Takahiro Sagawa

Entropy, Divergence, and Majorization in Classical and Quantum Thermodynamics

SpringerBriefs in Mathematical Physics

Volume 16

SpringerBriefs are characterized in general by their size (50–125 pages) and fast production time (2–3 months compared to 6 months for a monograph).

Briefs are available in print but are intended as a primarily electronic publication to be included in Springer's e-book package.

Typical works might include:

- An extended survey of a field
- A link between new research papers published in journal articles
- A presentation of core concepts that doctoral students must understand in order to make independent contributions
- Lecture notes making a specialist topic accessible for non-specialist readers.

SpringerBriefs in Mathematical Physics showcase, in a compact format, topics of current relevance in the field of mathematical physics. Published titles will encompass all areas of theoretical and mathematical physics. This series is intended for mathematicians, physicists, and other scientists, as well as doctoral students in related areas.

Editorial Board

- Nathanaël Berestycki (University of Cambridge, UK)
- Mihalis Dafermos (University of Cambridge, UK / Princeton University, US)
- Atsuo Kuniba (University of Tokyo, Japan)
- Matilde Marcolli (CALTECH, US)
- Bruno Nachtergaele (UC Davis, US)
- Hal Tasaki (Gakushuin University, Japan)
- 50 – 125 published pages, including all tables, figures, and references
- Softcover binding
- Copyright to remain in author's name
- Versions in print, eBook, and MyCopy

More information about this series at https://link.springer.com/bookseries/11953

Takahiro Sagawa

Entropy, Divergence, and Majorization in Classical and Quantum Thermodynamics

Takahiro Sagawa
Department of Applied Physics
The University of Tokyo
Tokyo, Japan

Preface

In these decades, it has been revealed that there is rich information-theoretic structure in thermodynamics of out-of-equilibrium systems in both the classical and quantum regimes. This has led to the fruitful interplay among statistical physics, quantum information theory, and mathematical theories including matrix analysis and asymptotic probability theory.

The main purpose of this book is to clarify how information theory works behind thermodynamics and to shed modern light on it. I will focus on both of purely information-theoretic concepts and their physical implications: I present self-contained and rigorous proofs of several fundamental properties of entropies, divergences, and majorization. I also discuss the modern formulations of thermodynamics, especially from the perspectives of stochastic thermodynamics and resource theory of thermodynamics. Resource theory is a recently-developed field as a branch of quantum information theory in order to quantify (energetically or information-theoretically) "useful resources." I show that resource theory has an intrinsic connection to various fundamental ideas of mathematics and information theory.

This book is not intended to be a comprehensive review of the field but would serve as a concise introduction to several important ingredients of the information-theoretic formulation of thermodynamics. I hope that the readers would grasp a compact overview on physics and mathematics of entropy-like quantities from the modern point of view.

Tokyo, Japan   Takahiro Sagawa
May 2021

Acknowledgements

Contents

Chapter 1
Introduction

Thermodynamics was originally established as a phenomenological theory of thermal equilibrium [1]. Entropy is at the core of thermodynamics, as it provides a complete characterization of state convertibility through the second law of thermodynamics; In the absence of a heat bath, a state transformation between equilibrium states is possible *if and only if* the entropy does not decrease. This remarkable feature of entropy has been formalized by a rigorous axiomatic theory at the purely phenomenological level by Lieb and Yngvason [2].

Once one goes beyond conventional equilibrium situations, however, the notion of entropy becomes more subtle. In fact, a proper definition of entropy of nonequilibrium states has been a long-standing problem in statistical physics. In light of progress in theories of modern thermodynamics, it has been revealed that information-theoretic entropies play significant roles in thermodynamics of out-of-equilibrium and even quantum systems.

There is a long history of researches of such entropy-like quantities in classical and quantum information theories [3–6] as well as matrix analysis [7, 8], where it has been shown that several information-theoretic quantities exhibit universal features that resemble the second law of thermodynamics. A particularly important concept is *divergence* as well as information-theoretic *entropy*. Divergence is also referred to as relative entropy and corresponds to free energy of thermodynamics.

In this book, we will discuss the fundamental properties of entropies and divergences both in the classical and quantum regimes, including the Shannon entropy [9], the von Neumann entropy [10], the Kullback-Leibler (KL) divergence [11] and its quantum generalization [12], the Rényi entropy and divergence [13], the f-divergence [14, 15], and general quantum divergence-like quantities called the Petz's quasi-entropies [16, 17]. In addition, we will briefly discuss the classical and quantum Fisher information [18–20], which has a close connection to divergences.

We will see that the monotonicity properties of these divergences under stochastic or quantum dynamics bring us to the information-theoretic foundation of the second