1}^{r}\log p_j + \log\prod_{i,j:i<j}(p_i - p_j) - \log\prod_{j} p_j^{\delta_{\sigma(j)}}. \end{aligned} \tag{6.37}$$

Further, due to the discussion in the proof of Lemma 6.5, the LHS of (6.35) equals the weighted sum of the expectation of (6.37) under the probability distribution $\prod_{l=1}^{r} p_l^{n_l^\sigma} \frac{n!}{\boldsymbol{n}^\sigma !}$ with the weight coefficient $\frac{\mathrm{sgn}(\sigma)\prod_i p_i^{\delta_{\sigma(i)}}}{\prod_{i,j:i<j}(p_i - p_j)}$. Since (6.37) does not depend on σ except for the final term, using (6.25), we obtain (6.35). ∎

6.3 Estimation of State in Qubit System

Section 6.2.2 has addressed the estimation only of the eigenvalues of the density matrix. In the following, in the case of $\mathbb{C}^2$, we deal with the problem to estimate the unitary direction of the density matrix as well as the eigenvalues to decide the complete form of the density matrix. Hence, the parameter space is the set of density matrix on $\mathbb{C}^2$. Now, we focus on the following estimation procedure. Firstly, we perform the measurement $\{P_{\boldsymbol{n}}\}$ given in Sect. 6.2.2 and estimate the unitary direction on each irreducible component. Letting $\boldsymbol{n}$ be the measurement outcome of the first step, we perform a measurement on the system $\mathcal{W}_{\boldsymbol{n}}$ in the second step, in which the state is $\mathsf{f}_{\boldsymbol{n}}(\rho) \otimes I_{\mathcal{V}_{\boldsymbol{n}}}$. In the second step, since the system $\mathcal{V}_{\boldsymbol{n}}$ has no information, the partial trace with respect to $\mathcal{V}_{\boldsymbol{n}}$ does not cause any information loss. Hence, the problem in the second step is to estimate the density matrix $\mathsf{f}_{\boldsymbol{n}}(\rho)$ on the system $\mathcal{U}_{\boldsymbol{n}}$. Since the eigenvalues are almost estimated in the first step, we estimate only the unitary direction of the density matrix $\mathsf{f}_{\boldsymbol{n}}(\rho)$ in the second step, which is the same problem as the problem discussed in Example 4.3 in Sect. 4.2. Hence, we choose the POVM generated by the highest weight vector $|\frac{n_1-n_2}{2}, \frac{n_1-n_2}{2}\rangle$ as the measurement in the second step. That is, we apply the measurement $M_{\boldsymbol{n},\zeta} := \mathsf{f}_{\boldsymbol{n}}(g(\zeta))|\frac{n_1-n_2}{2}; \frac{n_1-n_2}{2}\rangle\langle\frac{n_1-n_2}{2}; \frac{n_1-n_2}{2}|\mathsf{f}_{\boldsymbol{n}}(g(\zeta))^\dagger \otimes I_{\mathcal{V}_{\boldsymbol{n}}} = |\frac{n_1-n_2}{2} : \zeta\rangle\langle\frac{n_1-n_2}{2} : \zeta| \otimes I_{\mathcal{V}_{\boldsymbol{n}}}$ and output the density matrix $g(\zeta)\rho(\frac{\boldsymbol{n}}{n})g(\zeta)^\dagger$ as our estimate, where $\mathrm{SU}(2)/H$ is the quotient space of $\mathrm{SU}(2)$ divided by the stabilizer H stabilizing $\rho(\boldsymbol{n})$ and $g(\zeta)$ is a representative of $\zeta \in \mathrm{SU}(2)/H$.

Hence, the fidelity $F(\rho, g(\zeta)\rho(\frac{\boldsymbol{n}}{n})g(\zeta)^\dagger)$ between the true state ρ and the estimate $g(\zeta)\rho(\frac{\boldsymbol{n}}{n})g(\zeta)^\dagger$ is given in the following calculation. In the following, we discuss the case when the density matrix ρ is a diagonal matrix and choose r, t, c_1 such that $\rho = \rho_{(r,0,0)} = c_1 e^{tE_{0,1}}$. By letting $c(\lambda) = \frac{e^t - 1}{e^{t(\lambda+1)} - e^{-t\lambda}}$, the relation (4.30) implies that

$$
\begin{aligned}
&\sum_{\boldsymbol{n}\in\mathcal{Y}_n^2}\int_{S^2} F(\rho, g(\zeta)\rho(\frac{\boldsymbol{n}}{n})g(\zeta)^\dagger)^2 \operatorname{Tr}\rho^{\otimes n} M_{\boldsymbol{n},\theta}\mu_{S^2}(d\zeta)\\
=&\sum_{k=0}^{\lfloor n/2\rfloor} Q[\rho^{\otimes n}](\frac{n-k}{n},\frac{k}{n})\\
&\cdot\int_{S^2} F(\rho, g(\zeta)\rho(\frac{n-k}{n},\frac{k}{n})g(\zeta)^\dagger)^2 c(\frac{n-2k}{2})\\
&\cdot\left\langle\frac{n-2k}{2}:\zeta\middle| e^{t\mathsf{f}_{\frac{n-2k}{2}}(E_{0,1})}\middle|\frac{n-2k}{2}:\zeta\right\rangle\mu_{S^2}(d\zeta)\\
=&\sum_{k=0}^{\lfloor n/2\rfloor} Q[\rho^{\otimes n}](\frac{n-k}{n},\frac{k}{n})\\
&\cdot\left(\frac{1}{2}+\frac{n-2k}{n}\left(\frac{(n-2k+1)((1+r)^{n-2k+2}-(1-r)^{n-2k+2})}{2(n-2k+2)((1+r)^{n-2k+1}-(1-r)^{n-2k+1})}-\frac{1}{2}\right)\right.\\
&\left.+\sqrt{1-(\frac{n-2k}{n})^2}\frac{\sqrt{1-r^2}}{2}\right).
\end{aligned}
\tag{6.38}
$$

Using Lemma 6.6, we choose the constant K such that the expectation of $\frac{n-2k}{n}$ is $r+\frac{K}{n}+o(\frac{1}{n})$. Then, the expectation of $\frac{1}{2}+\frac{n-2k}{n}\left(\frac{(n-2k+1)((1+r)^{n-2k+2}-(1-r)^{n-2k+2})}{2(n-2k+2)((1+r)^{n-2k+1}-(1-r)^{n-2k+1})}-\frac{1}{2}\right)$ is $\frac{1-r^2}{2}+\frac{r}{2}\frac{K}{n}-\frac{1+r}{2n}+o(\frac{1}{n})$. On the other hand, letting $\Delta(r):=\frac{n-2k}{n}-r$, we have

$$
\sqrt{1-(\frac{n-2k}{n})^2}\frac{\sqrt{1-r^2}}{2}=\frac{1-r^2}{2}-\frac{r}{2}\Delta(r)-\frac{1}{4(1-r^2)}\Delta(r)^2+o(\Delta(r)^2).
\tag{6.39}
$$

Since the expectation of $\Delta(r)^2$ is $\frac{1-r^2}{n}$, the expectation of (6.39) is $\frac{1-r^2}{2}-\frac{r}{2}\frac{K}{n}-\frac{1}{4n}+o(\frac{1}{n})$. That is, (6.38) is calculated to $1-\frac{3+2r}{4n}+o(\frac{1}{n})$ [2, 49].

6.4 Universal Approximation of Quantum State

Now, we consider the universal approximation of quantum state with respect to a function $f(\rho_1,\rho_2)$ to describe the degree of the difference between two states ρ_1 and ρ_2 given in Sect. 2.4 when $f(\rho_1,\rho_2)$ satisfies the additivity with respect to the tensor product. This concept is a key to understand the universal information processing. In the following, under the function $f(\rho_1,\rho_2)$, we discuss the approximation of quantum state for a set $\mathcal{S}$ of density matrices. When a sequence $\{\rho_n\}$ of density matrices on $\mathcal{H}^{\otimes n}$ satisfies the following condition, the sequence $\{\rho_n\}$ is called a **universal approximation** for $\mathcal{S}$ with respect to f.

$$\lim_{n\to\infty}\frac{1}{n}f(\rho^{\otimes n},\rho_n)=0,\quad \forall\rho\in\mathcal{S}.$$

In fact, when the function f satisfies the axiom of the distance or given as a strictly monotone increasing function of a function satisfying the axiom of the distance, we can show that there is no universal approximation (Exercise 6.6). This section shows an example of universal approximation, which is a preparation for the latter section.

For example, when $\mathcal{H}=\mathbb{C}^r$, defining $\rho_{\boldsymbol{n}}:=\frac{1}{\dim\mathcal{W}_{\boldsymbol{n}}}P(\mathcal{W}_{\boldsymbol{n}})$ and $\rho_{U,n}:=\sum_{\boldsymbol{n}\in\mathcal{Y}_n^r}\frac{1}{|\mathcal{Y}_n^r|}\rho_{\boldsymbol{n}}$, we have

$$[\rho_{U,n}\otimes\rho_{U,m},\rho_{U,n+m}]=0, \tag{6.40}$$

because the subgroup $S_n\times S_M$ of S_{n+m} has an irreducible representation that is a direct sum decomposition of an irreducible representation of S_{n+m}. Then, the following theorem holds [55].

Theorem 6.1 *The sequence $\{\rho_{U,n}\}$ of states satisfies*

$$D_{\max}(\rho^{\otimes n}\|\rho_{U,n})\le\frac{(r+2)(r-1)}{2}\log(n+1). \tag{6.41}$$

Hence, it is a universal approximation for $\mathcal{S}(\mathcal{H})$ with respect to relative max-entropy.

Proof Since $\frac{\dim\mathcal{U}_{\boldsymbol{n}}}{\dim\mathcal{W}_{\boldsymbol{n}}}P(\mathcal{W}_{\boldsymbol{n}})\ge P(\mathcal{W}_{\boldsymbol{n}})\rho^{\otimes n}P(\mathcal{W}_{\boldsymbol{n}})$, we have $|\mathcal{Y}_n^r|\max_{\boldsymbol{n}\in\mathcal{Y}_n^r}\dim\mathcal{U}_{\boldsymbol{n}}\rho_{U,n}\ge\rho^{\otimes n}$. Since the relation (6.16) and (6.18) yields the inequality $|\mathcal{Y}_n^r|\max_{\boldsymbol{n}\in\mathcal{Y}_n^r}\dim\mathcal{U}_{\boldsymbol{n}}\le(n+1)^{(r+2)(r-1)/2}$, we obtain (6.41). ■

Further, we have the following theorem for relative entropy [54].

Theorem 6.2 *When $\boldsymbol{p}$ is the probability distribution composed of the eigenvalues of ρ, the sequence $\{\rho_{U,n}\}$ of states satisfies*

$$D(\rho^{\otimes n}\|\rho_{U,n})=\frac{r^2-1}{2}\log n+C_r-\log r!(r-1)!+C(\boldsymbol{p})+o(1), \tag{6.42}$$

where C_r and $C(\boldsymbol{p})$ are defined as

$$\begin{aligned}C_r:=&-\frac{r-1}{2}\log 2\pi e-\sum_{j=1}^{r}(r-j)\log j\\ C(\boldsymbol{p}):=&2\log\prod_{i<j}(p_i-p_j)-\frac{1}{2}\sum_j\log p_j\\ &-\sum_{\sigma\in S_r}\frac{\operatorname{sgn}(\sigma)(\prod_j p_j^{\delta_{\sigma(j)}})\log(\prod_j p_j^{\delta_{\sigma(j)}})}{\prod_{i<j}(p_i-p_j)}.\end{aligned}$$

Proof First, we notice that

$$\begin{aligned} & D(\rho^{\otimes n} \| \rho_{U,n}) = -\operatorname{Tr} \rho^{\otimes n} \log \rho_{U,n} - nH(\boldsymbol{p}) \\ = & \sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} Q[\rho^{\otimes n}](\boldsymbol{n})(\log |\mathcal{Y}_n^r| + \log \dim \mathcal{U}_{\boldsymbol{n}} + \log \dim \mathcal{V}_{\boldsymbol{n}}) - nH(\boldsymbol{p}). \end{aligned} \tag{6.43}$$

The relation (6.17) implies

$$\begin{aligned} & \sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} Q[\rho^{\otimes n}](\boldsymbol{n}) \log \dim \mathcal{V}_{\boldsymbol{n}} \\ = & (r-1) \log n + \sum_{j<l} \log(p_l - p_j) - \sum_{j=1}^{r} (r-j) \log j + o(1). \end{aligned} \tag{6.44}$$

Hence, together with (6.43) and (6.44), the relations (6.19) and (6.35) imply (6.42). ■

Now, we optimize our sequence of states in the sense of minmax of the universal approximation with respect to relative entropy. That is, the optimization of the worst case of the constant term of the RHS of (6.42) can be formulated as

$$D_{\min-\max}(r) := \min_{\{\sigma_n\}} \sup_{\rho} \lim_{n \to \infty} D(\rho^{\otimes n} \| \sigma_n) - \frac{r^2 - 1}{2} \log n. \tag{6.45}$$

The following theorem holds for this quantity [54].

Theorem 6.3 *The minimum $D_{\min-\max}(r)$ is realized when σ_n is $\sigma_{J,n} := \sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} \frac{J_n(\boldsymbol{n})}{\dim \mathcal{W}_{\boldsymbol{n}}} P(\mathcal{W}_{\boldsymbol{n}})$ or $\tilde{\sigma}_{J,n} := \int_{\mathcal{Y}(r)} \int_{\mathrm{SU}(r)} (U \rho(\boldsymbol{p}) U^{\dagger})^{\otimes n} \mu(dU) J(\boldsymbol{p}) d\boldsymbol{p}$. This value is calculated as $D_{\min-\max}(r) = C_r + \log \int_{\mathcal{Y}_r} e^{C(\boldsymbol{p})} d\boldsymbol{p}$, where $J_n(\boldsymbol{n}) := \frac{e^{C(\frac{\boldsymbol{n}}{n})}}{\sum_{\boldsymbol{n}' \in \mathcal{Y}_n^r} e^{C(\frac{\boldsymbol{n}'}{n})}}$ and $J(\boldsymbol{p}) := \frac{e^{C(\boldsymbol{p})}}{\int_{\mathcal{Y}(r)} e^{C(\boldsymbol{p})} d\boldsymbol{p}}$.*

In the classical system, a solution ρ_n of the above mini-max problem is given as the mixture of the n-fold independent and identical distribution of the parametrized distribution with respect to the Jeffreys prior distribution. Hence, the above probability distribution $\mu(dU) J(\boldsymbol{p}) d\boldsymbol{p}$ on the set of density matrices can be regarded as the quantum analogue of Jeffreys prior distribution.

Proof Since the density matrix $\rho^{\otimes n}$ is invariant for the representation f_n of the permutation group S_n given in [44, (4.61)]. For any element $g \in S_n$, $D(\rho^{\otimes n} \| \sigma_n) = -\operatorname{Tr}[\rho^{\otimes n} \log \sigma_n] - nH(\rho)$ equals $-\operatorname{Tr}[\rho^{\otimes n} \log \mathsf{f}_n(g) \sigma_n \mathsf{f}_n(g)^{\dagger}] - nH(\rho)$ Since the function $x \mapsto -\log x$ is matrix convex, we have

$$
\begin{aligned}
& - \mathrm{Tr}[\rho^{\otimes n} \log \sigma_n] - nH(\rho) \\
= & \sum_{g \in S_n} \frac{1}{|S_n|} \mathrm{Tr}[\rho^{\otimes n}(-\log \mathsf{f}_n(g)\sigma_n \mathsf{f}_n(g)^\dagger)] - nH(\rho) \\
\geq & - \mathrm{Tr}[\rho^{\otimes n} \log \sum_{g \in S_n} \frac{1}{|S_n|} \mathsf{f}_n(g)\sigma_n \mathsf{f}_n(g)^\dagger] - nH(\rho).
\end{aligned}
$$

Since the density matrix $\sum_{g \in S_n} \frac{1}{|S_n|} \mathsf{f}_n(g)\sigma_n \mathsf{f}_n(g)^\dagger$ is invariant for the representation of the permutation group S_n, we can restrict our density matrix σ_n in (6.45) to a density matrix invariant for the representation of the permutation group S_n.

Due to the matrix concavity of the function $x \mapsto -\log x$, the invariant measure μ on $\mathrm{SU}(r)$ satisfies

$$
\begin{aligned}
& \sup_{\rho} \overline{\lim_{n\to\infty}} \left(D(\rho^{\otimes n} \| \sigma_n) - \frac{r^2-1}{2} \log n \right) \\
= & \sup_{\boldsymbol{p}} \sup_{U \in \mathrm{SU}(r)} \overline{\lim_{n\to\infty}} -\mathrm{Tr}\, \rho(\boldsymbol{p})^{\otimes n} \log \Big[U^{\otimes n} \sigma_n U^{\otimes n \dagger} \Big] - nH(\boldsymbol{p}) - \frac{r^2-1}{2} \log n \\
\geq & \sup_{\boldsymbol{p}} \overline{\lim_{n\to\infty}} \Bigg[- \int_{\mathrm{SU}(r)} \mathrm{Tr}\, \rho(\boldsymbol{p})^{\otimes n} \log \Big[U^{\otimes n} \sigma_n U^{\otimes n \dagger} \Big] \mu(dU) \\
& \qquad - nH(\boldsymbol{p}) - \frac{r^2-1}{2} \log n \Bigg] \\
= & \sup_{\boldsymbol{p}} \overline{\lim_{n\to\infty}} \Bigg[- \mathrm{Tr}\, \rho(\boldsymbol{p})^{\otimes n} \Big[\int_{\mathrm{SU}(r)} \log \Big[U^{\otimes n} \sigma_n U^{\otimes n \dagger} \Big] \mu(dU) \Big] \\
& \qquad - nH(\boldsymbol{p}) - \frac{r^2-1}{2} \log n \Bigg] \\
\geq & \sup_{\boldsymbol{p}} \lim_{n\to\infty} \Bigg[- \mathrm{Tr}\, \rho(\boldsymbol{p})^{\otimes n} \log \Big[\int_{\mathrm{SU}(r)} U^{\otimes n} \sigma_n U^{\otimes n \dagger} \mu(dU) \Big] \\
& \qquad - nH(\boldsymbol{p}) - \frac{r^2-1}{2} \log n \Bigg].
\end{aligned}
$$

Since the density matrix $\left(\int_{\mathrm{SU}(r)} U^{\otimes n} \sigma_n U^{\otimes n \dagger} \mu(dU) \right)$ is invariant for the representation of $\mathrm{SU}(r)$, without loss of generality, we can restrict the density matrix σ_n in (6.45) to a density matrix invariant for the representation of $\mathrm{SU}(r)$. Hence, we have

$$
\begin{aligned}
& \inf_{\{\sigma_n\}} \sup_{\rho} \overline{\lim_{n\to\infty}} \left(D(\rho^{\otimes n} \| \sigma_n) - \frac{r^2-1}{2} \log n \right) \\
= & \inf_{\{P_n\}} \sup_{\rho} \overline{\lim_{n\to\infty}} \left(D(\rho^{\otimes n} \| \sigma_{P_n,n}) - \frac{r^2-1}{2} \log n \right),
\end{aligned}
$$

where P_n is a probability distribution on $\mathcal{Y}_n^r$ and $\sigma_{P_n,n}$ is defined as $\sigma_{P_n,n} := \sum_{\boldsymbol{n}\in\mathcal{Y}_n^r} P_n(\boldsymbol{n})\rho_{\boldsymbol{n}}$. Since $\frac{\sum_{\boldsymbol{n}'\in\mathcal{Y}_n^r} e^{C(\frac{\boldsymbol{n}'}{n})}}{n^{r-1}} \to \int_{\mathcal{Y}(r)} e^{C(\boldsymbol{p})}d\boldsymbol{p}$, we obtain

$$\begin{aligned}
&D(\rho(\boldsymbol{p})^{\otimes n}\|\sigma_{P_n,n})\\
\cong&\frac{r^2-1}{2}\log n + C_r - \log r!(r-1)! + C(\boldsymbol{p}) - \log\frac{P_n(\boldsymbol{n})}{\frac{1}{|\mathcal{Y}_n^r|}}\\
\cong&\frac{r^2-1}{2}\log n + C_r - \log r!(r-1)! + \log(\int_{\mathcal{Y}(r)} e^{C(\boldsymbol{p}')}d\boldsymbol{p}')\\
&+\log J(\boldsymbol{p}) - \sum_{\boldsymbol{n}} Q_{\boldsymbol{p}}(\boldsymbol{n})\log P_n(\boldsymbol{n})\frac{n^{r-1}}{r!(r-1)!}\\
\cong&\frac{r^2-1}{2}\log n + C_r + \log\int_{\mathcal{Y}(r)} e^{C(\boldsymbol{p}')}d\boldsymbol{p}'\\
&+\sum_{\boldsymbol{n}} Q_{\boldsymbol{p}}(\boldsymbol{n})(\log J(\boldsymbol{p}) - \log P_n(\boldsymbol{n}) - (r-1)\log n).
\end{aligned}\tag{6.46}$$

Here, only the final in (6.46) depends on $\boldsymbol{p}$. To evaluate the final term, we consider the joint distribution $Q_{J,n}(\boldsymbol{p},\boldsymbol{n}) := J(\boldsymbol{p})Q_{\boldsymbol{p}}(\boldsymbol{n})$.

Since Lemma 6.5 guarantees that the random variable $\frac{\boldsymbol{n}}{n}$ converges to $\boldsymbol{p}$ in probability under the probability distribution $Q_{\boldsymbol{p}}$ and that $J(\boldsymbol{p})$ is continuous for $\boldsymbol{p}$, the marginal distribution $Q_{J,n}(\boldsymbol{n})(=\int_{\mathcal{Y}(r)} Q_{J,n}(\boldsymbol{p},\boldsymbol{n})d\boldsymbol{p})$ approaches to $J_n(\boldsymbol{n})$. Since the random variable $\boldsymbol{p}$ converges to $\frac{\boldsymbol{n}}{n}$ in probability under the conditional distribution $Q_{J,n}(\boldsymbol{p}|\boldsymbol{n})$, we obtain

$$\begin{aligned}
&\int_{\mathcal{Y}(r)} \log J(\boldsymbol{p})\frac{Q_{\boldsymbol{p}}(\boldsymbol{n})J_n(\boldsymbol{p})}{\int_{\mathcal{Y}(r)} J_n(\boldsymbol{p})\mathrm{P}_{\boldsymbol{p}}(\boldsymbol{n})d\boldsymbol{p}}d\boldsymbol{p} \cong \log J(\frac{\boldsymbol{n}}{n})\\
=&\log J_n(\boldsymbol{n}) + \log\frac{\sum_{\boldsymbol{n}'\in\mathcal{Y}_n^r} e^{C(\frac{\boldsymbol{n}'}{n})}}{\int_{\mathcal{Y}(r)} e^{C(\boldsymbol{p})}d\boldsymbol{p}} \cong \log J_n(\boldsymbol{n}) + (r-1)\log n.
\end{aligned}\tag{6.47}$$

On the other hand, since the random variable $\frac{\boldsymbol{n}}{n}$ converges to $\boldsymbol{p}$ in probability under the probability distribution $Q_{\boldsymbol{p}}(\boldsymbol{n})$, the random variable $\log J_n(\boldsymbol{n}) - (r-1)\log n$ converges to $\log J(\boldsymbol{p})$ in probability under the same probability distribution $Q_{\boldsymbol{p}}(\boldsymbol{n})$. That is, for any distribution $\boldsymbol{p}$, we have

$$\sum_{\boldsymbol{n}} Q_{\boldsymbol{p}}(\boldsymbol{n})(\log J(\boldsymbol{p}) - \log J_n(\boldsymbol{n}) - (r-1)\log n) \to 0. \tag{6.48}$$

Using (6.47), we maximize the final term of (6.46) with respect to $\boldsymbol{p}$. Then, we have

$$\begin{aligned}
&\sup_{\boldsymbol{p}} \sum_{\boldsymbol{n}} Q_{\boldsymbol{p}}(\boldsymbol{n})(\log J(\boldsymbol{p}) - \log P_n(\boldsymbol{n}) - (r-1)\log n) \\
\geq &\int_{\mathcal{Y}(r)} \sum_{\boldsymbol{n}} Q_{\boldsymbol{p}}(\boldsymbol{n})(\log J(\boldsymbol{p}) - \log P_n(\boldsymbol{n}) - (r-1)\log n) J(\boldsymbol{p}) d\boldsymbol{p} \\
\cong &\sum_{\boldsymbol{n}} J_n(\boldsymbol{n})(\log J_n(\boldsymbol{n}) - \log P_n(\boldsymbol{n})) = D(J_n \| P_n) \geq 0,
\end{aligned}$$

which implies that

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{\boldsymbol{n}} P_n(\boldsymbol{n})(D(\rho(\frac{\boldsymbol{n}}{n})^{\otimes n} \| \sigma_{P_n,n}) - \frac{r^2-1}{2}\log n) \\
\geq &C_r + \log \int_{\mathcal{Y}(r)} e^{C(\boldsymbol{p})} d\boldsymbol{p}.
\end{aligned}$$

Due to the relation (6.48), the equality in the above inequality holds when $P_n(\boldsymbol{n}) = J_n(\boldsymbol{n})$. Thus,

$$\min_{\{\sigma_n\}} \sup_{\rho} \lim_{n\to\infty} \left(D(\rho^{\otimes n} \| \sigma_n) - \frac{r^2-1}{2}\log n \right) = C_r + \log \int_{\mathcal{Y}(r)} e^{C(\boldsymbol{p})} d\boldsymbol{p}.$$

Since the density matrix $\overline{\rho(\boldsymbol{p})^{\otimes n}}$ is written as $\sum_{\boldsymbol{n}\in\mathcal{Y}_n^r} Q_{\boldsymbol{p}}(\boldsymbol{n})\rho_{\boldsymbol{n}}$, the density matrix $\tilde{\sigma}_{J,n}$ is written as $\sum_{\boldsymbol{n}\in\mathcal{Y}_n^r} Q_{J,n}(\boldsymbol{n})\rho_{\boldsymbol{n}}$. Since $Q_{J,n}(\boldsymbol{n})$ converges to $J_n(\boldsymbol{n})$, the density matrix $\tilde{\sigma}_{J,n}$ also attains the mini-max value in (6.45). ■

Exercise 6.6 Show that there is no universal approximation state σ_n when the function f satisfies the axiom of the distance or the function f is given as a strictly monotone increasing function of a function satisfying the axiom of the distance in addition to the additive condition $f(\rho^{\otimes n}, \sigma^{\otimes n}) = nf(\rho, \sigma)$.

6.5 Entanglement Concentration

6.5.1 Case with Known Basis of Reduced Density

To realize several useful quantum protocols like quantum teleportation introduced in Sect. 3.2, we need the maximally entangled state. However, in a realistic situation, it is often difficult to perfectly generate the maximally entangled state, and it is possible to generate an imperfect entangled state. A quantum protocol is called **entanglement distillation** when it generates maximally entangled state(s) from several copies of an imperfect entangled state, as Fig. 3.5. In particular, when the initial imperfect entangled state is pure, the entanglement distillation is called **entanglement concentration**.

We assume that we are given n independent copies of a pure state $|a\rangle$ across two systems $\mathcal{H}_A$ and $\mathcal{H}_B$ [4]. Using the Schmidt decomposition of a pure state $|a\rangle$;

$$|a\rangle = \sum_{j=1}^{r} \sqrt{p_j} |v_j\rangle \otimes |u_j\rangle, \tag{6.49}$$

we give our protocol. Hence, the reduced density matrices on the system $\mathcal{H}_A$ and $\mathcal{H}_B$ are calculated to $\sum_{j=1}^r p_j |v_j\rangle\langle v_j|$ and $\sum_{j=1}^r p_j |u_j\rangle\langle u_j|$, respectively.

Then, for $\boldsymbol{n} \in \mathcal{T}_n^r$, using the Schmidt bases $\mathcal{B}_A := \{|v_j\rangle\}$ and $\mathcal{B}_B := \{|u_j\rangle\}$, we define the projections on $\mathcal{H}_A$ and $\mathcal{H}_B$ as

$$P_{\boldsymbol{n}}^{\mathcal{B}_A} := \sum_{\vec{x} \in T_{\boldsymbol{n}}} |\vec{v}(\vec{x})\rangle\langle \vec{v}(\vec{x})|, \quad P_{\boldsymbol{n}}^{\mathcal{B}_B} := \sum_{\vec{x} \in T_{\boldsymbol{n}}} |\vec{u}(\vec{x})\rangle\langle \vec{u}(\vec{x})|.$$

Then, we can define the PVMs $\{P_{\boldsymbol{n}}^{\mathcal{B}_A}\}_{\boldsymbol{n}}$ and $\{P_{\boldsymbol{n}}^{\mathcal{B}_B}\}_{\boldsymbol{n}}$ on the system $\mathcal{H}_A^{\otimes n}$ and $\mathcal{H}_B^{\otimes n}$, respectively. The reduced density matrices on the system $\mathcal{H}_A^{\otimes n}$ and $\mathcal{H}_B^{\otimes n}$ of the pure state $|a\rangle^{\otimes n}$ on the composite system $\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$ are $\sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} \prod_{j=1}^r p_j^{n_j} P_{\boldsymbol{n}}^{\mathcal{B}_A}$ and $\sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} \prod_{j=1}^r p_j^{n_j} P_{\boldsymbol{n}}^{\mathcal{B}_B}$, respectively. Hence, the pure state $|a\rangle^{\otimes n}$ on the composite system is written as

$$|a\rangle^{\otimes n} = \sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} \prod_{j=1}^r p_j^{n_j/2} \sum_{\vec{x} \in T_{\boldsymbol{n}}} |\vec{v}(\vec{x})\rangle \otimes |\vec{u}(\vec{x})\rangle.$$

When we perform the measurements corresponding to the PVMs $\{P_{\boldsymbol{n},A}\}_{\boldsymbol{n}}$ and $\{P_{\boldsymbol{n},B}\}_{\boldsymbol{n}}$ on respective systems, the probability that the respective outcomes are $\boldsymbol{n}$ and $\boldsymbol{n}'$ is calculated to

$$\begin{aligned}
&\|P_{\boldsymbol{n}}^{\mathcal{B}_A} \otimes P_{\boldsymbol{n}'}^{\mathcal{B}_B} |a\rangle^{\otimes n}\|^2 = \|(I \otimes P_{\boldsymbol{n}'}^{\mathcal{B}_B})(P_{\boldsymbol{n}}^{\mathcal{B}_A} \otimes I)|a\rangle^{\otimes n}\|^2 \\
=&\|(I \otimes P_{\boldsymbol{n}'}^{\mathcal{B}_B})(P_{\boldsymbol{n}}^{\mathcal{B}_A} \otimes I) \sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} \prod_{j=1}^r p_j^{n_j/2} \sum_{\vec{x} \in T_{\boldsymbol{n}}} |\vec{v}(\vec{x})\rangle \otimes |\vec{u}(\vec{x})\rangle\|^2 \\
=&\|(I \otimes P_{\boldsymbol{n}'}^{\mathcal{B}_B}) \sum_{\boldsymbol{n} \in \mathcal{Y}_n^r} \prod_{j=1}^r p_j^{n_j/2} \sum_{\vec{x} \in T_{\boldsymbol{n}}} |\vec{v}(\vec{x})\rangle \otimes |\vec{u}(\vec{x})\rangle\|^2 \\
=&\delta_{\boldsymbol{n},\boldsymbol{n}'} \prod_{j=1}^r p_j^{n_j} |T_{\boldsymbol{n}}| = \delta_{\boldsymbol{n},\boldsymbol{n}'} \operatorname{Mul}[\boldsymbol{p}, n](\boldsymbol{n}).
\end{aligned}$$

Thus, these two outcomes coincide with each other. The final state is $\frac{1}{\sqrt{|T_{\boldsymbol{n}}|}}\sum_{\vec{x}\in T_{\boldsymbol{n}}} |\vec{v}(\vec{x})\rangle \otimes |\vec{u}(\vec{x})\rangle$, which is the maximally entangled state with length $\log |T_{\boldsymbol{n}}|$. This construction of the protocol depends on Schmidt basis given in (6.49).

However, the performance depends on Schmidt coefficient $\sqrt{p_j}$. The average of the number of generated maximally entangled state is $\mathrm{E}_{\mathrm{Mul}[\boldsymbol{p},n]}\Big[\log \frac{n!}{\boldsymbol{n}!}\Big] = \sum_{\boldsymbol{n}\in\mathcal{T}_n^r} \mathrm{Mul}[\boldsymbol{p},n](\boldsymbol{n}) \log \frac{n!}{\boldsymbol{n}!}$. The asymptotic behavior of this value is given by (6.11) of Lemma 6.4. That is, the number of generated maximally entangled state is almost equal to $nH(\boldsymbol{p})$. The entropy $H(\boldsymbol{p})$ is the asymptotically optimal generation rate of maximally entangled state even though the number of generated maximally entangled state is fixed. This fact shows that even though this number is not priorly determined, we can attain the optimal generation rate.

Exercise 6.7 Calculate the limit of the probability $\mathrm{Mul}[\boldsymbol{p},n]\{\frac{1}{\sqrt{n}}(\log \frac{n!}{\boldsymbol{n}!} - nH(\boldsymbol{p})) \geq R_2\}$ for any real number R_2. Here, R_2 is the second order generation rate because $\log \frac{n!}{\boldsymbol{n}!}$ expresses the size of the generated maximally entangled state.

6.5.2 Case with Unknown Basis of Reduced Density

The above protocol depends on Schmidt basis. In general, it is often the case that we do not know the Schmidt basis priorly. In such a situation, we need a protocol for entanglement concentration that does not depend on the Schmidt basis. Such a protocol is called universal entanglement concentration protocol.

In the following, we propose a protocol for entanglement concentration that works for $|X\rangle\rangle$ across two systems $\mathcal{H}_A$ and $\mathcal{H}_B$ independently of their Schmidt bases [93]. The respective tensor product spaces are decomposed to

$$\mathcal{H}_A^{\otimes n} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r(A)}} \mathcal{W}_{\boldsymbol{n},A} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r(A)}} \mathcal{U}_{\boldsymbol{n},A} \otimes \mathcal{V}_{\boldsymbol{n},A} \tag{6.50}$$

$$\mathcal{H}_B^{\otimes n} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r(B)}} \mathcal{W}_{\boldsymbol{n},B} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r(B)}} \mathcal{U}_{\boldsymbol{n},B} \otimes \mathcal{V}_{\boldsymbol{n},B}. \tag{6.51}$$

For simplicity, we assume that X is invertible. This assumption implies that these two systems $\mathcal{H}_A$ and $\mathcal{H}_B$ have the same dimension. Hence, when r is the dimension of $\mathcal{H}_A$, for $\boldsymbol{n} \in \mathcal{Y}_n^r$, $\mathcal{U}_{\boldsymbol{n},A}$ is isometric to $\mathcal{U}_{\boldsymbol{n},B}$. Thus, the matrix $\mathsf{f}_{\boldsymbol{n}}(X)$ on $\mathcal{U}_{\boldsymbol{n},A}$ can be regarded as a matrix from $\mathcal{U}_{\boldsymbol{n},A}$ to $\mathcal{U}_{\boldsymbol{n},B}$. Using this fact, the pure state $|X\rangle\rangle^{\otimes n}$ on the composite system can be written as

$$|X\rangle\rangle^{\otimes n} = \sum_{\boldsymbol{n}\in\mathcal{Y}_n^r} |\mathsf{f}_{\boldsymbol{n}}(X)\rangle\rangle \otimes |I_{\mathcal{V}_{\boldsymbol{n}}}\rangle\rangle. \tag{6.52}$$

When X is not invertible, we focus on the ranges $\mathcal{K}_A$ and $\mathcal{K}_B$ on both systems of $|X\rangle\rangle$. Then, the tensor product spaces of these ranges can be decomposed as

$$\mathcal{K}_A^{\otimes n} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r'}} \mathcal{W}'_{\boldsymbol{n},A} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r'}} \mathcal{U}'_{\boldsymbol{n},A} \otimes \mathcal{V}_{\boldsymbol{n},A} \tag{6.53}$$

$$\mathcal{H}_B^{\otimes n} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r'}} \mathcal{W}'_{\boldsymbol{n},B} = \oplus_{\boldsymbol{n}\in\mathcal{Y}_n^{r'}} \mathcal{U}'_{\boldsymbol{n},B} \otimes \mathcal{V}_{\boldsymbol{n},B}, \tag{6.54}$$

where r' is the rank of the reduced density matrix of X. The pure state $|X\rangle\!\rangle^{\otimes n}$ on the composite system is written as

$$|X\rangle\!\rangle^{\otimes n} = \sum_{\boldsymbol{n}\in\mathcal{Y}_n^{r'}} |\mathsf{f}_{\boldsymbol{n}}(X)\rangle\!\rangle \otimes |I_{\mathcal{V}_{\boldsymbol{n}}}\rangle\!\rangle. \tag{6.55}$$

Since $\mathcal{U}'_{\boldsymbol{n},A}$ and $\mathcal{U}'_{\boldsymbol{n},B}$ are subspaces of $\mathcal{U}_{\boldsymbol{n},A}$ and $\mathcal{U}_{\boldsymbol{n},B}$, the matrix $\mathsf{f}_{\boldsymbol{n}}(X)$ from $\mathcal{U}'_{\boldsymbol{n},A}$ to $\mathcal{U}'_{\boldsymbol{n},B}$ can be regarded as the matrix $\mathsf{f}_{\boldsymbol{n}}(X)$ from $\mathcal{U}_{\boldsymbol{n},A}$ to $\mathcal{U}_{\boldsymbol{n},B}$. Due to this characterization, the vector $|\mathsf{f}_{\boldsymbol{n}}(X)\rangle\!\rangle \otimes |I_{\mathcal{V}_{\boldsymbol{n}}}\rangle\!\rangle$ can be regarded as a state on $\mathcal{U}_{\boldsymbol{n},A} \otimes \mathcal{U}_{\boldsymbol{n},B} \otimes \mathcal{V}_{\boldsymbol{n},A} \otimes \mathcal{V}_{\boldsymbol{n},B}$.

In the following, we denote the projections to $\mathcal{W}_{\boldsymbol{n},A}$ and $\mathcal{W}_{\boldsymbol{n},B}$ by $P_{\boldsymbol{n},A}$ and $P_{\boldsymbol{n},B}$, respectively. When we perform the measurements $\{P_{\boldsymbol{n},A}\}_{\boldsymbol{n}}$ and $\{P_{\boldsymbol{n},B}\}_{\boldsymbol{n}}$ to the systems $\mathcal{H}_A^{\otimes n}$ and $\mathcal{H}_B^{\otimes n}$, respectively, the probability to obtain the outcomes $\boldsymbol{n}$ and $\boldsymbol{n}'$ is calculated as

$$\begin{aligned}
&\|P_{\boldsymbol{n},A} \otimes P_{\boldsymbol{n}',B}|A\rangle\!\rangle^{\otimes n}\|^2 = \|(I \otimes P_{\boldsymbol{n}',B})(P_{\boldsymbol{n},A} \otimes I)|A\rangle\!\rangle^{\otimes n}\|^2\\
=&\|(I \otimes P_{\boldsymbol{n}',B})(P_{\boldsymbol{n},A} \otimes I) \sum_{\boldsymbol{n}\in\mathcal{Y}_n^r} |\mathsf{f}_{\boldsymbol{n}}(A)\rangle\!\rangle \otimes |I_{\mathcal{V}_{\boldsymbol{n}}}\rangle\!\rangle\|^2\\
=&\|(I \otimes P_{\boldsymbol{n}',B})|\mathsf{f}_{\boldsymbol{n}}(A)\rangle\!\rangle \otimes |I_{\mathcal{V}_{\boldsymbol{n}}}\rangle\!\rangle\|^2 = \delta_{\boldsymbol{n},\boldsymbol{n}'} Q[\rho^{\otimes n}](\boldsymbol{n}),
\end{aligned}$$

where ρ is the reduced density matrix of $|A\rangle\!\rangle$ on $\mathcal{H}_A$. Hence, these two outcomes coincide with each other. The final state is $\frac{1}{\sqrt{Q[\rho^{\otimes n}](\boldsymbol{n})}}|\mathsf{f}_{\boldsymbol{n}}(A)\rangle\!\rangle \otimes |I_{\mathcal{V}_{\boldsymbol{n}}}\rangle\!\rangle$, which is a state on the system $\mathcal{U}_{\boldsymbol{n},A} \otimes \mathcal{U}_{\boldsymbol{n},B} \otimes \mathcal{V}_{\boldsymbol{n},A} \otimes \mathcal{V}_{\boldsymbol{n},B}$. Taking the partial trace with respect to the system $\mathcal{U}_{\boldsymbol{n},A} \otimes \mathcal{U}_{\boldsymbol{n},B}$, the final state is $\frac{1}{\sqrt{\dim \mathcal{W}_{\boldsymbol{n}}}}|I_{\mathcal{V}_{\boldsymbol{n}}}\rangle\!\rangle$. That is, we obtain the maximally entangled state with the length $\log\dim\mathcal{V}_{\boldsymbol{n}}$. Under this protocol, the average of the length of the generated maximally entangled state is $\sum_{\boldsymbol{n}\in\mathcal{Y}_n^r} Q[\rho^{\otimes n}](\boldsymbol{n}) \log\dim\mathcal{V}_{\boldsymbol{n}}$. The asymptotic behavior of this value is given as (6.35). This value is smaller than the average length $\sum_{\boldsymbol{n}\in\mathcal{T}_n^r} \mathrm{Mul}[\boldsymbol{p},n](\boldsymbol{n}) \log\frac{n!}{\boldsymbol{n}!}$ of the protocol with the knowledge of the Schmidt bases given in Sect. 6.5.2. Due to the relation (6.35) in Lemma 6.9, the limit of the difference is calculated to

$$\frac{\sum_{\sigma\in S_r} \mathrm{sgn}(\sigma) \prod_i p_i^{\delta_{\sigma(i)}} \log \prod_j p_j^{\delta_{\sigma(j)}}}{\prod_{i,j:i<j}(p_i-p_j)} - \log \prod_{i,j:i<j} (p_i - p_j).$$

That is, the knowledge of the Schmidt bases produces only negligible advantage for the length of generated maximally entangled state.

Exercise 6.8 Calculate the limit of the probability $Q[\rho^{\otimes n}]\{\frac{1}{\sqrt{n}}(\log\dim\mathcal{V}_{\boldsymbol{n}} - nH(\boldsymbol{p})) \geq R_2\}$ for any real number R_2.

6.6 Quantum Data Compression

Data compression has two types, the fixed-length type and the variable-length type. In the classical case, the fixed-length type is a method to determine the compression rate independently of the input data, and the variable-length type is a method to determine the compression rate dependently of the input data. Hence, the fixed-length type has a possibility to fail to recover the original data. However, when we choose a suitable compression method with a suitable compression rate based on the probability distribution of the input data, the probability of the above failure can be reduced. On the other hand, in the variable-length type, we can perfectly recover the original data when we employ a suitable compression method with a suitable compression rate. However, there is a possibility to have a large compression rate, i.e., the case when the compressed data is larger than the original data. When we suitably choose the compression protocol and the compression rate dependently of the probability distribution of the input data, the averaged compression rate can be reduced. In summary, in the fixed-length type, the success or the failure of the decoding is a random variable, and in the variable-length type, the length of compressed data is a random variable. Since the success of the decoding is the absolute requirement in practice, the modern computer employs the variable-length type of data compression.

6.6.1 Fixed-Length Data Compression

This subsection gives a protocol for fixed-length data compression in the quantum system [48, 80]. In fixed-length data compression, an encoding process is given as a TP-CP map E from the input quantum system $\mathcal{H}$ to the quantum memory system $\mathcal{K}$, and a decoding process is given as a TP-CP map D from the quantum memory system $\mathcal{K}$ to the input quantum system $\mathcal{H}$, as Fig. 6.4. The information source is described as a probability distribution $\{\mathrm{P}(i), |x_j\rangle\langle x_j|\}$ on the set of pure states on the input system $\mathcal{H}$. Then, the performance of the code (E, D) is characterized as a kind of error;

$$\varepsilon_{\mathrm{P}}(E, D) := \sum_j \mathrm{P}(j)(1 - \langle x_j|D \circ E(|x_j\rangle\langle x_j|)|x_j\rangle). \tag{6.56}$$

Hence, our problem is the trade-off between $\varepsilon_{\mathrm{P}}(E, D)$ and $\dim \mathcal{K}$. In general, due to (2.48), using **averaged density matrix** $\rho_{\mathrm{P}} := \sum_j \mathrm{P}(j)|x_j\rangle\langle x_j|$ and entanglement fidelity defined in (2.46), we can evaluate this value as $\varepsilon_{\mathrm{P}}(E, D) \le 1 - F_e^2(D \circ E, \rho_{\mathrm{P}})$.

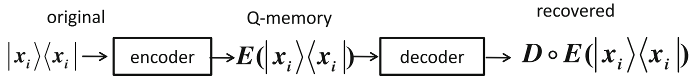

original
Q-memory
recovered
$|x_i\rangle\langle x_i|$
encoder
$E(|x_i\rangle\langle x_i|)$
decoder
$D \circ E(|x_i\rangle\langle x_i|)$

Fig. 6.4 Fixed length quantum data compression

That is, it is possible to evaluate $\varepsilon_{\mathrm{P}}(E, D)$ independently of the ensemble in the information source and dependently only of the averaged density matrix ρ_{P}.

Next, based on an isometry U from $\mathcal{K}$ to $\mathcal{H}$, we define the encoder E_U and the decoder D_U as

$$E_U(\rho) := U^\dagger \rho U + (\operatorname{Tr} \rho(I - U^\dagger U))\rho_{\mathrm{mix},\mathcal{K}}, \quad D_U(\rho) := U\rho U^\dagger. \tag{6.57}$$

Then, using $P_U := UU^\dagger$, we have $D_U \circ E_U(\rho) = P_U \rho P_U + (\operatorname{Tr} \rho(I - P_U))U\rho_{\mathrm{mix},\mathcal{K}} U^\dagger \le \Lambda[P_U](\rho)$. Since the purification $|X\rangle\rangle$ of ρ satisfies $\langle\langle X|P_U \otimes I|X\rangle\rangle = \operatorname{Tr} \rho_{\mathrm{P}} P_U$,

$$\begin{aligned} &1 - F_e^2(D_U \circ E_U, \rho_{\mathrm{P}}) \le 1 - F_e^2(\Lambda[P_U], \rho_{\mathrm{P}}) \\ =&1 - \langle\langle X|P_U \otimes I|X\rangle\rangle \langle\langle X|P_U \otimes I|X\rangle\rangle = 1 - (\operatorname{Tr} \rho_{\mathrm{P}} P_U)^2 \\ =&2(1 - \operatorname{Tr} \rho_{\mathrm{P}} P_U) - (1 - \operatorname{Tr} \rho_{\mathrm{P}} P_U)^2 \le 2 \operatorname{Tr} \rho_{\mathrm{P}}(I - P_U). \end{aligned} \tag{6.58}$$

When restrict our encoder and our decoder into the above types of encoders and decoders, our problem can be regarded as the trade-off to reduce both of the upper bound $\operatorname{Tr} \rho_{\mathrm{P}}(I - P)$ of our error and the dimension $\operatorname{Tr} P$ of the memory, where P is the projection to the range of U.

Next, we consider the case when the information source is subject to the n-fold independent and identical distribution of P. Then, the averaged density matrix on the input system $\mathcal{H}^{\otimes n}$ is given as $\rho_{\mathrm{P}}^{\otimes n}$. Let $\boldsymbol{p}$ be the vector composed of the eigenvalues of ρ_{P}. Given $R > 0$, we define the projection $Q[R, n] := \sum_{\boldsymbol{n}:H(\frac{\boldsymbol{n}}{n}) \le R} P_{\boldsymbol{n}}$. The relations (6.18), (6.20), and (6.23) yield

$$\operatorname{Tr} Q[R, n] \le (n + 1)^{r-1} e^{nR}, \tag{6.59}$$

$$\operatorname{Tr} \rho_{\mathrm{P}}^{\otimes n} Q[R, n] \le \max_{\boldsymbol{q}:H(q) \le R} (n + 1)^{r(r+1)/2} e^{-nD(\boldsymbol{q}\|\boldsymbol{p})}. \tag{6.60}$$

By letting $\mathcal{K}_n$ be the range of $Q[R, n]$ and U_n be the isometric embedding from $\mathcal{K}_n$ to $\mathcal{H}^{\otimes n}$, the encoder E_{U_n} and the decoder D_{U_n} satisfy

$$\lim_{n\to\infty} \frac{-1}{n} \log \varepsilon_{\mathrm{P}}(E_{U_n}, D_{U_n}) = \min_{\boldsymbol{q}:H(q) \le R} D(\boldsymbol{q}\|\boldsymbol{p}) \tag{6.61}$$

$$\lim_{n\to\infty} \frac{-1}{n} \log \dim \mathcal{K}_n = R. \tag{6.62}$$

That is, when $H(\boldsymbol{p}) = H(\rho) > R$, (6.61) is strictly positive, which implies that $\varepsilon_{\mathrm{P}}(E_{U_n}, D_{U_n})$ goes to zero exponentially. Due to this fact, once the condition $H(\rho) > R$ holds, it is possible to make a fixed-length data compression code whose error goes to zero exponentially independently of the form of the information source.

Now, we consider slightly different scenario. Assume that we have n copies of an entangle state $|X\rangle\rangle$ on $\mathcal{H}_A \otimes \mathcal{H}_B$. To save the size in the system $\mathcal{H}_A^{\otimes n}$, we apply the above compression protocol (the encoder and the decoder) only on the system $\mathcal{H}_A^{\otimes n}$. In this case, the fidelity is evaluated by the entanglement fidelity

$F_e^2(D \circ E, \mathrm{Tr}_B |X\rangle\!\rangle\langle\!\langle X|)$ defined in (2.46). Since the relation (6.58) evaluates this value by using $\mathrm{Tr}_A \mathrm{Tr}_B |X\rangle\!\rangle\langle\!\langle X| P_U$, the evaluation of this subsection can be applied to this scenario.

Exercise 6.9 Calculate the limits of $\frac{1}{\sqrt{n}}(\log \mathrm{Tr}\, Q[H(\boldsymbol{p}) + \frac{R_2}{\sqrt{n}}, n] - H(\boldsymbol{p}))$ and $\mathrm{Tr}\, \rho_{\mathrm{P}}^{\otimes n} Q[H(\boldsymbol{p}) + \frac{R_2}{\sqrt{n}}, n]$ for any real number R_2. This calculation gives the second order compression rate.

6.6.2 Variable-Length Data Compression I: Determined Compression Rate

The discussion in Sect. 6.6.1 shows that we can construct a code to attain the entropy rate $H(\rho_{\mathrm{P}})$ of the information source when the entropy rate is known. However, when the entropy rate is unknown, we cannot employ the protocol for fixed-length data compression given in Sect. 6.6.1. This situation needs variable-length data compression.

In the classical variable-length data compression, the compression rate is determined dependently of the input data. The quantum variable-length data compression, the length of stored memory is determined according to the state of the input system. To decide the length, we need the quantum measurement. Such a type of data compression is called a determined-compression-rate-type quantum variable-length data compression. Since a state is demolished due to the measurement, the original quantum state is not necessarily recovered in determined-compression-rate-type unlike the classical case. The state demolition causes loss of a part of information. Since the knowledge of the entropy rate $H(\rho_{\mathrm{P}})$ enables us to compress the quantum information source with negligible error as shown in Sect. 6.6.1, we need to address the trade-off between the estimation of the entropy rate $H(\rho_{\mathrm{P}})$ and the state demolition.

Hence, we need to describe the encoding process by an instrument $\{\Lambda_\omega\}_\omega$ that has the compression rate and the state change. Here, the compression length $R(\omega)$ is determined from the measurement outcome ω. Dependently of the compression length $R(\omega)$, we define the decoder D_ω, which is a TP-CP map. Such a protocol is described as $\Phi = (\{\Lambda_\omega\}_\omega, \{R(\omega)\}_\omega, \{D_\omega\})$. When the information source is described by the probability distribution $\{\mathrm{P}(i), |x_j\rangle\langle x_j|\}$ on the set of pure states on the input system $\mathcal{H}$, the performance of the code Φ is given by the behavior of the random variable $R(\omega)$ under the probability distribution $\sum_i \mathrm{P}(i)\, \mathrm{Tr}\, \Lambda_\omega(|x_i\rangle\langle x_i|)$ and the expectation $F^2(\{\Lambda_\omega\}_\omega, \{D_\omega\}) := \sum_i \mathrm{P}(i) \sum_\omega \mathrm{Tr}\, \Lambda_\omega(|x_j\rangle\langle x_j|)\langle x_j| \frac{D_\omega(\Lambda_\omega(|x_j\rangle\langle x_j|))}{\Lambda_\omega(|x_j\rangle\langle x_j|)} |x_j\rangle = \sum_i \mathrm{P}(i) \sum_\omega \langle x_j| D_\omega(\Lambda_\omega(|x_j\rangle\langle x_j|))|x_j\rangle$ of the square of the fidelity between the final and initial states. In the following, we address the case when the instrument $\{\Lambda_\omega\}_\omega$ is given as $\Lambda_\omega(\rho) = \sqrt{M_\omega}\rho\sqrt{M_\omega}$ with a POVM $\{M_\omega\}_\omega$, the memory system is the range of M_ω, and D_ω is the identity map. When the input system is $\mathcal{H}$, $R(\omega)$ is $\log \mathrm{rank}\, M_\omega$ and we have

$$F^2(\{\Lambda_\omega\}_\omega, \{D_\omega\}) = \sum_i \mathrm{P}(i) \sum_\omega \langle x_i|\sqrt{M_\omega}|x_i\rangle\langle x_i|\sqrt{M_\omega}|x_i\rangle$$
$$= \sum_i \mathrm{P}(i) \sum_\omega \langle x_i|\sqrt{M_\omega}|x_i\rangle^2. \tag{6.63}$$

Here, we consider the case when the input system is $\mathbb{C}^{r\otimes n}$ and the input information source is given as the n-fold independent and identical distribution P^n of a distribution P on the set of pure states on $\mathbb{C}^r$. When the set of projections $\{P_{\boldsymbol{n}}\}_{\boldsymbol{n}}$ are used as our POVM $\{M_\omega\}$, the compression rate $\frac{\log \operatorname{rank} P_{\boldsymbol{n}}}{n}$ converges to the entropy rate $H(\rho_{\mathrm{P}})$ in probability. However, the expectation of the square of the fidelity does not converge to 1 so that the error remains even in the asymptotic limit. To overcome this issue, we introduce a measurement to estimate the entropy rate $H(\rho_{\mathrm{P}})$ only with negligible state demolition so that the fidelity goes to 1 in the asymptotic limit.

Given a real number $\epsilon > 0$, using $U_{x,\epsilon} := (x-\epsilon, x+\epsilon]$, we define the matrix $M_{x,\epsilon}$ as $M_{x,\epsilon} := \sum_{\boldsymbol{n}:H(\frac{\boldsymbol{n}}{n})\in U_{x,\epsilon}} P_{\boldsymbol{n}}$. Then, we have $\sum_{j\in\mathbb{Z}} M_{\epsilon j,\epsilon} = I$. Hence, for a positive integer l, we have $\sum_{j\in\mathbb{Z}} M_{\epsilon j,\epsilon l} = lI$. Defining $M_{\epsilon j}^{\epsilon,l} := \frac{1}{l} M_{\epsilon j,\epsilon l}$, we have a POVM $M^{\epsilon;l} := \{M_{\epsilon j}^{\epsilon,l}\}_{j\in\mathbb{Z}}$ with measurement outcome j, which will be employed to construct our compression protocol. Notice that the measurement outcome of the POVM $M^{\epsilon;l}$ can be simulated by the PVM $\{P_{\boldsymbol{n}}\}_{\boldsymbol{n}}$. That is, the POVM $\{M_{\epsilon j}^{\epsilon,l}\}_{j\in\mathbb{Z}}$ can be realized by the combination of the following defined probability transition matrix $Q(j|\boldsymbol{n})$ and the PVM $\{P_{\boldsymbol{n}}\}_{\boldsymbol{n}}$;

$$Q(j|\boldsymbol{n}) := \begin{cases} \frac{1}{l} \text{ when } \epsilon j - \epsilon l < H(\frac{\boldsymbol{n}}{n}) \le \epsilon j + \epsilon l \\ 0 \text{ otherwise.} \end{cases} \tag{6.64}$$

Since $\sqrt{M_{\epsilon j}^{\epsilon,l}} = \frac{1}{\sqrt{l}} M_{\epsilon j,\epsilon l}$, Jensen inequality (Lemma 2.5) yields the following evaluation of (6.63);

$$\sum_{\boldsymbol{i}} \mathrm{P}^n(\boldsymbol{i}) \sum_{j\in\mathbb{Z}} \langle \boldsymbol{x_i}|\sqrt{M_{\epsilon j}^{\epsilon,l}}|\boldsymbol{x_i}\rangle^2 = \sum_{\boldsymbol{i}} \mathrm{P}^n(\boldsymbol{i}) \sum_{j\in\mathbb{Z}} \frac{1}{l} \langle \boldsymbol{x_i}|M_{\epsilon j,\epsilon l}|\boldsymbol{x_i}\rangle^2$$
$$\ge \sum_{j\in\mathbb{Z}} \frac{1}{l} (\sum_{\boldsymbol{i}} \mathrm{P}^n(\boldsymbol{i}) \langle \boldsymbol{x_i}|M_{\epsilon j,\epsilon l}|\boldsymbol{x_i}\rangle)^2 = \sum_{j\in\mathbb{Z}} \frac{1}{l} (\operatorname{Tr} \rho_{\mathrm{P}}^{\otimes n} M_{\epsilon j,\epsilon l})^2. \tag{6.65}$$

When $j \in \mathbb{Z}$ satisfies the condition $\epsilon j \in U_{H(\rho_{\mathrm{P}}),\epsilon(l-1)}$, the relation $U_{H(\rho_{\mathrm{P}}),\epsilon} \subset U_{\epsilon j,\epsilon l}$ holds. Hence, we have

$$\operatorname{Tr} \rho_{\mathrm{P}}^{\otimes n} M_{\epsilon j,\epsilon l} \ge \sum_{H(\frac{\boldsymbol{n}}{n})\in U_{H(\rho_{\mathrm{P}}),\epsilon}} \operatorname{Tr} \rho_{\mathrm{P}}^{\otimes n} P_{\boldsymbol{n}}. \tag{6.66}$$

Since the number of integers $j \in \mathbb{Z}$ satisfying the condition $\epsilon j \in U_{H(\rho_{\mathrm{P}}),\epsilon(l-1)}$ is $l-1$, we have

$$\sum_{j\in\mathbb{Z}} \frac{1}{l}(\mathrm{Tr}\,\rho_{\mathrm{P}}^{\otimes n} M_{\epsilon j,\epsilon l})^2 \ge \frac{l-1}{l}\Big(\sum_{H(\frac{\boldsymbol{n}}{n})\in U_{H(\rho_{\mathrm{P}}),\epsilon}} \mathrm{Tr}\,\rho_{\mathrm{P}}^{\otimes n} P_{\boldsymbol{n}}\Big)^2. \tag{6.67}$$

When ϵ_n satisfies the condition $\sqrt{n}\epsilon_n \to \infty$, the discussion in the initial part of Sect. 6.2.4 guarantees that the value $\sum_{H(\frac{\boldsymbol{n}}{n})\in U_{H(\rho_{\mathrm{P}}),\epsilon_n}} \mathrm{Tr}\,\rho_{\mathrm{P}}^{\otimes n} P_{\boldsymbol{n}}$ converges to 1. When ϵ is chosen to be ϵ_n and l is chosen to be an integer l_n satisfying the conditions $l_n \to \infty$ and $\epsilon_n l_n \to 0$, the RHS of (6.67) converges to 1. That is, due to (6.65), the fidelity between the final and initial states given in (6.63) converges to 1.

Notice that the compression rate is $\frac{1}{n}\log\operatorname{rank} M_{\epsilon_n j}^{\epsilon_n,l_n}$, which is decided by the measurement outcome j. The relation (6.18) and the discussion of the initial part of Sect. 6.2.4 guarantee that the compression rate $\frac{1}{n}\log\operatorname{rank} M_{\epsilon_n j}^{\epsilon_n,l_n}$ asymptotically approaches to $\epsilon_n j$ under the limit $n \to \infty$. That is, the value $\max_{j\in\mathbb{Z}} |\frac{1}{n}\log M_{\epsilon_n j}^{\epsilon_n,l_n} - \epsilon_n j|$ converges to 0.

Now, we virtually consider that we firstly apply PVM $E = \{P_{\boldsymbol{n}}\}_{\boldsymbol{n}}$ and apply the transition matrix $Q(j|\boldsymbol{n})$ instead of the direct application of $M^{\epsilon_n;l_n}$. The condition $\epsilon_n l_n \to 0$ and the relation (6.64) guarantee that the difference $\epsilon_n j - H(\frac{\boldsymbol{n}}{n})$ converges to zero in probability. Since the random variable $H(\frac{\boldsymbol{n}}{n})$ converges to $H(\rho_{\mathrm{P}})$ in probability, the random variable $\epsilon_n j$ decided by the measurement outcome converges to $H(\rho_{\mathrm{P}})$ in probability. Hence, the random variable $\frac{1}{n}\log M_{\epsilon_n j}^{\epsilon_n,l_n}$ converges to $H(\rho_{\mathrm{P}})$ in probability. In this way, we can construct a determined-compression-rate-type universal variable-length data compression [50, 51].

6.6.3 Prefix Code and Kraft Inequality

Although the quantum variable-length data compression given in the above subsection has asymptotically zero error, it has non-zero error in the finite-length setting. However, in the classical case, it is possible to construct an error-free variable-length data compression. When the information source takes values in the set $\{1,\dots,d\}$, an **error-free variable-length code** is given as an injective map from $\{1,\dots,d\}$ to $\{0,1\}^* := \cup_{n\ge 0}\{0,1\}^n$. For any element $i \in \{1,\dots,d\}$, $\phi(i)$ is called a **codeword** and the integer n satisfying $\phi(i) \in \{0,1\}^n$ is called codeword length and is denoted by $|\phi(i)|$. For example, when the information source is subject to the probability distribution P, the **averaged codeword length** is given as $\sum_{i=1}^d \mathrm{P}(i)|\phi(i)|$. Then, an error-free variable-length code ϕ on $\{1,\dots,d\}$ satisfies

$$\sum_{i=1}^{d} \frac{1}{2^{l_i}} \le \lceil \log_2 d \rceil, \tag{6.68}$$

where $l_i := |\phi(i)|$.

Proof of (6.68) Let $f(M)$ be the maximum value of $\sum_{i=1}^M 2^{-|\phi(i)|}$ under the condition that M codewords $\phi(1),\dots,\phi(M)$ are distinct elements in $\{0,1\}^* := \cup_{n=1}^\infty\{0,1\}^n$.

When $M = 2^1 + 2^2 + \cdots + 2^m$, it is calculated to be $f(2^1 + 2^2 + \cdots + 2^m) = m$. Hence, we have $f(d) \le f(2^1 + 2^2 + \cdots + 2^{\lceil \log_2 d \rceil}) \le \lceil \log_2 d \rceil$, which yields the desired statement. ■

When we concatenate two error-free variable-length codes ϕ_1 and ϕ_2, a map $\phi_1 \cdot \phi_2$ from $\{1, \ldots, d\}^2$ to $\{0, 1\}^*$ is defined as $\phi_1 \cdot \phi_2(i, j) := \phi_1(i)\phi_2(j)$. However, even when ϕ_1 and ϕ_2 are injective, the map $\phi_1 \cdot \phi_2$ is not necessarily injective (See Exercise 6.10). An error-free variable-length code ϕ is called **uniquely decodable** when the map $\phi^n := \overbrace{\phi \cdot \cdots \cdot \phi}^{n}$ is injective for any positive integer n. For an error-free variable-length code ϕ and integers $i \in \{1, \ldots, d\}$ and $j \le |\phi(i)|$, the initial j elements in the binary sequence $\phi(i)$ is called a **prefix** of the codeword $\phi(i)$. For example, when $\phi(i) = 0111$, 0, 01, and 011 are prefixes of the codeword $\phi(i)$. An error-free variable-length code ϕ is called a **prefix code** when any codeword is different from any prefix of another codeword. When ϕ_1 and ϕ_2 are prefix codes, the map $\phi_1 \cdot \phi_2$, i.e., the concatenation of ϕ_1 and ϕ_2, is an injective map from $\{1, \ldots, d\}^2$ to $\cup_{n \ge 0}\{0, 1\}^n$. In this case, the map $\phi_1 \cdot \phi_2$ is also a prefix code (Exercise 6.13). Repeating this discussion, we find that the map ϕ^n is a prefix code for any prefix code ϕ. Hence, any prefix code ϕ is uniquely decodable. However, a uniquely decodable error-free variable-length code is not necessarily a prefix code. (See Exercise 6.11 or [79, Example 1.14].) Any prefix code ϕ on $\{1, \ldots, d\}$ satisfies **Kraft inequality** as follows

$$\sum_{i=1}^{d} \frac{1}{2^{l_i}} \le 1, \tag{6.69}$$

where $l_i := |\phi(i)|$. In fact, a uniquely decodable error-free variable-length code ϕ satisfies **McMillian inequality** [37, Theorem 3.8] as

$$\sum_{i=1}^{d} \frac{1}{2^{l_i}} \le 1 \tag{6.70}$$

even though it is not necessarily a prefix code. Kraft inequality (6.69) follows from McMillian inequality.

Proof of (6.70) Any code ϕ^n satisfies $(\sum_{i=1}^{d} \frac{1}{2^{l_i}})^n = \sum_{i=1}^{d^n} \frac{1}{2^{\phi^n(i)}}$. Applying the inequality (6.68) to the error-free variable-length code ϕ^n, we have $\sum_{i=1}^{d^n} \frac{1}{2^{\phi^n(i)}} \le \lceil n \log_2 d \rceil$. That is, $n \log(\sum_{i=1}^{d} \frac{1}{2^{l_i}}) \le \log \lceil n \log_2 d \rceil$. Dividing both sides by n and taking the limit $n \to \infty$, we obtain $\sum_{i=1}^{d} \frac{1}{2^{l_i}} \le 1$. ■

Conversely, when a set $\{l_i\}$ of integers satisfies (6.69), there exists a prefix code ϕ such that $l_i := |\phi(i)|$. The construction of such a prefix code ϕ is given in [79, Sect. 1.4]. That is, when a probability distribution given as the LHS of (6.69), there exists a prefix code corresponding to the probability distribution in the sense of Kraft inequality. In other word, there exists a prefix code whose averaged codeword length is $\sum_{i=1}^{d} \mathrm{P}(i) l_i$ under a probability distribution P. For any probability distribution

Q, the integers $l_i := \lceil -\log_2 \mathrm{Q}(i) \rceil$ satisfy the condition (6.69). So, there exists a prefix code whose averaged codeword length is less than $-\sum_{i=1}^{d} \mathrm{P}(i) \log_2 \mathrm{Q}(i) + 1$ for any probability distribution P. Thus, the averaged codeword length is given as $-\sum_{i=1}^{d} \mathrm{P}(i) \log_2 \mathrm{Q}(i)$ within error 1 by using a probability distribution Q corresponding to the prefix code and another probability distribution P corresponding to the information source. Since $D(\mathrm{P}\|\mathrm{Q}) \geq 0$, this value is calculated to

$$-\sum_{i=1}^{d} \mathrm{P}(i) \log_2 \mathrm{Q}(i) = \frac{H(\mathrm{P}) + D(\mathrm{P}\|\mathrm{Q})}{\log 2} \geq \frac{H(\mathrm{P})}{\log 2}. \tag{6.71}$$

When the information source is given by the probability distribution P, the minimum averaged codeword length is $\frac{H(\mathrm{P})}{\log 2}$. Due to the equality condition of the above inequality, it is attained when the prefix code is given by the distribution P.

Exercise 6.10 Consider an error-free variable-length code ϕ defined as $(1, 2) \mapsto ((0), (0, 0))$. Answer whether the map ϕ^2 from $\{1, 2\}^2$ to $\{0, 1\}^*$ is injective or not.

Exercise 6.11 Consider the error-free variable-length code ϕ_1 defined as $(1, 2) \mapsto ((0), (0, 1))$. Show that ϕ_1 is uniquely decodable.

Exercise 6.12 Find a prefix code ϕ_2 such that $\phi_1 \cdot \phi_2$ is not invertible.

Exercise 6.13 Show that the map $\phi_1 \cdot \phi_2$ is also an prefix code when ϕ_1 and ϕ_2 are prefix codes.

Exercise 6.14 Make a prefix code corresponding to the distribution (1/4, 1/8, 1/8, 1/2).

6.6.4 Variable-Length Data Compression II: Undetermined Compression Rate

In Sect. 6.6.2, we have constructed a determined-compression-rate-type universal variable-length data compression. However, it has non-zero error while the error goes to zero asymptotically. Koashi-Imoto [85] showed that the state demolition cannot be vanished perfectly when the information source contains density matrices non-commutative to each other. Hence, if we need to make a variable-length data compression that has completely zero error like a classical code discussed in Sect. 6.6.3, we have to give up to determine the compression rate. Since such a code needs to keep a superposition of distinct coding rates, to make such a code, we need to consider the system $\mathcal{H}_{\oplus} := \bigoplus_{k=0}^{\infty} (\mathbb{C}^2)^{\otimes k}$ and define a quantum error-free variable-length code as an isometric map U from the d-dimensional quantum system $\mathcal{H}$ to $\mathcal{H}_{\oplus}$ [54]. Now, let P_k be the projection to $(\mathbb{C}^2)^{\otimes k}$ and define the operator $H := \sum_{k=0}^{\infty} k P_k$. The **averaged codeword length** is given as $\mathrm{Tr}\, \rho U^{\dagger} H U$ when the averaged density matrix of the information source is ρ_{P}. In this scenario with undermined compression rate, it is natural that k is considered to be the energy rather than the length of the memory. In general, it is better to store the information so that the averaged energy is smaller. Hence, we adopt a strategy to reduce the averaged energy for the information storage. Then, the following lemma holds [54].

Lemma 6.10 *A quantum error-free variable-length code U on the d-dimensional system $\mathcal{H}$ satisfies*

$$\operatorname{Tr} U^\dagger 2^{-H} U \le \lceil \log_2 d \rceil. \tag{6.72}$$

Then, the density matrix $\sigma(U) := \frac{1}{\operatorname{Tr} U^\dagger 2^{-H} U} U^\dagger 2^{-H} U$ on the system $\mathcal{H}$ satisfies

$$-\operatorname{Tr} \rho \log_2 \sigma(U) - \log\lceil \log_2 d \rceil \le \operatorname{Tr} HU\rho U^\dagger, \forall \rho \in \mathcal{S}(\mathcal{H}). \tag{6.73}$$

Proof Letting $f(M) := \max_{\operatorname{rank} P = M} \operatorname{Tr} 2^{-H} P$, we have $\operatorname{Tr} U^\dagger 2^{-H} U = \operatorname{Tr} 2^{-H} UU^\dagger \le f(d)$. The same discussion as the proof of (6.68) yields the inequality $f(d) \le \lceil \log_2 d \rceil$. Hence, we obtain (6.72).

Further, the matrix concavity of the function $x \mapsto \log x$ implies the inequality $-\log_2 U^\dagger 2^{-H} U \le U^\dagger H U$. So, the inequality (6.72) guarantees the inequality (6.73). ■

As explained for a classical code, a concatenation of two quantum error-free variable-length codes is not necessarily a quantum error-free variable-length code. In the classical case, a concatenation of two error-free variable-length codes can be defined as a map. However, in the quantum case, it is not trivial to define a concatenation of two quantum error-free variable-length codes U_1 and U_2 even as a map. A quantum error-free variable-length code U on a d-dimensional system $\mathcal{H}$ is called a **uniquely decodable** when there exist a classical uniquely decodable error-free variable-length code ϕ and a basis $\{e_1, \ldots, e_d\}$ of the d-dimensional system $\mathcal{H}$ such that

$$U = \sum_{i=1}^{d} |\phi(i)\rangle\langle e_i|. \tag{6.74}$$

Then, the n-concatenation of U, i.e., the quantum error-free variable-length code $U^{(n)}$ is defined as

$$U^{(n)} = \sum_{i_1=1}^{d} \cdots \sum_{i_n=1}^{d} |\phi(i_1) \cdots \phi(i_n)\rangle\langle e_{i_1} \cdots e_{i_n}|. \tag{6.75}$$

Especially, when ϕ is a prefix code, U is called a **quantum prefix code**.

McMillian inequality (6.70) guarantees that a uniquely decodable quantum error-free variable-length code U on the d-dimensional quantum system $\mathcal{H}$ satisfies the relations

$$\operatorname{Tr} U^{\dagger} 2^{-H} U = \sum_{i=1}^{d} 2^{-|\phi(i)|} \leq 1. \tag{6.76}$$

Conversely, for a density matrix σ on the quantum system $\mathcal{H}$, there exists a quantum prefix code U such that

$$U^{\dagger} H U \leq -\log \sigma + I. \tag{6.77}$$

Proof of (6.77) Consider the diagonalization $\sigma = \sum_{i=1}^{d} \mathrm{Q}(i)|e_i\rangle\langle e_i|$ of σ. Since the integers $l_i = \lceil -\log_2 \mathrm{Q}(i) \rceil$ satisfy the condition (6.69), there exists a classical prefix code $l_i = \phi(i)$. When the quantum prefix code U is defined by (6.74), we obtain (6.77). ∎

Within error 1, the averaged codeword length is given as $-\operatorname{Tr} \rho \log_2 \sigma$ by using the density matrix σ corresponding to the quantum prefix code and the density matrix ρ corresponding to the information source. Since $D(\rho\|\sigma) \geq 0$, this value is calculated to

$$-\operatorname{Tr} \rho \log_2 \sigma = \frac{H(\rho) + D(\rho\|\sigma)}{\log 2} \geq \frac{H(\rho)}{\log 2}. \tag{6.78}$$

When the averaged density matrix of the information source is ρ, the minimum of the averaged codeword length is $\frac{H(\rho)}{\log 2}$. Due to the equality condition of the above inequality, we find that the minimum is realized when the quantum prefix code corresponds to the density matrix ρ. Then, the quantum relative entropy $D(\rho\|\sigma)$ can be regarded as **redundancy**.

When the averaged density matrix of the information source is unknown and we have several candidates of the unknown state, how can we find a density matrix σ to reduce the redundancy $D(\rho\|\sigma)$ uniformly? In general, we rarely have no assumption for the candidates of the density matrix ρ. Now, we assume that the physical system of our interest is given as the tensor product space $\mathcal{H}^{\otimes n}$ of $\mathcal{H} := \mathbb{C}^d$ and the averaged density matrix of the information source is given as the tensor product state $\rho^{\otimes n}$. We consider the problem to find a density matrix σ to uniformly reduce the redundancy $D(\rho^{\otimes n}\|\sigma)$ independently of ρ. This problem can be regarded as the problem to uniformly approximate the tensor product state $\rho^{\otimes n}$ in the sense of quantum relative entropy. Due to the discussion in Sect. 6.4, there exists a density matrix σ_n on $\mathcal{H}^{\otimes n}$ such that

$$D(\rho^{\otimes n}\|\sigma_n) \cong \frac{d^2-1}{2} \log n + C_{\min-\max}(d). \tag{6.79}$$

Since $d^2 - 1$ is the dimension of the set of density matrices, this leading therm can be regarded as the same term as in the classical case. The detailed optimization for the constant term has been discussed in Theorem 6.3.

Whatever distribution P describes the information source in the sense of its n-fold independent and identical distribution P^n, the averaged density matrix has the form $\rho_{\mathrm{P}}^{\otimes n}$. Hence, we can construct a quantum error-free variable-length code that universally works by using the density matrix $\rho_{J,n}$ or $\tilde{\rho}_{J,n}$ given in Theorem 6.3 [54].

6.7 Classical-Quantum Channel Coding

6.7.1 Formulation

Chapter 5 has dealt with the problem to transmit a quantum state via noisy quantum channel. This section addresses the problem to transmit a classical message via noisy quantum channel, which is easier task than the above task. Using this advantage, we construct a pair of an encoder and a decoder that universally works, i.e., whose construction does not depend on the noisy channel. In this problem, we consider a quantum channel Λ from the input system $\mathcal{H}'$ to the output system $\mathcal{H} = \mathbb{C}^d$ and priorly choose a set of classical symbols $\mathcal{X} := \{x_1, \ldots, x_k\}$ (alphabet) and a map $x \mapsto \rho_x$ from the alphabet to the set $\mathcal{S}(\mathcal{H})$ of quantum state on the input system $\mathcal{H}'$. When the input element of the alphabet is i, the state on the output system (the output state) is $W(x) := \Lambda(\rho_x)$. In the following, the map from the input element in the alphabet x to the output state $W(x)$ is called a **classical-quantum channel**. Once a classical-quantum channel is given, we can discuss the problem to transmit a classical message via a noisy quantum channel. When the set of massages to be transmitted is the set $\{1, \ldots, M\}$, the encoder is given as a map ϕ from $\{1, \ldots, M\}$ to $\mathcal{X}$. A decoder is given as a POVM $\{Y_i\}_{i=1}^M$ taking values in the message $\{1, \ldots, M\}$. Hence, the triplet $(M, \phi, \{Y_i\}_{i=1}^M)$ is called a code and is denoted by Φ. The performance of the code Φ is characterized by the size $|\Phi| := M$ of the message and the averaged error probability $\varepsilon(\Phi) := 1 - \sum_{i=1}^M \operatorname{Tr} W(\phi(i))Y_i$ (Fig. 6.5).

In the following, we address the problem to transmit the classical message via n uses of the same classical-quantum channel. Hence, we focus on the classical-quantum channel $W^{(n)}(\vec{x}) := W(x_1) \otimes \cdots \otimes W(x_n)$ from the alphabet $\mathcal{X}^n$ to the quantum system $\mathcal{H}^{\otimes n}$. Using the **quantum transmission information** $I(\boldsymbol{p}, W)$, we can describe the rate of the transmittable message (the transmission rate) as follows. (For example, a proof is available in [45, Chap. 4].)

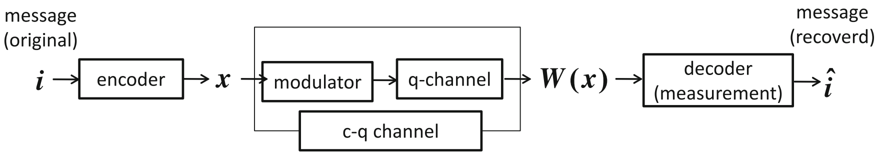


Fig. 6.5 Classical-quantum channel coding

Theorem 6.4 (Holevo [71], Schumacher, Westmoreland [113]) *Let $\{\Phi_n\}$ be a sequence of codes for the above defined sequence $\{W^{(n)}\}$ of classical-quantum channels. Then, the optimal performance can be characterized as follows.*

$$C(W) := \sup_{\{\Phi_n\}}\{\limsup_{n\to\infty}\frac{1}{n}\log|\Phi_n| | \varepsilon(\Phi_n)\to 0\} = \max_{p} I(\boldsymbol{p}, W). \tag{6.80}$$

Due to the above definition, $C(W)$ can be regarded as the supremum of the transmission rate under the constraint that the averaged error probability goes to zero asymptotically. Unfortunately, it is impossible to construct, independently of the classical-quantum channel W, a code that realizes the transmission rate $C(W)$ under the constraint that the averaged error probability asymptotically goes to zero. Instead of this requirement, it is possible to construct, independently of the classical-quantum channel W, a code whose transmission rate is $I(\boldsymbol{p}, W)$ under the same constraint. Such a sequence of codes is called a universal channel code for classical-quantum channel. There exists a sequence $\{\Phi_n\}_{n=1}^{\infty}$ of codes such that the relations

$$\lim_{n\to\infty}\frac{-1}{n}\log\varepsilon[\Phi_n, W] \geq \max_{0\leq t\leq 1}\frac{\phi_{W,\boldsymbol{p}}(t)-tR}{1+t} \tag{6.81}$$

$$\lim_{n\to\infty}\frac{1}{n}\log|\Phi_n| = R \tag{6.82}$$

hold for any probability distribution $\boldsymbol{p}$ on $\mathcal{X}$, where $\phi_{W,\boldsymbol{p}}(t)$ is defined to be $-(1-t)\log\mathrm{Tr}(\sum_{x=1}^{k}p(x)W(x)^{1-t})^{\frac{1}{1-t}}$ in (2.83) [55]. As shown in the end of Sect. 2.8, $\phi_{W,\boldsymbol{p}}(t)$ is a concave function for $t\in(0,1)$. Since the derivative of $\phi_{W,\boldsymbol{p}}(t)$ satisfies that $\phi'_{W,\boldsymbol{p}}(0) = I(\boldsymbol{p}, W)$, when the transmission rate R is smaller than the quantum transmission information $I(\boldsymbol{p}, W)$, there exists a real number $t\in(0,1)$ such that $\phi_{W,\boldsymbol{p}}(t)-tR>0$, which implies that $\max_{0\leq t\leq 1}\frac{\phi_{W,\boldsymbol{p}}(t)-tR}{1+t}>0$. So, the averaged error probability goes to zero.

6.7.2 Construction of Code

To construct a universal channel code satisfying (6.81) and (6.82), we extend the concept of type introduced in Sect. 6.1 to the case with the conditional distribution. For this purpose, we consider an probability space $\mathcal{Z} := \{z_1,\ldots,z_l\}$ composed of finite elements in addition to the input alphabet $\mathcal{X} = \{x_1,\ldots,x_k\}$. Given a type $\boldsymbol{n} = (n_1,\ldots,n_k)\in\mathcal{T}_n(\mathcal{X})$ on $\mathcal{X}$, the tuple of k types on $\mathcal{Z}$, $\boldsymbol{V} = (\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k)\in\mathcal{T}_{n_1}(\mathcal{Z})\times\cdots\times\mathcal{T}_{n_k}(\mathcal{Z})$, is called a conditional type associated with $\boldsymbol{n}\in\mathcal{T}_n(\mathcal{X})$. We denote the set of conditional types on $\mathcal{Z}$ associated with $\boldsymbol{n}\in\mathcal{T}_n(\mathcal{X})$ by $V(\boldsymbol{n},\mathcal{Z})$. In particular, a conditional type $\boldsymbol{V}$ can be naturally regarded as a type on $\mathcal{X}\times\mathcal{Z}$. Given an element $\vec{x}\in\mathcal{X}^n$ and a conditional type $\boldsymbol{V}$ on $\mathcal{Z}$ associated with $\boldsymbol{n}(\vec{x})$, $(\frac{\boldsymbol{v}_1}{n_1},\ldots,\frac{\boldsymbol{v}_k}{n_k})$

is also a conditional type and is denoted by $\frac{\boldsymbol{V}}{\boldsymbol{n}}$. For $\boldsymbol{V} \in V(\boldsymbol{n}(\vec{x}), \mathcal{Z})$, we define the set $T_{\boldsymbol{V}}(\vec{x}) := \{\vec{y} \in \mathcal{Z}^n | (x_1, y_1), \dots, (x_n, y_n) \in T_{\boldsymbol{V}}\}$, which is called V shell.

For the construction of our code, we prepare the following lemma.

Lemma 6.11 (Packing Lemma [25]) *Given an arbitrary distribution $\boldsymbol{p} \in \mathcal{T}_n(\mathcal{X})$ and a real number $R < H(\boldsymbol{p})$, we define $M_n := e^{n(R-\frac{1}{\sqrt{n}})}$. There exist M_n elements $\mathcal{M}_n := \{\vec{x}_1, \dots, \vec{x}_{M_n}\}$ of $T_{n\boldsymbol{p}} \subset \mathcal{X}^n$ such that*

$$|T_{\boldsymbol{V}}(\vec{x}) \cap (\mathcal{M}_n \setminus \{\vec{x}\})| \le |T_{\boldsymbol{V}}(\vec{x})| e^{-n(H(\boldsymbol{p})-R)}$$

for $\forall \vec{x} \in \mathcal{M}_n$ and $\forall \boldsymbol{V} \in V(\boldsymbol{n}(\vec{x}), \mathcal{X})$.

Then, we choose a code ϕ such that the image of ϕ is $\mathcal{M}_n$ given in the above lemma. Given an element $\vec{x} \in \mathcal{X}^n$, we define the subgroup $S_{\vec{x}} := \{g \in S_n | g(\vec{x}) = \vec{x}\}$ of the permutation S_n. For another element $\vec{x}' \neq \vec{x} \in \mathcal{X}^n$, we choose $\boldsymbol{V} \in V(\boldsymbol{n}(\vec{x}), \mathcal{X})$ such that $\vec{x}' \in T_{\boldsymbol{V}}(\vec{x})$. Then, the transition matrix corresponding to the conditional distribution $\frac{\boldsymbol{V}}{\boldsymbol{n}(\vec{x})}$ is not the identity matrix. Since $\vec{x} \notin T_{\boldsymbol{V}}(\vec{x})$, due to Lemma 6.11, the uniform distribution $p_{\mathcal{M}_n}$ on $\mathcal{M}_n$ satisfies

$$\begin{aligned}\sum_{g \in S_{\vec{x}}} \frac{1}{|S_{\vec{x}}|} p_{\mathcal{M}_n} \circ g(\vec{x}') = \frac{|T_{\boldsymbol{V}}(\vec{x}) \cap \mathcal{M}_n|}{|T_{\boldsymbol{V}}(\vec{x})|} \frac{1}{|\mathcal{M}_n|} = \frac{|T_{\boldsymbol{V}}(\vec{x}) \cap (\mathcal{M}_n \setminus \{\vec{x}\})|}{|T_{\boldsymbol{V}}(\vec{x})| \cdot |\mathcal{M}_n|} \\ \le e^{-nH(\boldsymbol{p})} e^{\sqrt{n}} = \boldsymbol{p}^n(\vec{x}') e^{\sqrt{n}},\end{aligned} \tag{6.83}$$

where the first equation follows from the fact that $g(\vec{x})$ is chosen according to the uniform distribution on $T_{\boldsymbol{V}}(\vec{x})$ and the final equation follows from the equation $e^{-nH(\boldsymbol{p})} = \boldsymbol{p}^n(\vec{x}')$ for $\vec{x}' \in T_{\boldsymbol{p}}$.

Next, for a POVM for the decoding, we construct a density matrix $\rho_{\vec{x}}$ that uniformly approximates $W^{(n)}(\vec{x})$ for $\vec{x} \in \mathcal{X}^n$. For simplicity, we consider the case when $\vec{x}' = (\underbrace{1, \dots, 1}_{m_1}, \underbrace{2, \dots, 2}_{m_2}, \dots, \underbrace{k, \dots, k}_{m_k})$. In this case, we define $\rho_{\vec{x}'} := \rho_{U,m_1} \otimes \rho_{U,m_2} \otimes \cdots \otimes \rho_{U,m_k}$. For a general element $\vec{x} \in \mathcal{X}^n$, we define $\rho_{\vec{x}}$ as the permutation of $\rho_{\vec{x}'}$ with the above special element $\vec{x}'$ satisfying $\boldsymbol{n}(\vec{x}) = \boldsymbol{n}(\vec{x}')$. Then, the relation (6.40) guarantees that $\rho_{\vec{x}}$ is commutative with $\rho_{U,n}$. Further, the relation (6.41) implies that

$$(n+1)^{\frac{k(r+2)(r-1)}{2}} \rho_{\vec{x}} \ge W_n(\vec{x}). \tag{6.84}$$

Then, for an arbitrary real number $C_n > 0$, we define the projection $P(\vec{x}) := \{\rho_{\vec{x}} - C_n \rho_{U,n} \ge 0\}$ and define the decoder as

$$Y_{\vec{x}'} := \sqrt{\sum_{\vec{x} \in \mathcal{M}_n} P(\vec{x})}^{-1} P(\vec{x}') \sqrt{\sum_{\vec{x} \in \mathcal{M}_n} P(\vec{x})}^{-1}.$$

Here, we notice that the code $\Phi_{U,n}(\boldsymbol{p}, R) := (e^{nR-\sqrt{n}}, \mathcal{M}_n, \{Y_{\vec{x}}\}_{\vec{x}\in\mathcal{M}_n})$ does not depend on the classical-quantum channel W and depends only on the probability distribution p on the input alphabet.

Exercise 6.15 List up all of sequences in $(\{1,2\} \times \{1,2\})^5$ corresponding to the conditional type $((1,1),(1,2))$.

6.7.3 Evaluation of Performance

First, we prepare an important inequality for evaluation of the averaged error probability [45, Chap. 4].

Lemma 6.12 *When two Hermitian matrices S and T satisfy $I \ge S \ge 0$ and $T \ge 0$, the matrix inequality $I - \sqrt{S+T}^{-1} S \sqrt{S+T}^{-1} \le 2(I-S) + 4T$ holds.*

This inequality implies that

$$I - Y_{\vec{x}'} \le 2(I - P(\vec{x}')) + 4 \sum_{\vec{x}(\neq \vec{x}')\in\mathcal{M}_n} P(\vec{x}).$$

Hence, we can evaluate the averaged error probability of the code $\Phi_{U,n}(\boldsymbol{p}, R)$ under the classical-quantum channel $W^{(n)}$ as

$$\begin{aligned}
&\frac{1}{|\mathcal{M}_n|} \sum_{\vec{x}'\in\mathcal{M}_n} \operatorname{Tr} W^{(n)}(\vec{x}')(I - Y_{\vec{x}'}) \\
\le &\frac{2}{|\mathcal{M}_n|} \sum_{\vec{x}'\in\mathcal{M}_n} \operatorname{Tr} W^{(n)}(\vec{x}')(I - P(\vec{x}')) \\
&+ \frac{4}{|\mathcal{M}_n|} \sum_{\vec{x}'\in\mathcal{M}_n} \operatorname{Tr} W^{(n)}(\vec{x}') \sum_{\vec{x}(\neq \vec{x}')\in\mathcal{M}_n} P(\vec{x}) \\
= &\frac{2}{|\mathcal{M}_n|} \sum_{\vec{x}\in\mathcal{M}_n} \operatorname{Tr} W^{(n)}(\vec{x})(I - P(\vec{x})) \\
&+ 4 \operatorname{Tr} \left[\sum_{\vec{x}\in\mathcal{M}_n} P(\vec{x}) \left(\frac{1}{|\mathcal{M}_n|} \sum_{\vec{x}'(\neq\vec{x})\in\mathcal{M}_n} W^{(n)}(\vec{x}') \right) \right].
\end{aligned} \tag{6.85}$$

Since the density matrix $\rho_{\vec{x}}$ is commutative with the density matrix $\rho_{U,n}$, for $0 \le t \le 1$, we have

$$(I - P(\vec{x})) = \{\rho_{\vec{x}} - C_n \rho_{U,n} < 0\} \le \rho_{\vec{x}}^{-t} C_n^t \rho_{U,n}^t. \tag{6.86}$$

Since the density matrix $\rho_{\vec{x}}$ is also commutative with the density matrix $W^{(n)}(\vec{x})$, the matrix $W^{(n)}(\vec{x})\rho_{\vec{x}}^{-t}$ is a Hermitian matrix. Further, the inequality (6.84) implies that

$$W^{(n)}(\vec{x})\rho_{\vec{x}}^{-t} \le (n+1)^{\frac{kt(r+2)(r-1)}{2}} W^{(n)}(\vec{x})^{1-t}. \tag{6.87}$$

The inequalities (6.86) and (6.87) yield

$$\begin{aligned} &\operatorname{Tr} W^{(n)}(\vec{x})(I - P(\vec{x})) \le \operatorname{Tr} W^{(n)}(\vec{x})\rho_{\vec{x}}^{-t}\rho_{U,n}^{t} C_n^t \\ \le &(n+1)^{\frac{kt(r+2)(r-1)}{2}} C_n^t \operatorname{Tr} W^{(n)}(\vec{x})^{1-t}\rho_{U,n}^{t}. \end{aligned} \tag{6.88}$$

Since the value $\operatorname{Tr} W^{(n)}(\vec{x})(I - P(\vec{x}))$ is invariant for the permutation on $\vec{x}$, the evaluation (6.4) for $\frac{n!}{(n\boldsymbol{p})!} = |T_{n\boldsymbol{p}}|$ guarantees

$$\boldsymbol{p}^n(\vec{x}) = e^{-nH(\boldsymbol{p})} \ge \frac{(n+1)^{-r}}{|T_{n\boldsymbol{p}}|} \tag{6.89}$$

for $\vec{x} \in T_{\boldsymbol{p}}$. Hence, we have

$$\begin{aligned} &\operatorname{Tr} W^{(n)}(\vec{x})(I - P(\vec{x})) = \frac{1}{|T_{\boldsymbol{p}}|} \sum_{\vec{x}' \in T_{\boldsymbol{p}}} \operatorname{Tr} W^{(n)}(\vec{x}')(I - P(\vec{x}')) \\ \le &(n+1)^r \sum_{\vec{x}' \in \mathcal{X}^n} \boldsymbol{p}^n(\vec{x}') \operatorname{Tr} W^{(n)}(\vec{x}')(I - P(\vec{x}')) && (6.90) \\ \le &(n+1)^{r + \frac{kt(r+2)(r-1)}{2}} C_n^t \operatorname{Tr}\Big(\sum_{\vec{x}' \in \mathcal{X}^n} \boldsymbol{p}^n(\vec{x}') W^{(n)}(\vec{x}')^{1-t} \Big)\rho_{U,n}^{t} && (6.91) \\ \le &(n+1)^{r + \frac{kt(r+2)(r-1)}{2}} C_n^t \max_{\sigma} \operatorname{Tr} \Big[\sum_{x \in \mathcal{X}} \boldsymbol{p}(x) W(x)^{1-t} \Big]^{\otimes n} \sigma^t \\ \le &(n+1)^{r + \frac{kt(r+2)(r-1)}{2}} C_n^t \left(\operatorname{Tr} \left(\Big[\sum_{x \in \mathcal{X}} \boldsymbol{p}(x) W(x)^{1-t} \Big]^{\otimes n} \right)^{\frac{1}{1-t}} \right)^{1-t} && (6.92) \\ = &(n+1)^{r + \frac{kt(r+2)(r-1)}{2}} C_n^t \left(\operatorname{Tr} \left(\sum_{x \in \mathcal{X}} \boldsymbol{p}(x) W(x)^{1-t} \right)^{\frac{1}{1-t}} \right)^{n(1-t)} \\ = &(n+1)^{r + \frac{kt(r+2)(r-1)}{2}} C_n^t e^{-n\phi_{W,\boldsymbol{p}}(t)}, && (6.93) \end{aligned}$$

where the inequalities (6.90), (6.91), and (6.92) follow from (6.89), (6.88), and (2.24), respectively.

Next, we evaluate the second term in (6.85) by using the invariance with respect to the subgroup $S_{\vec{x}}$ as

$$
\begin{aligned}
&\operatorname{Tr}\left[P(\vec{x})\left(\frac{1}{|\mathcal{M}_n|}\sum_{\vec{x}'(\neq\vec{x})\in\mathcal{M}_n} W^{(n)}(\vec{x}')\right)\right]\\
=&\operatorname{Tr}\left[P(\vec{x})\sum_{\vec{x}'(\neq\vec{x})\in\mathcal{M}_n} p_{\mathcal{M}_n}(\vec{x}')W^{(n)}(\vec{x}')\right]\\
=&\operatorname{Tr}\left[P(\vec{x})\sum_{s\in S_{\vec{x}}}\frac{1}{|S_{\vec{x}}|}\sum_{\vec{x}'(\neq\vec{x})\in\mathcal{M}_n} p_{\mathcal{M}_n}(\vec{x}')V_s W^{(n)}(\vec{x}')V_s^*\right]\\
=&\operatorname{Tr}\left[P(\vec{x})\sum_{\vec{x}'(\neq\vec{x})\in\mathcal{M}_n}\sum_{s\in S_{\vec{x}}}\frac{1}{|S_{\vec{x}}|}p_{\mathcal{M}_n}\circ s^{-1}(\vec{x}')W^{(n)}(\vec{x}')\right]\\
\le&\operatorname{Tr}\left[P(\vec{x})\sum_{\vec{x}'(\neq\vec{x})\in\mathcal{M}_n}\boldsymbol{p}^n(\vec{x}')e^{\sqrt{n}}W^{(n)}(\vec{x}')\right] &(6.94)\\
=&e^{\sqrt{n}}\operatorname{Tr}\left[P(\vec{x})W_{\boldsymbol{p}}^{\otimes n}\right]\\
\le&e^{\sqrt{n}}\operatorname{Tr}\left[P(\vec{x})(n+1)^{\frac{(r+2)(r-1)}{2}}\rho_{U,n}\right] &(6.95)\\
\le&e^{\sqrt{n}}\operatorname{Tr}\left[P(\vec{x})(n+1)^{\frac{(r+2)(r-1)}{2}}C_n^{-1}\rho_{\vec{x}}\right] &(6.96)\\
\le&e^{\sqrt{n}}\operatorname{Tr}\left[(n+1)^{\frac{(r+2)(r-1)}{2}}C_n^{-1}\rho_{\vec{x}}\right]=e^{\sqrt{n}}(n+1)^{\frac{(r+2)(r-1)}{2}}C_n^{-1}, &(6.97)
\end{aligned}
$$

where the inequalities (6.94), (6.95), and (6.96) follow from (6.83), (6.41), and the inequality $P(\vec{x})(\rho_{U,n}-C_n^{-1}\rho_{\vec{x}})\le 0$, respectively.

Now, given $t\in(0,1)$, $R'>R>0$, we choose $M_n:=e^{nR-\sqrt{n}}$ and $C_n:=e^{nR'}$. The inequalities (6.85), (6.93), and (6.97) yield the following evaluation of the averaged error probability as

$$
\begin{aligned}
&\varepsilon(\Phi_{U,n}(\boldsymbol{p},R),W)\\
\le&2(n+1)^{r+\frac{kt(r+2)(r-1)}{2}}e^{-n(\phi_{W,\boldsymbol{p}}(t)-tR')}+4(n+1)^{\frac{(r+2)(r-1)}{2}}e^{-n(R'-R)}. \qquad (6.98)
\end{aligned}
$$

Hence, the exponential decreasing rate is evaluated as

$$
\lim_{n\to\infty}\frac{-1}{n}\log\varepsilon(\Phi_{U,n}(\boldsymbol{p},R),W)\ge\min\{\phi_{W,\boldsymbol{p}}(t)-tR',R'-R\}.
$$

So, the remaining task is the choice of R'. For this choice, we introduce the set Θ of possible classical-quantum channel, and prepare the following lemma, which will be shown in the next subsection.

Lemma 6.13 ([63, Lemma 1]) *The function $\phi_{W,\boldsymbol{p}}(t)$ satisfies*

$$\begin{aligned}&\max_{R'} \inf_{W\in\Theta} \min(\max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR'), R'-R)\\ =& \inf_{W\in\Theta} \max_{R'} \min(\max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR'), R'-R)\\ =& \inf_{W\in\Theta} \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s)-sR). \end{aligned} \tag{6.99}$$

The maximum value $\max_{R'} \min(\max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR'), R'-R)$ *is attained when* $R' = R + \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s)-sR)$*. In particular, the maximum value* $\max_{R'} \inf_{W\in\Theta} \min(\max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR'), R'-R)$ *is attained when*

$$R' = R + \inf_{W\in\Theta} \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s)-sR). \tag{6.100}$$

That is, when we choose $|\mathcal{M}_n| := e^{nR-\sqrt{n}}$ and $C_n := e^{nR'}$ with (6.100), the relation

$$\lim_{n\to\infty} \frac{-1}{n} \log \varepsilon(\Phi_{U,n}(\boldsymbol{p},R), W) \ge \inf_{W\in\Theta} \max_{t\in(0,1)} \frac{\phi_{W,\boldsymbol{p}}(t)-tR}{1+t}$$

holds for any classical-quantum channel $W \in \Theta$.

6.7.4 Proof of Lemma 6.13

Now, we show Lemma 6.13 according to [63]. When $\inf_{W\in\Theta} I(\boldsymbol{p}, W) \le R$, all terms in (6.99) are zero. So, we can assume that $\inf_{W\in\Theta} I(\boldsymbol{p}, W) > R$ without loss of generality.

Firstly, we show that

$$\begin{aligned}&\max_{R'} \min(\max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR'), R'-R)\\ =& \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s)-sR). \end{aligned} \tag{6.101}$$

Since the function $R' \mapsto \max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR')$ is monotone decreasing and continuous and the function $R' \mapsto R'-R$ is monotone increasing and continuous, there exists a real number $R^* > R$ such that $\max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR^*) = R^*-R$. We can choose $s^* := \operatorname{argmax}_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR^*)$ because $\phi_{W,\boldsymbol{p}}(s)$ is a concave function. Here, we assume that $s^* \in (0,1)$. s^* satisfies $\frac{d\phi_{W,\boldsymbol{p}}(s)}{ds}|_{s=s^*} = R^*$. Since $(\phi_{W,\boldsymbol{p}}(s^*)-s^*R^*) = R^*-R$, we have $R^* = \frac{R+\phi_{W,\boldsymbol{p}}(s^*)}{1+s^*}$. So, $\max_{s\in[0,1]}(\phi_{W,\boldsymbol{p}}(s)-sR^*) = \frac{(\phi_{W,\boldsymbol{p}}(s^*)-s^*R)}{1+s^*}$ and $\frac{d\phi_{W,\boldsymbol{p}}(s)}{ds}|_{s=s^*} = \frac{R+\phi_{W,\boldsymbol{p}}(s^*)}{1+s^*}$.

Since the first derivative of $\frac{(\phi_{W,\boldsymbol{p}}(s)-sR)}{1+s}$ with respect to s is $\frac{(1+s)\frac{d\phi_{W,\boldsymbol{p}}(s)}{ds}-\phi_{W,\boldsymbol{p}}(s)-R}{(1+s)^2}$ and $(1+s)\frac{d\phi_{W,\boldsymbol{p}}(s)}{ds} - \phi_{W,\boldsymbol{p}}(s) - R$ is monotone decreasing for $s \in [0,1]$, $s_* := \mathrm{argmax}_{s\in[0,1]} \frac{(\phi_{W,\boldsymbol{p}}(s)-sR)}{1+s}$ satisfies the same condition $\frac{d\phi_{W,\boldsymbol{p}}(s)}{ds}|_{s=s_*} = \frac{R+\phi_{W,\boldsymbol{p}}(s_*)}{1+s_*}$. So, we find that $\max_{s\in[0,1]} \frac{(\phi_{W,\boldsymbol{p}}(s)-sR)}{1+s} = \frac{(\phi_{W,\boldsymbol{p}}(s^*)-s^*R)}{1+s^*}$. Thus, we obtain (6.101) when $s^* \in (0,1)$. When $s^* = 0$, we can show $s_* = 0$, which implies (6.101). Similarly, we can show (6.101) when $s^* = 1$.

Since

$$\begin{aligned}&\max_{R'} \inf_{W\in\Theta} \min(\max_{s\in[0,1]} (\phi_{W,\boldsymbol{p}}(s) - sR'), R' - R)\\ \le& \inf_{W\in\Theta} \max_{R'} \min(\max_{s\in[0,1]} (\phi_{W,\boldsymbol{p}}(s) - sR'), R' - R),\end{aligned} \tag{6.102}$$

it is sufficient to show there exists R' such that

$$\begin{aligned}&\inf_{W\in\Theta} \min(\max_{s\in[0,1]} (\phi_{W,\boldsymbol{p}}(s) - sR'), R' - R)\\ \ge& \inf_{W\in\Theta} \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s) - sR).\end{aligned} \tag{6.103}$$

We choose R' to be $R + \inf_{W\in\Theta} \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s) - sR)$. Given a parameter $W \in \Theta$, using the function $f(s,W) := \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s) - sR)$ and $s_W := \mathrm{argmax}_{s\in[0,1]} f(s,W)$, we have

$$f(s_W, W) \ge \inf_{W'\in\Theta} \max_{s'\in[0,1]} f(s', W'), \tag{6.104}$$

which implies that

$$\begin{aligned}&\max_{s\in[0,1]} (\phi_{W,\boldsymbol{p}}(s) - sR') \ge (\phi_{W,\boldsymbol{p}}(s_W) - s_W R')\\ =&(\phi_{W,\boldsymbol{p}}(s_W) - s_W R) + s_W \inf_{W'\in\Theta} \max_{s'\in[0,1]} f(s', W')\\ =&f(s_W, W) + s_W(f(s_W, W) - \inf_{W'\in\Theta} \max_{s'\in[0,1]} f(s', W')) \ge f(s_W, W)\\ \ge& \inf_{W\in\Theta} \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s) - sR) = R' - R.\end{aligned} \tag{6.105}$$

Thus,

$$\min(\max_{s\in[0,1]} (\phi_{W,\boldsymbol{p}}(s) - sR'), R' - R) = \inf_{W\in\Theta} \max_{s\in[0,1]} \frac{1}{1+s}(\phi_{W,\boldsymbol{p}}(s) - sR), \tag{6.106}$$

which implies (6.103). ■

Appendix A
Solutions of Exercises

Exercise 1.1
$P^{\rho}_{M}(1)=\frac{9}{16}$, $P^{\rho}_{M}(2)=\frac{1}{16}$, and $P^{\rho}_{M}(3)=\frac{3}{8}$.

Exercise 1.2
$A=E_1-E_2$, where $E_1=\frac{1}{2}\begin{pmatrix}1&1\\1&1\end{pmatrix}$ and $E_2=\frac{1}{2}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}$.

Exercise 1.3
We can calculate it in two ways. (1) $1\times\frac{1}{2}+(-1)\times\frac{1}{2}=0$. (2) $\mathrm{Tr}\,\rho A=0$.

Exercise 1.4
We can calculate it in two ways. (1) $1^2\times\frac{1}{2}+(-1)^2\times\frac{1}{2}=1$. (2) $\mathrm{Tr}\,\rho A^2=1$.

Exercise 1.5
Since $\begin{pmatrix}\frac{1}{2}&\frac{1}{4}\\\frac{1}{4}&\frac{1}{2}\end{pmatrix}=\frac{3}{4}\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}\end{pmatrix}+\frac{1}{4}\begin{pmatrix}\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{2}\end{pmatrix}$, we have

$$\begin{aligned}\begin{pmatrix}\frac{1}{2}&\frac{1}{4}\\\frac{1}{4}&\frac{1}{2}\end{pmatrix}^{-1/2}&=(\frac{3}{4})^{-1/2}\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}\end{pmatrix}+(\frac{1}{4})^{-1/2}\begin{pmatrix}\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{2}\end{pmatrix}\\&=\frac{2}{\sqrt{3}}\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}\end{pmatrix}+2\begin{pmatrix}\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{2}\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt{3}}+1&\frac{1}{\sqrt{3}}-1\\\frac{1}{\sqrt{3}}-1&\frac{1}{\sqrt{3}}+1\end{pmatrix}.\end{aligned}$$

Hence, the POVM $\{M_i\}_{i=1}^{3}$ is given as

$$\begin{aligned}M_1=&\begin{pmatrix}\frac{1}{\sqrt{3}}+1&\frac{1}{\sqrt{3}}-1\\\frac{1}{\sqrt{3}}-1&\frac{1}{\sqrt{3}}+1\end{pmatrix}\frac{1}{2}\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{3}}+1&\frac{1}{\sqrt{3}}-1\\\frac{1}{\sqrt{3}}-1&\frac{1}{\sqrt{3}}+1\end{pmatrix}\\=&\begin{pmatrix}\frac{1}{3}&\frac{1}{3}\\\frac{1}{3}&\frac{1}{3}\end{pmatrix},\end{aligned}$$

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8

$$\begin{aligned}
M_2 &= \begin{pmatrix} \frac{1}{\sqrt{3}}+1 & \frac{1}{\sqrt{3}}-1 \\ \frac{1}{\sqrt{3}}-1 & \frac{1}{\sqrt{3}}+1 \end{pmatrix} \frac{1}{4} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}}+1 & \frac{1}{\sqrt{3}}-1 \\ \frac{1}{\sqrt{3}}-1 & \frac{1}{\sqrt{3}}+1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{3}+\frac{1}{2\sqrt{3}} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{3}-\frac{1}{2\sqrt{3}} \end{pmatrix}, \\
M_3 &= \begin{pmatrix} \frac{1}{\sqrt{3}}+1 & \frac{1}{\sqrt{3}}-1 \\ \frac{1}{\sqrt{3}}-1 & \frac{1}{\sqrt{3}}+1 \end{pmatrix} \frac{1}{4} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}}+1 & \frac{1}{\sqrt{3}}-1 \\ \frac{1}{\sqrt{3}}-1 & \frac{1}{\sqrt{3}}+1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{3}-\frac{1}{2\sqrt{3}} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{3}+\frac{1}{2\sqrt{3}} \end{pmatrix}.
\end{aligned}$$

Exercise 1.6
Since $XX^\dagger = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the Schmidt rank is 1. The Schmidt coefficients are $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

Exercise 1.7

$$Y \otimes Z|X\rangle\rangle_{A,B} = \sum_{k,j} \sum_{k',j'} y_{k,k'} z_{j,j'} x_{k',j'} |k\rangle \otimes |j\rangle = |YXZ^T\rangle\rangle_{A,B}. \tag{A.1}$$

Exercise 1.8

$$\mathrm{Tr}_B\, |X\rangle\rangle_{A,B}\ {}_{A,B}\langle\langle Y| = \mathrm{Tr}_B \sum_{k,j} x_{k,j} |k\rangle \otimes |j\rangle \sum_{k',j'} \overline{y_{k',j'}} \langle k'| \otimes \langle j'| \tag{A.2}$$

$$= \sum_{k',k,j} x_{k,j} \overline{y_{k',j}} |k\rangle\langle k'| = XY^\dagger. \tag{A.3}$$

Exercise 1.9
$\sqrt{p}|0\rangle|0\rangle + \sqrt{1-p}|1\rangle|1\rangle$.

Exercise 1.10
Define the POVM $M_i := \overline{X} p_i \rho_i^T (X^T)^{-1}$. Using Exercise 1.8, we have

$$\begin{aligned}
&\mathrm{Tr}_B\, M_i |X\rangle\rangle\langle\langle X| = \mathrm{Tr}_B\, \overline{X} p_i \rho_i^T (X^T)^{-1} |X\rangle\rangle\langle\langle X| \\
&= \mathrm{Tr}_B\, |XX^{-1} p_i \rho_i (X^\dagger)^{-1}\rangle\rangle\langle\langle X| = XX^{-1} p_i \rho_i (X^\dagger)^{-1} X^\dagger = p_i \rho_i.
\end{aligned}$$

Exercise 1.11
Choose the system $\mathcal{H}_B$ as the space spanned by the CONSs $\{|i\rangle\}_i$. Then, we define $|X\rangle\rangle$ as $\sum_i \sqrt{p_i}|x_i\rangle|i\rangle$. Define the PVM E as $E_i := |i\rangle\langle i|$. So, we can check the desired condition.

Exercise 1.12
Since $\mathrm{Tr}_A \frac{1}{d}|U_i\rangle\rangle\langle\langle U_i||\rho_A \otimes I = \mathrm{Tr}_A(\rho_A \otimes I)\frac{1}{d}|U_i\rangle\rangle\langle\langle U_i| = \mathrm{Tr}_A \frac{1}{d}|\rho_A U_i\rangle\rangle\langle\langle U_i| = \frac{1}{d}U_i^\dagger \rho_A U_i$, we have $\mathrm{Tr}\, \frac{1}{d}|U_i\rangle\rangle\langle\langle U_i|\rho_A \otimes \rho_{\mathrm{mix},B} = \mathrm{Tr}_B \frac{1}{d}U_i^\dagger \rho_A U_i \rho_{\mathrm{mix},B} = \frac{1}{d^2}$.

Exercise 2.1
Let x_1 and x_2 be two elements of the domain of f. For an arbitrary number $\lambda \in [0, 1]$, the linearity of f implies that $f(\lambda x_1(1-\lambda)x_2) = \lambda f(x_1)+(1-\lambda)f(x_2)$. Since g is a concave function, we have $g(\lambda f(x_1)+(1-\lambda)f(x_2)) \le \lambda g(f(x_1))+(1-\lambda)g(f(x_2))$. Combining these two inequalities, we obtain $g(f(\lambda x_1(1-\lambda)x_2)) \le \lambda g(f(x_1)) + (1-\lambda)g(f(x_2))$.

Exercise 2.2
For a square matrix X, we choose two Hermitian matrices X_1 and X_2 such that $X = X_1 + X_2 i$. Then, we have $|\mathrm{Tr}\, X\rho|^2 = (\mathrm{Tr}\, X_1\rho)^2 + (\mathrm{Tr}\, X_2\rho)^2$. Exercise 2.1 guarantees the convexity of the functions $(\mathrm{Tr}\, X_1\rho)^2$ and $(\mathrm{Tr}\, X_2\rho)^2$. Since a sum of convex functions is a convex function, we obtain the desired statement.

Exercise 2.3
We fix two elements $x_1, x_2 \in \mathcal{X}$ and $\lambda \in (0, 1)$. For any $\epsilon > 0$, we choose a' such that $f_{a'}(\lambda x_1 + (1-\lambda)x_2) \ge \sup_a f_a(\lambda x_1 + (1-\lambda)x_2) - \epsilon$. Since

$$\begin{aligned} f_{a'}(\lambda x_1 + (1-\lambda)x_2) &\le \lambda f_{a'}(x_1) + (1-\lambda)f_{a'}(x_2) \\ &\le \lambda(\sup_a f_a(x_1)) + (1-\lambda)(\sup_a f_a(x_2)), \end{aligned}$$

we have

$$\sup_a f_a(\lambda x_1 + (1-\lambda)x_2) - \epsilon \le \lambda(\sup_a f_a(x_1)) + (1-\lambda)(\sup_a f_a(x_2)).$$

Since $\epsilon > 0$ is arbitrary, we have

$$\sup_a f_a(\lambda x_1 + (1-\lambda)x_2) \le \lambda(\sup_a f_a(x_1)) + (1-\lambda)(\sup_a f_a(x_2)),$$

which is the desired statement.

Exercise 2.4
Let $\{p_i\}$ be an arbitrary distribution on $\{1, \ldots, d\}$. So, we have $\sum_{i=1}^k p_i^\downarrow \ge \frac{k}{d}$, which implies the desired statement.

Exercise 2.5

$$
\begin{aligned}
& I_\rho(A:B) = H(\rho_A) + H(\rho_B) - H(\rho) \\
= & -\operatorname{Tr}\rho(\log\rho_A)\otimes I_B - \operatorname{Tr}\rho I_A \otimes (\log\rho_B) + \operatorname{Tr}\rho\log\rho \\
= & -\operatorname{Tr}\rho\log(\rho_A\otimes\rho_B) + \operatorname{Tr}\rho\log\rho = D(\rho\|\rho_A\otimes\rho_B).
\end{aligned}
$$

Exercise 2.6

We focus on the eigenvectors of the density matrix ρ whose eigenvalue is not λ. We move such eigenvectors cyclically and obtain another density matrix. We take the average among such density matrices with equal weight. Such the averaged density matrix has von Neumann entropy $(1-\lambda)\log\dim\mathcal{H} + h(\lambda)$. Hence, the concavity of von Neumann entropy yields the desired statement.

Exercise 2.7

The (2.44) can be shown as follows.

$$
\begin{aligned}
& D(\rho\|\sigma_A\otimes\sigma_B) = \operatorname{Tr}\rho(\log\rho - \log\sigma_A\otimes\sigma_B) \\
= & \operatorname{Tr}\rho(\log\rho - (\log\sigma_A)\otimes I_B - I_A\otimes(\log\sigma_B)) \\
= & \operatorname{Tr}\rho(\log\rho - (\log\sigma_A)\otimes I_B - I_A\otimes(\log\rho_B)) \\
& + \operatorname{Tr}\rho(I_A\otimes(\log\rho_B) - I_A\otimes(\log\sigma_B)) \\
= & \operatorname{Tr}\rho(\log\rho - \log(\sigma_A\otimes\rho_B)) + \operatorname{Tr}\rho_B(\log\rho_B - \log\sigma_B) \\
= & D(\rho\|\sigma_A\otimes\rho_B) + D(\rho_B\|\sigma_B).
\end{aligned}
\tag{A.4}
$$

Exercise 2.8

We denote the spectral decomposition E by $\{E_i\}$. We define a unitary representation $\mathsf{f}(\theta_1,\ldots,\theta_n) := \sum_{j=1}^n e^{i\theta_j}E_j$ of the group $\mathbb{R}^n$. The pinching Λ_E with respect to the PVM E satisfies that $\Lambda_E(\rho) = \overline{\rho}$. Due to Theorem 2.9, the image $\Lambda_E(\rho)$ of the pinching is the point closest to ρ among the invariant density matrices with respect to this representation in the sense of relative entropy. That is, Theorem 2.9 yields (2.45).

Exercise 2.9

Since $\langle\Phi|F_m\otimes I|\Phi\rangle = \operatorname{Tr}F_m\rho$, we have

$$
\begin{aligned}
& F_e^2(\rho,\Lambda) = \langle\Phi|(\Lambda\otimes id)(|\Phi\rangle\langle\Phi|)|\Phi\rangle \\
= & \langle\Phi|\sum_m (F_m\otimes I)|\Phi\rangle\langle\Phi|(F_m\otimes I)^\dagger|\Phi\rangle \\
= & \sum_m |\langle\Phi|F_m\otimes I|\Phi\rangle|^2 = \sum_m |\operatorname{Tr}F_m\rho|^2.
\end{aligned}
\tag{A.5}
$$

Exercise 2.10
Since $\langle \Phi|(F_m \otimes I)|\Phi\rangle = \mathrm{Tr}\, F_m\rho$, we have

$$\begin{aligned} & F_e^2(\rho, \Lambda) = \langle \Phi|(\Lambda \otimes id)(|\Phi\rangle\langle\Phi|)|\Phi\rangle \\ =& \langle \Phi| \sum_m (F_m \otimes I)|\Phi\rangle\langle\Phi|(F_m \otimes I)^\dagger|\Phi\rangle \\ =& \sum_m \langle \Phi|(F_m \otimes I)|\Phi\rangle\langle\Phi|(F_m \otimes I)^\dagger|\Phi\rangle = \sum_m |\mathrm{Tr}\, F_m\rho|^2. \end{aligned}$$

Exercise 2.11
Based on Exercise 1.11, we choose the reference system $\mathcal{H}_R$, a purification $|X\rangle\rangle$ of ρ on $\mathcal{H}_B$, and a PVM $E = \{E_i\}_i$ on $\mathcal{H}_R$ such that $p_i|x_i\rangle\langle x_i = \mathrm{Tr}_B\, E_i|X\rangle\rangle\langle\langle X|$. Hence,

$$\begin{aligned} & F_e^2(\rho, \Lambda) = \langle \Phi|(\Lambda \otimes id)(|\Phi\rangle\langle\Phi|)|\Phi\rangle \\ \le & F^2(\Lambda_E(|\Phi\rangle\langle\Phi|), \Lambda_E((id \otimes M_i)(\Lambda \otimes id)(|\Phi\rangle\langle\Phi|))) \\ = & \sum_j p_j\langle x_j|\Lambda(|x_j\rangle\langle x_j|)|x_j\rangle. \end{aligned} \tag{A.6}$$

Exercise 2.12
$H_{1+s}(\rho_{\boldsymbol{x}})$ is calculated to be $\frac{1}{s}\log((\frac{1+\|\boldsymbol{x}\|}{2})^{1+s} + (\frac{1-\|\boldsymbol{x}\|}{2})^{1+s})$.

Exercise 2.13
To show 2.82, it is enough to show that

$$D_{1+s}(\rho\|\rho_A \otimes \sigma_B) \ge D_{1+s}(\rho\|\rho_A \otimes \sigma_B^*(1+s)) \tag{A.7}$$

for any state σ_B. This inequality is equivalent to

$$e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B)} \le e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B^*(1+s))} \tag{A.8}$$

for $s \in (-1, 0)$ and

$$e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B)} \ge e^{sD_{1+s}(\rho\|\rho_A\otimes\sigma_B^*(1+s))} \tag{A.9}$$

for $s \in (0, \infty)$.

(A.8) is equivalent to the following inequality

$$\mathrm{Tr}\, \rho^{1+s}\rho_A^{-s} \otimes \sigma_B^{-s} = \mathrm{Tr}_B(\mathrm{Tr}_A\, \rho^{1+s}\rho_A^{-s})\sigma_B^{-s} \le (\mathrm{Tr}_B(\mathrm{Tr}_A\, \rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}})^{1+s}. \tag{A.10}$$

We apply matrix Hölder inequality (2.21) to the case when $A = (\mathrm{Tr}_A\, \rho^{1+s}\rho_A^{-s})$, $B = \sigma_B^{-s}$, $p = \frac{1}{1+s} > 0$ and $q = \frac{1}{-s} > 0$ with $s \in (-1, 0)$. Since the RHS of (2.21) equals the RHS of (A.10), we obtain (A.10).

Similarly, (A.9) is equivalent to the following inequality

$$\mathrm{Tr}_B(\mathrm{Tr}_A\, \rho^{1+s}\rho_A^{-s})\sigma_B^{-s} \geq (\mathrm{Tr}_B(\mathrm{Tr}_A\, \rho^{1+s}\rho_A^{-s})^{\frac{1}{1+s}})^{1+s}. \tag{A.11}$$

We obtain (A.11) by applying matrix reverse Hölder inequality (2.22) to the case when $A = (\mathrm{Tr}_A\, \rho^{1+s}\rho_A^{-s})$, $B = \sigma_B^{-s}$, $p = \frac{1}{1+s} > 0$ and $q = \frac{1}{-s} > 0$ with $s \in (0, \infty)$.

Exercise 3.1
Due to Theorem 3.1, it is sufficient to show that $\mathrm{Tr}_B |\Psi\rangle\langle\Psi| \succ \mathrm{Tr}_B |\Phi\rangle\langle\Phi|$. This relation is trivial from the definition of majorization.

Exercise 3.2
Firstly, Bob applies the above measurement on the composite system $\mathcal{H}_B \otimes \mathcal{H}_{B'}$, and sends the outcome to Alice. Then, Alice applies unitary on the system $\mathcal{H}_A$ dependently of the measurement outcome. Then, the final state of the composite system $\mathcal{H}_A \otimes \mathcal{H}_C$ is a maximally entangled state.

Exercise 3.3
The commutation relation 3.1 guarantees that

$$\begin{aligned}
&(\mathsf{X}_{\mathbb{F}}(t_1) \otimes \cdots \otimes \mathsf{X}_{\mathbb{F}}(t_r))(\mathsf{Z}_{\mathbb{F}}(s_1) \otimes \cdots \otimes \mathsf{Z}_{\mathbb{F}}(s_r))\frac{1}{q^{r/2}}\sum_{\boldsymbol{x}\in\mathbb{F}_q^r}|\boldsymbol{x}\rangle \\
=&\omega_{\mathbb{F}}^{\sum_{i=1}^r trs_it_i}(\mathsf{Z}_{\mathbb{F}}(s_1) \otimes \cdots \otimes \mathsf{Z}_{\mathbb{F}}(s_r))(\mathsf{X}_{\mathbb{F}}(t_1) \otimes \cdots \otimes \mathsf{X}_{\mathbb{F}}(t_r))\frac{1}{q^{r/2}}\sum_{\boldsymbol{x}\in\mathbb{F}_q^r}|\boldsymbol{x}\rangle \\
=&\omega_{\mathbb{F}}^{\sum_{i=1}^r trs_it_i}(\mathsf{Z}_{\mathbb{F}}(s_1) \otimes \cdots \otimes \mathsf{Z}_{\mathbb{F}}(s_r))\frac{1}{q^{r/2}}\sum_{\boldsymbol{x}\in\mathbb{F}_q^r}|\boldsymbol{x}\rangle.
\end{aligned}$$

Hence, the definition (3.12) guarantees the desired statement.

Exercise 3.4
Using $|+\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle := \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, we have $|u_0^\zeta\rangle = \frac{1}{\sqrt{2}}(|0\rangle |+\rangle^{\otimes(r-1)} + |1\rangle|-\rangle^{\otimes(r-1)})$ when ζ is a star graph.

Exercise 3.5
We have $|u_0^\zeta\rangle = \frac{1}{2}(|0, 0, +\rangle + |0, 1, -\rangle + |1, 0, +\rangle - |1, 1, -\rangle)$ when ζ is 1D Cluster graph with $r = 3$.

Exercise 3.6
We have $|u_0^\zeta\rangle = \frac{1}{2}(|0, 0, +\rangle + |0, 1, -\rangle + |1, 0, -\rangle - |1, 1, +\rangle)$ when ζ is Ring graph with $r = 3$.

Exercise 3.7
Using 3.25

$$
\begin{aligned}
E_{G,2}(|\Phi\rangle\langle\Phi|) &= -\log \max_{|a,b\rangle\langle a,b|} \langle a,b|\Phi\rangle\langle\Phi|a,b\rangle \\
&= -\log \max_{|a,b\rangle\langle a,b|} |\langle a| \sum_i \sqrt{p_i}|i\rangle\langle i|b\rangle|^2 \\
&= -\log|\max_i \sqrt{p_i}|^2 = -\log \max_i p_i .
\end{aligned}
$$

Exercise 3.8
Consider the bipartite cut between the vertex connected to all of other vertex (1st vertex) and the remaining vertexes. Under this bipartite cut, we have $E_{G,2:1,(2,\ldots,r)} = \log 2$. Then, we find that $E_{G,2}$ is not greater than $\log 2$ due to (3.26). We employ the concrete form obtained in Exercise 3.4. This bound can be attained when we choose the separable state $|0\rangle|+\rangle^{\otimes(r-1)}$. So, we conclude that $E_{G,2} = \log 2$.

Exercise 3.9
We employ the concrete form obtained in Exercise 3.5. Consider the bipartite cut between the first vertex and the remaining vertexes. Under this bipartite cut, we have $E_{G,2:1,(2,3)} = \log 2$. Then, we find that $E_{G,2}$ is not greater than $\log 2$ due to (3.26). This bound can be attained when we choose the separable state $|+,0,+\rangle$. So, we conclude that $E_{G,2} = \log 2$.

Exercise 3.10
We employ the concrete form obtained in Exercise 3.6. To optimize the quantity $|\langle a,b,c|u_0^\zeta\rangle|$, we can restrict our vector $|a\rangle, |b\rangle, |c\rangle$ to real vectors. So, without loss of generality, we can assume that $|a\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$. Then,

$$
\begin{aligned}
\langle a|u_0^\zeta\rangle &= \frac{1}{2}(\cos\theta(|0,+\rangle + |1,-\rangle)\sin\theta(|0,-\rangle - |1,+\rangle)) \\
&= \frac{1}{2}(|0\rangle(\cos\theta|+\rangle + \sin\theta|-\rangle)|1\rangle(\cos\theta|-\rangle - \sin\theta|+\rangle)).
\end{aligned}
$$

Since $\cos\theta|+\rangle + \sin\theta|-\rangle$ is orthogonal to $\cos\theta|-\rangle - \sin\theta|+\rangle$, $\max_{b,c}\langle a,b,c|u_0^\zeta\rangle = \frac{1}{2}$. So, $\max_{a,b,c}\langle a,b,c|u_0^\zeta\rangle = \frac{1}{2}$. That is, $E_{G,2} = 2\log 2$.

Exercise 4.1
Since the maximum value is attained when M is $M_{d_\lambda|\lambda;\lambda\rangle\langle\lambda;\lambda|}$, Example 4.1 guarantees that the desired maximum value is

$$\begin{aligned}&\int_0^1 \epsilon \frac{dP_{\lambda_1,\lambda,\epsilon}}{d\epsilon} d\epsilon = \int_0^1 \epsilon \frac{d\epsilon^{\frac{2\lambda+1}{2\lambda_1}}}{d\epsilon} d\epsilon = \int_0^1 \epsilon \frac{2\lambda+1}{2\lambda_1}\epsilon^{\frac{2\lambda+1}{2\lambda_1}-1} d\epsilon \\ =& \int_0^1 \frac{2\lambda+1}{2\lambda_1}\epsilon^{\frac{2\lambda+1}{2\lambda_1}} d\epsilon = \frac{\frac{2\lambda+1}{2\lambda_1}}{\frac{2\lambda+1}{2\lambda_1}+1} = \frac{2\lambda+1}{2\lambda+2\lambda_1+1}.\end{aligned}$$

Exercise 4.2
Using (4.23), we have

$$\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \frac{d_\lambda}{d_{2\lambda}} = \frac{2\lambda+1}{4\lambda+1} \cong \frac{1}{2} + \frac{1}{8\lambda}. \tag{A.12}$$

Exercise 4.3

Using (4.23), we have the same asymptotic characterization as (A.12).

Exercise 4.4

$$\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \frac{d_{\boldsymbol{\lambda}}}{d_{2\boldsymbol{\lambda}}} = \frac{\lambda_Z}{2\lambda_Z} = \frac{1}{2} \tag{A.13}$$

Exercise 4.5

$$\begin{aligned}&\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \frac{d_{\boldsymbol{\lambda}}}{d_{2\boldsymbol{\lambda}}} = \frac{(n+r-1)!(r-1)!(2n)!}{(r-1)!n!(2n+r-1)!} = \frac{(n+r-1)!(2n)!}{n!(2n+r-1)!} \\ =& \frac{n+r-1}{2n+r-1}\cdots\frac{n+1}{2n+1} \\ \cong& \frac{1}{2^{r-1}}(1+\frac{r-1}{2n}+\frac{r-2}{2n}\cdots+\frac{1}{2n}) = \frac{1}{2^{r-1}}(1+\frac{r(r-1)}{4n})\end{aligned}$$

Exercise 4.6

$$\max_{\hat{M}} \mathcal{D}_F(\hat{M}) = \frac{d_{\boldsymbol{\lambda}}}{d_{2\boldsymbol{\lambda}}} = \frac{\lambda_Z^r}{(2\lambda_Z)^r} = \frac{1}{2^r}$$

Exercise 4.7 (4.24) can be calculated as follows.

$$\begin{aligned}&\frac{(2\lambda+1)(e^t-1)(e^{t(\lambda+\lambda_1+1)}-e^{-t(\lambda+\lambda_1)})}{(2(\lambda+\lambda_1)+1)(e^{t(\lambda_1+1)}-e^{-t\lambda_1})(e^{t(\lambda+1)}-e^{-t\lambda})} \\ \cong& \frac{(2\lambda+1)(1-e^{-t})}{2(\lambda+\lambda_1)+1} \cong \frac{1-e^{-t}}{2}(1+\frac{1}{4\lambda}).\end{aligned} \tag{A.14}$$

Exercise 4.8
The RHS (4.25) can be calculated in (A.14).

Exercise 4.9
The RHS of (4.26) equals $\frac{1-e^{-t}}{2}$.

Exercise 4.10
(4.37) shows that

$$\lim_{n\to\infty}\frac{1}{n}|\mathcal{F}^{-1}[\varphi_n](-nx)|^2 = \Big|\int_{-1}^{1}\frac{1}{\sqrt{2}}e^{-ixy}dy\Big|^2 = \Big|\frac{\sqrt{2}\sin x}{x}\Big|^2.$$

Exercise 4.11
$\cos\frac{\pi}{2n+2} \cong 1-\frac{1}{2}(\frac{\pi}{2n+2})^2+\frac{1}{24}(\frac{\pi}{2n+2})^4 \cong 1-\frac{\pi^2}{8}\frac{1}{n^2}+\frac{\pi^2}{4}\frac{1}{n^3}$.

Exercise 4.12
$\frac{1}{E}+\frac{1}{4E^3}$.

Exercise 5.1
When a coset contains $(a,0,0)$, it is $\{(a+b,b,b)\}_{b\in\mathbb{F}_q}$. The coset does not contain $(c,0,0)$ for $c\neq a$ nor $(0,a',0)$, $(0,0,a')$. So, we can choose the representatives $\{J(x)\}_{x\in\mathbb{F}_q}$ satisfying the requirement.

Exercise 5.2
Assume that the coset x contains (a,b,c). When $b=c$, $J_{\mathrm{ML}}(x)=(a-b,0,0)$. When $a=c$, $J_{\mathrm{ML}}(x)=(0,b-a,0)$. When $a=b$, $J_{\mathrm{ML}}(x)=(0,0,c-a)$. When $a>b>c$, $J_{\mathrm{ML}}(x)=(a-c,b-c,0)$. Other cases can be given in the same way.

Exercise 5.3
We have three patterns for correct decoding as follows.

(1) The noise is $(0,0,0)$. This probability is $(1-p)^3$.
(2) The noise is $(a,0,0)$, $(0,a,0)$, or $(0,0,a)$. This probability is $3(q-1)\frac{p}{q-1}(1-p)^2=3p(1-p)^2$.
(3) The noise is $(a,b,0)$, etc. Among this pattern, $q^2-1-3(q-1)=q^2-3q+2$ cases are correctable. Each case has probability $(\frac{p}{q-1})^2(1-p)=\frac{p^2(1-p)}{(q-1)^2}$. The probability of correctly decoding with this pattern is $(q^2-3q+2)\frac{p^2(1-p)}{(q-1)^2}=\frac{(q-2)p^2(1-p)}{q-1}$.

Totally, the whole correctly decoding probability is $(1-p)^3+3p(1-p)^2+\frac{(q-2)p^2(1-p)}{q-1}=(1-p)^2(1+2p)+\frac{(q-2)p^2(1-p)}{q-1}$.

Exercise 5.4
Then, the element of $C_{G,1}$ whose first, second, third, and fourth entries are zero is limited to $(0, 0, 0, 0, 0, 0, 0)^T$. Hence, we can choose representatives whose first, second, third, and fourth entries are zero. Since $\mathbb{F}_q^7/C_{G,1}$ is the three-dimensional space, all of elements of $\mathbb{F}_q^7/C_{G,1}$ are given as $[(0, 0, 0, 0, a, b, c)^T]_{a,b,c\in\mathbb{F}_q}$.

Exercise 5.5
Any non-zero element of $C_{G,1}$ has at least three non-zero entries. When $x \in \mathbb{F}_q^7$ has only one non-zero entry, any other element of $[x]$ has at least two non-zero entries. Hence, we can choose decoder satisfying the required condition.

Exercise 5.6
We have at least the following two patterns for correct decoding as follows.

(1) The noise is $(0, 0, 0)$. This probability is $(1 - p)^7$.
(2) The noise has only one non-zero entry. This probability is $7(q-1)\frac{p}{q-1}(1-p)^6 = 7p(1 - p)^6$.

Hence, the whole correctly decoding probability is larger than $(1 - p)^7 + 7p(1 - p)^6 = (1 - p)^6(1 + 6p)$.

Exercise 5.7
Since all of column vectors of G_2 are given as linear combinations of column vectors of G_1, the code space $C_{H,1}$ contains the code space $C_{H,2}$.

Exercise 5.8
Since $G_1G_3^T = 0$, we have $C_{G,1} \subset C_{G,3}^{\perp}$. Since the dimension of $C_{G,1}$ is 4 and the dimension of $C_{G,3}^{\perp}$ is $7 - 3 = 4$, $C_{G,1} = C_{G,3}^{\perp}$. Since $G_2G_4^T = 0$, we can show that $C_{G,2} = C_{G,4}^{\perp}$ in the same way.

Exercise 5.9
The desired statement can be shown in the same way as Exercise 5.6.

Exercise 5.10
The torsion condition is shown as follows. $C_{G,2} \subset C_{G,1} = C_{G,3}^{\perp}$.

Exercise 5.11
Due to Exercise 5.6 and Exercise 5.9, there exist classical decoders $\{\boldsymbol{s}_x\}_{x\in\mathbb{X}^r/C_{G,3}^{\perp}}$ and $\{\boldsymbol{t}_y\}_{y\in\mathbb{X}^r/C_{G,2}^{\perp}}$ such that

$$1 - \sum_{x\in\mathbb{X}^r/C_{G,3}^{\perp}} \mathrm{P}^{\mathbb{X}_1^r}(\boldsymbol{s}_x) \le 1 - (1 - p)^6(1 + 6p) \tag{A.15}$$

$$1 - \sum_{y\in\mathbb{X}^r/C_{G,2}^{\perp}} \mathrm{P}^{\mathbb{X}_2^r}(\boldsymbol{t}_y) \le 1 - (1 - p)^6(1 + 6p) \tag{A.16}$$

because $C_{G,1} = C_{G,3}^{\perp}$ and $C_{G,2} = C_{G,4}^{\perp}$. Combining both evaluations, we obtain the desired statement.

Exercise 6.1
(1, 1, 2, 2, 2, 2), (1, 2, 1, 2, 2, 2), (1, 2, 2, 1, 2, 2), (1, 2, 2, 2, 1, 2), (1, 2, 2, 2, 2, 1), (2, 1, 1, 2, 2, 2), (2, 1, 2, 1, 2, 2), (2, 1, 2, 2, 1, 2), (2, 1, 2, 2, 2, 1), (2, 2, 1, 1, 2, 2), (2, 2, 1, 2, 1, 2), (2, 2, 1, 2, 2, 1), (2, 2, 2, 1, 1, 2), (2, 2, 2, 1, 2, 1), (2, 2, 2, 2, 1, 1).

Exercise 6.2
When $n = 1$, the random variable $\boldsymbol{n}$ has covariance $C_{i,j}$. For general n, the random variable $\frac{\boldsymbol{n}}{n}$ is the sample mean of the above random variable with n independent trials. So, we can apply law of large number and the central limit theorem. So, we obtain Lemma 6.3.

Exercise 6.3
Since $\frac{\partial H(\boldsymbol{p})}{\partial p_i} = \log p_i - 1$, we have

$$\begin{aligned}
&\sum_{i,j} \frac{\partial H(\boldsymbol{p})}{\partial p_i} C_{i,j} \frac{\partial H(\boldsymbol{p})}{\partial p_j} \\
=&\sum_{i'} (\frac{\partial H(\boldsymbol{p})}{\partial p_{i'}})^2 p_{i'} - \sum_{i,j} \frac{\partial H(\boldsymbol{p})}{\partial p_i} p_i p_j \frac{\partial H(\boldsymbol{p})}{\partial p_j} \\
=&\sum_{i'} (\log p_{i'} - 1)^2 p_{i'} - \sum_{i,j} (\log p_i - 1) p_i p_j (\log p_j - 1) \\
=&(\sum_{i'} p_{i'} \log p_{i'}^2) - H(\boldsymbol{p}) + 1 - (H(\boldsymbol{p}) - 1)^2 \\
=&(\sum_{i'} p_{i'} \log p_{i'}^2) - H(\boldsymbol{p})^2 = V(\boldsymbol{p}).
\end{aligned}$$

Exercise 6.4
The relation (6.6) and Lemma 6.6 guarantee that the random variables $\sqrt{n}(H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p}))$ asymptotically obeys the normal distribution with average 0 and variance $\sum_{i,j} \frac{\partial H(\boldsymbol{p})}{\partial p_i} C_{i,j} \frac{\partial H(\boldsymbol{p})}{\partial p_j}$, which equals $V(\boldsymbol{p})$ as shown in (6.14). Further, as discussed in the end of Subsection 6.2.1, the differences $\sqrt{n}(H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p})) - \sqrt{n}(\frac{1}{n} \log \frac{n!}{\boldsymbol{n}!} - H(\boldsymbol{p}))$ and $\sqrt{n}(H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p})) - \sqrt{n}(\frac{1}{n} \log \dim \mathcal{V}_{\boldsymbol{n}} - H(\boldsymbol{p}))$ converges to zero in probability. So, the remaining random variables asymptotically obey the normal distribution with average 0 and variance $V(\boldsymbol{p})$.

Exercise 6.5
Recall (2.74) and (4.59) in [44] as

$$\begin{aligned}\dim \mathcal{U}_{\frac{n+k}{2},\frac{n-k}{2}}(S_n) &= \frac{k+1}{n+1}\binom{n+1}{\frac{n+k}{2}+1} \\ \chi_{\frac{n+k}{2},\frac{n-k}{2}}(\rho) &= \frac{p^{\frac{n+k}{2}+1}(1-p)^{\frac{n-k}{2}} - p^{\frac{n-k}{2}}(1-p)^{\frac{n+k}{2}+1}}{(2p-1)}.\end{aligned}$$

Thus, we obtain

$$\begin{aligned}&Q[\rho^{\otimes n}]\Big(\frac{n+k}{2},\frac{n-k}{2}\Big) = \dim \mathcal{U}_{\frac{n+k}{2},\frac{n-k}{2}}(S_n)\cdot \chi_{\frac{n+k}{2},\frac{n-k}{2}}(\rho) \\ =&\frac{k+1}{(2p-1)(n+1)}\Big(B\Big(n+1,p,\frac{n+k}{2}+1\Big) - B\Big(n+1,p,\frac{n-k}{2}+1\Big)\Big).\end{aligned}$$

Exercise 6.6
Assume that f satisfies the axiom of the distance as well as the additive condition $f(\rho^{\otimes n},\sigma^{\otimes n}) = nf(\rho,\sigma)$. Any state $\rho_{U,n}$ satisfies

$$nf(\rho,\sigma) = f(\rho^{\otimes n},\sigma^{\otimes n}) \le f(\rho^{\otimes n},\rho_{U,n}) + f(\sigma^{\otimes n},\rho_{U,n}).$$

Hence, at least one of $f(\rho^{\otimes n},\rho_{U,n})$ and $f(\sigma^{\otimes n},\rho_{U,n})$ increases linearly. So, no state is universal approximation with respect to f.

Exercise 6.7
Lemma 6.4 guarantees that the limit of the desired probability is $\frac{1}{\sqrt{2\pi}}\int_{\infty}^{\frac{R_2}{\sqrt{V(\boldsymbol{p})}}} e^{-\frac{x^2}{2}}dx$.

Exercise 6.8
Corollary 6.1 guarantees that the limit of the desired probability is $\frac{1}{\sqrt{2\pi}}\int_{\infty}^{\frac{R_2}{\sqrt{V(\boldsymbol{p})}}} e^{-\frac{x^2}{2}}dx$.

Exercise 6.9
The relation (6.59) shows that $\lim_{n\to\infty}\frac{1}{\sqrt{n}}(\log \operatorname{Tr} Q[H(\boldsymbol{p})+\frac{R_2}{\sqrt{n}},n] - H(\boldsymbol{p})) \le R_2$. Conversely, we have

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{\sqrt{n}}(\log \operatorname{Tr} Q[H(\boldsymbol{p})+\frac{R_2}{\sqrt{n}},n] - H(\boldsymbol{p})) \\ \ge &\lim_{n\to\infty}\frac{1}{\sqrt{n}}(\log \dim \mathcal{W}_{\boldsymbol{n}} - H(\boldsymbol{p})) \\ \ge &\lim_{n\to\infty}\frac{1}{\sqrt{n}}(\log \dim \mathcal{V}_{\boldsymbol{n}} - H(\boldsymbol{p})) = R_2,\end{aligned}$$

where $\boldsymbol{n}$ is chosen to satisfy $\log H(\frac{\boldsymbol{n}}{n}) = nH(\boldsymbol{p}) + \sqrt{n}R_2$. The final equation follows from (6.20) and (6.21). So, we obtain $\lim_{n\to\infty} \frac{1}{\sqrt{n}}(\log \mathrm{Tr}\, Q[H(\boldsymbol{p}) + \frac{R_2}{\sqrt{n}}, n] - H(\boldsymbol{p})) = R_2$.

Corollary 6.1 with respect to $\sqrt{n}(H(\frac{\boldsymbol{n}}{n}) - H(\boldsymbol{p}))$ guarantees that

$$\lim_{n\to\infty} \mathrm{Tr}\, \rho_{\mathrm{P}}^{\otimes n} Q[H(\boldsymbol{p}) + \frac{R_2}{\sqrt{n}}, n] = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\frac{R_2}{\sqrt{V(\rho_{\mathrm{P}})}}} e^{-\frac{x^2}{2}} dx. \tag{A.17}$$

Exercise 6.10
ϕ^2 is not injective because the map ϕ^2 maps $((1,1),(1,2),(2,1),(2,2)) \mapsto ((0,0),(0,0,0),(0,0,0),(0,0,0,0))$.

Exercise 6.11
When we replace $(0,1)$ by (1), ϕ_1 is modified as $(1,2) \mapsto ((0),(1))$. In this modification, ϕ_1^n is invertible. So, the original ϕ^n is also invertible. Hence, ϕ_1 is uniquely decodable.

Exercise 6.12
When ϕ_2 is defined as $(1,2) \mapsto ((0),(1,0))$, $\phi_1 \cdot \phi_2$ maps $((1,1),(1,2),(2,1),(2,2)) \mapsto ((0,0),(0,1,0),(0,1,0),(0,1,1,0))$. So, $\phi_1 \cdot \phi_2$ is not invertible.

Exercise 6.13
For any i_1 and $i_2 \neq j_2$, we have $\phi_1(i_1)\phi_2(i_2) \neq \phi_1(i_1)\phi_2(j_2)$. Also, for any $i_1 \neq j_1$ and i_2, we have $\phi_1(i_1)\phi_2(i_2) \neq \phi_1(j_1)\phi_2(i_2)$. Similarly, for any $i_1 \neq j_1$ and $i_2 \neq j_2$, we have $\phi_1(i_1)\phi_2(i_2) \neq \phi_1(j_1)\phi_2(j_2)$. So, $\phi_1 \cdot \phi_2$ is a prefix code.

Exercise 6.14
$(1,2,3,4) \mapsto ((1,0),(1,1,0),(1,1,1),(0))$.

Exercise 6.15
We list the sequences by grouping 6 sequences together as follows.

$$\begin{array}{ll}
((1,1),(1,2),(2,1),(2,2),(2,2)), & ((1,1),(1,2),(2,2),(2,1),(2,2)), \\
((1,1),(1,2),(2,2),(2,2),(2,1)), & ((1,2),(1,1),(2,1),(2,2),(2,2)), \\
((1,2),(1,1),(2,2),(2,1),(2,2)), & ((1,2),(1,1),(2,2),(2,2),(2,1)),
\end{array}$$

$$\begin{array}{ll}
((1,1),(2,1),(1,2),(2,2),(2,2)), & ((1,1),(2,2),(1,2),(2,1),(2,2)), \\
((1,1),(2,2),(1,2),(2,2),(2,1)), & ((1,2),(2,1),(1,1),(2,2),(2,2)), \\
((1,2),(2,2),(1,1),(2,1),(2,2)), & ((1,2),(2,2),(1,1),(2,2),(2,1)),
\end{array}$$

$$\begin{array}{ll}
((1,1),(2,1),(2,2),(1,2),(2,2)), & ((1,1),(2,2),(2,1),(1,2),(2,2)), \\
((1,1),(2,2),(2,2),(1,2),(2,1)), & ((1,2),(2,1),(2,2),(1,1),(2,2)), \\
((1,2),(2,2),(2,1),(1,1),(2,2)), & ((1,2),(2,2),(2,2),(1,1),(2,1)),
\end{array}$$

$$\begin{array}{ll}
((1,1),(2,1),(2,2),(2,2),(1,2)), & ((1,1),(2,2),(2,1),(2,2),(1,2)), \\
((1,1),(2,2),(2,2),(2,1),(1,2)), & ((1,2),(2,1),(2,2),(2,2),(1,1)),
\end{array}$$

$$((1,2),(2,2),(2,1),(2,2),(1,1)),\ ((1,2),(2,2),(2,2),(2,1),(1,1)),$$

$$\begin{gathered}
((2,1),(1,1),(1,2),(2,2),(2,2)),\ ((2,2),(1,1),(1,2),(2,1),(2,2)),\\
((2,2),(1,1),(1,2),(2,2),(2,1)),\ ((2,1),(1,2),(1,1),(2,2),(2,2)),\\
((2,2),(1,2),(1,1),(2,1),(2,2)),\ ((2,2),(1,2),(1,1),(2,2),(2,1)),
\end{gathered}$$

$$\begin{gathered}
((2,1),(1,1),(2,2),(1,2),(2,2)),\ ((2,2),(1,1),(2,1),(1,2),(2,2)),\\
((2,2),(1,1),(2,2),(1,2),(2,1)),\ ((2,1),(1,2),(2,2),(1,1),(2,2)),\\
((2,2),(1,2),(2,1),(1,1),(2,2)),\ ((2,2),(1,2),(2,2),(1,1),(2,1)),
\end{gathered}$$

$$\begin{gathered}
((2,1),(1,1),(2,2),(2,2),(1,2)),\ ((2,2),(1,1),(2,1),(2,2),(1,2)),\\
((2,2),(1,1),(2,2),(2,1),(1,2)),\ ((2,1),(1,2),(2,2),(2,2),(1,1)),\\
((2,2),(1,2),(2,1),(2,2),(1,1)),\ ((2,2),(1,2),(2,2),(2,1),(1,1)),
\end{gathered}$$

$$\begin{gathered}
((2,1),(2,2),(1,1),(1,2),(2,2)),\ ((2,2),(2,1),(1,1),(1,2),(2,2)),\\
((2,2),(2,2),(1,1),(1,2),(2,1)),\ ((2,1),(2,2),(1,2),(1,1),(2,2)),\\
((2,2),(2,1),(1,2),(1,1),(2,2)),\ ((2,2),(2,2),(1,2),(1,1),(2,1)),
\end{gathered}$$

$$\begin{gathered}
((2,1),(2,2),(1,1),(2,2),(1,2)),\ ((2,2),(2,1),(1,1),(2,2),(1,2)),\\
((2,2),(2,2),(1,1),(2,1),(1,2)),\ ((2,1),(2,2),(1,2),(2,2),(1,1)),\\
((2,2),(2,1),(1,2),(2,2),(1,1)),\ ((2,2),(2,2),(1,2),(2,1),(1,1)),
\end{gathered}$$

$$\begin{gathered}
((2,1),(2,2),(2,2),(1,1),(1,2)),\ ((2,2),(2,1),(2,2),(1,1),(1,2)),\\
((2,2),(2,2),(2,1),(1,1),(1,2)),\ ((2,1),(2,2),(2,2),(1,2),(1,1)),\\
((2,2),(2,1),(2,2),(1,2),(1,1)),\ ((2,2),(2,2),(2,1),(1,2),(1,1)).
\end{gathered}$$

References

1. E. Bagan, M. Baig, R. Munoz-Tapia, Quantum reverse-engineering and reference frame alignment without non-local correlations. Phys. Rev. A **70**, 030301 (2004)
2. E. Bagan, M.A. Ballester, R.D. Gill, A. Monras, R. Munoz-Tapia, Optimal full estimation of qubit mixed states. Phys. Rev. A **73**, 032301 (2006)
3. M. Bellare, S. Tessaro, A. Vardy, Semantic security for the wiretap channel, in *Proceedings of the 32nd Annual Cryptology Conference*, vol. 7417 (2012), pp. 294–311
4. C.H. Bennett, H.J. Bernstein, S. Popescu, B. Schumacher, Concentrating partial entanglement by local operations. Phys. Rev. A **53**, 2046 (1996)
5. C.H. Bennett, G. Brassard, Quantum cryptography: public key distribution and coin tossing, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* (Bangalore, India, 1984), pp. 175–179
6. C.H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, W.K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. Phys. Rev. Lett. **70**, 1895 (1993)
7. C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, W.K. Wootters, Mixed-state entanglement and quantum error correction. Phys. Rev. A **54**, 3824–3851 (1996)
8. C.H. Bennett, S.J. Wiesner, Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. Phys. Rev. Lett. **69**, 2881 (1992)
9. R. Bhatia, *Matrix Analysis* (Springer, New York, 1996)
10. I. Bjelakovic, J.-D. Deuschel, T. Kruger, R. Seiler, R. Siegmund-Schultze, A. Szkola, A quantum version of Sanov's theorem. Commun. Math. Phys. **260**(3), 659–671 (2005)
11. N.A. Bogomolov, Minimax measurements in a general statistical decision theory. Teor. Veroyatnost. i Primenen. **26**, 798–807 (1981) (English translation: Theory Probab. Appl. **26**, 4, 787–795 (1981))
12. D. Bouwmeester, A. Ekert, A. Zeilinger (eds.), *The Physics of Quantum Information: Quantum Cryptography* (Quantum Teleportation, Quantum Computation, Springer, 2000)
13. V. Bužek, R. Derka, S. Massar, Optimal quantum clocks. Phys. Rev. Lett. **82**, 2207 (1999)
14. A.R. Calderbank, E.M. Rains, P.W. Shor, N.J.A. Sloane, Quantum error correction and orthogonal geometry. Phys. Rev. Lett. **78**, 405–408 (1997)
15. A.R. Calderbank, E.M. Rains, P.W. Shor, N.J.A. Sloane, Quantum error correction via codes over GF(4). IEEE Trans. Inform. Theor. **44**, 1369–1387 (1998)
16. A.R. Calderbank, P.W. Shor, Good quantum error correcting codes exist. Phys. Rev. A **54**, 1098–1105 (1996)
17. J.L. Carter, M.N. Wegman, Universal classes of hash functions. J. Comput. Syst. Sci. **18**, 143–154 (1979)

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8

18. N.J. Cerf, A. Ipe, X. Rottenberg, Cloning of continuous quantum variables. Phys. Rev. Lett. **85**, 1754–1757 (2000)
19. M.-D. Choi, Completely positive linear maps on complex matrices. Linear Algebra Appl. 285–290 (1975)
20. G. Chiribella, G.M. D'Ariano, M.F. Sacchi, Optimal estimation of group transformations using entanglement. Phys. Rev. A **72**, 042338 (2005)
21. G. Chiribella, G.M. D'Ariano, P. Perinotti, M.F. Sacchi, Efficient use of quantum resources for the transmission of a reference frame. Phys. Rev. Lett. **93**, 180503 (2004)
22. M. Christandl, The Structure of Bipartite Quantum States Insights from Group Theory and Cryptography, PhD thesis, University of Cambridge, 2006
23. L. Collatz, U. Sinogowitz, Spektren endlicher Grafen. Abh. Math. Sem. Univ. Hamburg **21**, 63–77 (1957)
24. H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, 1946)
25. I. Csiszár, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Academic Press, 1981). (Cambridge University Press, 2nd edn., 2011)
26. M.J. Donald, M. Horodecki, Continuity of relative entropy of entanglement. Phys. Lett. A **264**, 257–260 (1999)
27. M. Donald, M. Horodecki, O. Rudolph, The uniqueness theorem for entanglement measures. J. Math. Phys. **43**, 4252–4272 (2002)
28. R.G. Gallager, *Information Theory and Reliable Communication* (Wiley, 1968)
29. N. Gisin, S. Massar, Optimal quantum cloning machines. Phys. Rev. Lett. **79**, 2153 (1997)
30. D. Gottesman, Class of quantum error-correcting codes saturating the quantum Hamming bound. Phys. Rev. A **54**, 1862–1868 (1996)
31. D. Gottesman, Theory of fault-tolerant quantum computation. Phys. Rev. A **57**, 127–37 (1998)
32. D. Gottesman, Fault-tolerant quantum computation with higher-dimensional systems. Chaos, Solitons Fractals **10**, 1749–1758 (1999)
33. M. Grassl, M. Rotteler, T. Beth, Efficient quantum circuits for non-qubit quantum error-correcting codes. Int. J. Found. Comput. Sci. (IJFCS) **14**(5), 757–775 (2003)
34. M. Hamada, Teleportation and entanglement distillation in the presence of correlation among bipartite mixed states. Phys. Rev. A **68**, 012301 (2003)
35. M. Hamada, Notes on the fidelity of symplectic quantum error-correcting codes. Int. J. Quantum Inf. **1**(4), 443–463 (2003)
36. M. Hamada, Reliability of Calderbank-Shor-Steane codes and security of quantum key distribution. J. Phys. A: Math. Gen. **37**(34), 8303–8328 (2004)
37. T.S. Han, K. Kobayashi, *Mathematics of Information and Coding* (American Mathematical Society, 2001)
38. A.W. Harrow, M.A. Nielsen, How robust is a quantum gate in the presence of noise? Phys. Rev. A **68**, 012308 (2003)
39. T. Hashimoto, A. Hayashi, M. Hayashi, M. Horibe, Unitary-process discrimination with error margin. Phys. Rev. A **81**, 062327 (2010)
40. M.B. Hastings, A counterexample to additivity of minimum output entropy. Nat. Phys. **5**, 255 (2009)
41. A. Hayashi, T. Hashimoto, M. Horibe, Extended quantum color coding. Phys. Rev. A **71**, 012326 (2005)
42. M. Hayashi, *Quantum Information: An Introduction* (Springer, 2006) (Originally published in Japanese in 2004)
43. M. Hayashi, S. Ishizaka, A. Kawachi, G. Kimura, T. Ogawa, *Introduction to Quantum Information Science*, Graduate Texts in Physics (Springer, 2015) (Originally published in Japanese in 2012)
44. M. Hayashi, *Group Representations for Quantum Theory*, (Springer, 2016) (Originally published in Japanese in 2014)
45. M. Hayashi, *Quantum Information: An Introduction* (Springer, 2006)
46. M. Hayashi, Asymptotic estimation theory for a finite dimensional pure state model. J. Phys. A: Math. Gen. **31**, 4633–4655 (1998). (It is also appeared as Chapter 23 of *Asymptotic Theory of Quantum Statistical Inference,* M. Hayashi eds.)

47. M. Hayashi, Optimal sequence of POVMs in the sense of Stein's lemma in quantum hypothesis. J. Phys. A: Math. Gen. **35**, 10759–10773 (2002)
48. M. Hayashi, Exponents of quantum fixed-length pure state source coding. Phys. Rev. A **66**, 032321 (2002)
49. M. Hayashi, K. Matsumoto, Asymptotic performance of optimal state estimation in quantum two level system. arXiv:quant-ph/0411073
50. M. Hayashi, K. Matsumoto, Quantum universal variable-length source coding. Phys. Rev. A **66**, 022311 (2002)
51. M. Hayashi, K. Matsumoto, Simple construction of quantum universal variable-length source coding. Quant. Inf. Comput. **2**, Special Issue, 519–529 (2002). arXiv:quant-ph/0209124
52. M. Hayashi, Parallel treatment of estimation of SU(2) and phase estimation. Phys. Lett. A **354**(3), 183–189 (2006)
53. M. Hayashi, Practical evaluation of security for quantum key distribution. Phys. Rev. A **74**, 022307 (2006)
54. M. Hayashi, Universal approximation of multi-copy states and universal quantum lossless data compression. Commun. Math. Phys. **293**(1), 171–183 (2010)
55. M. Hayashi, Universal coding for classical-quantum channel. Commun. Math. Phys. **289**(3), 1087–1098 (2009)
56. M. Hayashi, Upper bounds of eavesdropper's performances in finite-length code with the decoy method. Phys. Rev. A **76**, 012329 (2007)
57. M. Hayashi, General theory for decoy-state quantum key distribution with an arbitrary number of intensities. New J. Phys. **9**, 284 (2007)
58. M. Hayashi, D. Markham, M. Murao, M. Owari, S. Virmani, Bounds on multipartite entangled orthogonal state discrimination using local operations and classical communication. Phys. Rev. Lett. **96**, 040501 (2006)
59. M. Hayashi, D. Markham, M. Murao, M. Owari, S. Virmani, The geometric measure of entanglement for a symmetric pure state with positive amplitudes. J. Math. Phys. **50**, 122104 (2009)
60. M. Hayashi, Exponential decreasing rate of leaked information in universal random privacy amplification. IEEE Trans. Inf. Theor. **57**, 3989–4001 (2011)
61. M. Hayashi, T. Tsurumaru, Concise and tight security analysis of the Bennett-Brassard 1984 protocol with finite key lengths. New J. Phys. **14**, 093014 (2012)
62. M. Hayashi, Fourier Analytic Approach to Quantum Estimation of Group Action. Commun. Math. Phys. **347**(1), 3–82 (2009)
63. M. Hayashi, Universal channel coding for general output alphabet (2016). arXiv:1502.02218v2
64. P.M. Hayden, M. Horodecki, B.M. Terhal, The asymptotic entanglement cost of preparing a quantum state. J. Phys. A: Math. Gen. **34**, 6891–6898 (2001)
65. M. Hein, J. Eisert, H.J. Briegel, Multi-party entanglement in graph states. Phys. Rev. A **69**, 062311 (2004)
66. M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest, H.-J. Briegel, Entanglement in graph states and its applications, in *Proceedings of the International School of Physics "Enrico Fermi"*, vol. 162 ed. by G. Casati, D.L. Shepelyansky, P. Zoller, G. Benenti (IOP Press, 2006), pp. 115–218
67. F. Hiai, Matrix analysis: matrix monotone functions, matrix means, and majorization. Interdiscip. Inf. Sci. **16**, 139–248 (2010)
68. F. Hiai, D. Petz, The proper formula for relative entropy and its asymptotics in quantum probability. Commun. Math. Phys. **143**, 99–114 (1991)
69. T. Hiroshima, Majorization criterion for distillability of a bipartite quantum state. Phys. Rev. Lett. **91**, 057902 (2003)
70. A.S. Holevo, Covariant measurements and uncertainty relations. Rep. Math. Phys. **16**, 385–400 (1979)
71. A.S. Holevo, The capacity of the quantum channel with general signal states. IEEE Trans. Inf. Theor. **44**, 269 (1998)

72. A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982) Originally published in Russian (1980) (2nd Edition, Springer 2012)
73. P. Horodecki, Separability criterion and inseparable mixed states with positive partial transposition. Phys. Lett. A **232**, 333 (1997)
74. M. Horodecki, P. Horodecki, Reduction criterion of separability and limits for a class of distillation protocols. Phys. Rev. A **59**, 4206 (1999)
75. M. Horodecki, P. Horodecki, R. Horodecki, Separability of mixed states: necessary and sufficient conditions. Phys. Lett. A **223**, 1–8 (1996)
76. R. Howe, E.C. Tan, *Non-Abelian Harmonic Analysis: Applications of SL(2, R)* (Springer, 1992)
77. R. Hübener, M. Kleinmann, T.C. Wei, C. González-Guillén, O. Gühne, The geometric measure of entanglement for symmetric states. Phys. Rev. A **80**, 032324 (2009)
78. A. Jamiolkowski, Linear transformations which preserve trace and positive semidefiniteness of operators. Rep. Math. Phys. **3**, 275 (1972)
79. G.A. Jones, J.M. Jones, *Information and Coding Theory* (Springer, 2000)
80. R. Jozsa, M. Horodecki, P. Horodecki, R. Horodecki, Phys. Rev. Lett. **81**, 1714 (1998)
81. A. Klappenecker, M. Rötteler, Beyond stabilizer codes II: Clifford codes. IEEE Trans. Inform. Theor. **48**(8), 2396–2399 (2002)
82. A. Klappenecker, P.K. Sarvepalli, Clifford code constructions of operator quantum error-correcting codes. IEEE Trans. Inform. Theor. **54**(12), 5760–5765 (2008)
83. A. Klappenecker, P.K. Sarvepalli, Encoding subsystem codes. Int. J. Adv. Secur. **2**(2 & 3), 142–155 (2009)
84. E. Knill, Group representations, error bases and quantum codes (1996). arXiv:quant-ph/9608049
85. M. Koashi, N. Imoto, Compressibility of mixed-state signals. Phys. Rev. Lett. **87**, 017902 (2001)
86. J. Von Korff, J. Kempe, Quantum advantage in transmitting a permutation. Phys. Rev. Lett. **93**(26), 260502 (2004)
87. M. Keyl, R.F. Werner, Estimating the spectrum of a density operator. Phys. Rev. A **64**, 052311 (2001)
88. K. Kraus, *States, Effects and Operations* (Springer, 1983)
89. A. Luis, J. Perina, Optimum phase-shift estimation and the quantum description of the phase difference. Phys. Rev. A **54**, 4564 (1996)
90. D. Markham, A. Miyake, S. Virmani, Entanglement and local information access for graph states. New J. Phys. **9**, 194 (2007)
91. A.W. Marshall, I. Olkin, *Inequalities: Theory of Majorization and Its Applications* (Academic Press, 1979)
92. R. Matsumoto, Conversion of a general quantum stabilizer code to an entanglement distillation protocol. J. Phys. Math. Gen. **36**(29), 8113–8127 (2003)
93. K. Matsumoto, M. Hayashi, Universal distortion-free entanglement concentration. Phys. Rev. A **75**, 062338 (2007)
94. K. Matsumoto, T. Shimono, A. Winter, Remarks on additivity of the Holevo channel capacity and of the entanglement of formation. Comm. Math. Phys. **246**(3), 427–442 (2004)
95. D. Mayers, Quantum key distribution and string oblivious transfer in noisy channels, in *Advances in Cryptography Proceedings of Crypto'96* (1996), pp. 343–357
96. D. Mayers, Unconditional security in quantum cryptography. J. Assoc. Comp. Mach. **48**, 351–406 (2001)
97. N.D. Mermin, *Quantum Computer Science: An Introduction* (Cambridge University Press, 2007)
98. T. Miyadera, Information-disturbance theorem for mutually unbiased observables. Phys. Rev. A **73**, 042317 (2006)
99. M.A. Naĭmark, Comptes rendus (Doklady) de l'Acadenie des science de l'URSS, 41. **9**, 359 (1943)

100. M.A. Nielsen, Conditions for a class of entanglement transformations. Phys. Rev. Lett. **83**, 436 (1999)
101. M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000)
102. M. Ohya, D. Petz, *Quantum Entropy and Its Use* (Springer, New York, 1993)
103. M. Owari, M.B. Plenio, E.S. Polzik, A. Serafini, M.M. Wolf, Squeezing the limit: quantum benchmarks for the teleportation and storage of squeezed states. New J. Phys. **10**, 113014 (2008)
104. M. Ozawa, On the noncommutative theory of statistical decisions. Res. Rep. Inf. Sci. **A-74** (1980)
105. M. Ozawa, Quantum state reduction and the quantum Bayes principle, in *Quantum Communication, Computing, and Measurement*, ed. by O. Hirota, A.S. Holevo, C.M. Caves (Plenum, New York, 1997), pp. 233–241
106. M. Ozawa, An Operational approach to quantum state reduction. Ann. Phys. **259**, 121–137 (1997)
107. M. Ozawa, Quantum state reduction: an operational approach. Fortschr. Phys. **46**, 615–625 (1998)
108. S. Popescu, Bell's inequalities versus teleportation: what is nonlocality? Phys. Rev. Lett. **72**, 797–799 (1994)
109. M. Reed, B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic Press, 1980)
110. R. Renner, Security of quantum key distribution, PhD thesis, ETH Zurich, 2005. arXiv:quant-ph/0512258
111. B. Schumacher, Sending entanglement through noisy quantum channels. Phys. Rev. A **54**, 2614–2628 (1996)
112. B. Schumacher, M.A. Nielsen, Quantum data processing and error correction. Phys. Rev. A **54**, 2629 (1996)
113. B. Schumacher, M.D. Westmoreland, Sending classical information via noisy quantum channels. Phys. Rev. A **56**, 131 (1997)
114. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379– 423 and 623–656 (1948)
115. N. Sharma, N.A. Warsi, Fundamental bound on the reliability of quantum information transmission. Phys. Rev. Lett. **110**(8), 080501 (2013)
116. A. Shimony, Ann. N.Y. Acad. Sci. **755**, 675 (1995)
117. P. Shor, J. Preskill, Simple proof of security of the BB84 quantumkey distribution protocol. Phys. Rev. Lett. **85**, 441–444 (2000)
118. A.M. Steane, Multiple particle interference and quantum error correction. Proc. Roy. Soc. Lond. A **452**, 2551–2577 (1996)
119. W.F. Stinespring, Positive functions on C^* algebras. Proc. Am. Math. Soc. **6**, 211 (1955)
120. D.R. Stinson, Universal hash families and the leftover hash lemma, and applications to cryptography and computing. J. Combin. Math. Combin. Comput. **42**, 3–31 (2002)
121. A. Streltsov, H. Kampermann, D. Bruß, Linking a distance measure of entanglement to its convex roof. New J. Phys. **12**, 123004 (2010)
122. T. Tsurumaru, M. Hayashi, Dual universality of hash functions and its applications to classical and quantum cryptography. IEEE Trans. Inf. Theor. **59**(7), 4700–4717 (2013)
123. V. Vedral, M.B. Plenio, Phys. Rev. A **57**, 1619 (1998)
124. G. Vidal, R. Tarrach, Robustness of entanglement. Phys. Rev. A **59**, 141 (1999)
125. X.B. Wang, T. Hiroshima, A. Tomita, M. Hayashi, Quantum Information with Gaussian States. Phys. Rep.: Rev. Sect. Phys. Lett. **448**, 1–111 (2007)
126. S. Watanabe, T. Matsumoto, T. Uyematsu, Noise tolerance of the BB84 protocol with random privacy amplification. Int. J. Quantum Inf. **4**(6), 935–946 (2006)
127. T.-C. Wei, P.M. Goldbart, Phys. Rev. A **68**, 042307 (2003)
128. R.F. Werner, Optimal cloning of pure states. Phys. Rev. A **58**, 1827 (1998)
129. M.M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2013)

130. H.M. Wiseman, R.B. Killip, Adaptive single-shot phase measurements: a semiclassical approach. Phys. Rev. A **56**, 944–957 (1997)
131. A.D. Wyner, The wire-tap channel. Bell Syst. Tech. J. **54**, 1355–1387 (1975)
132. H. Zhu, L. Chen, M. Hayashi, Additivity and non-additivity of multipartite entanglement measures. New J. Phys. **12**, 083002 (2010)

Index

M. Hayashi, *A Group Theoretic Approach to Quantum Information*,
DOI 10.1007/978-3-319-45241-8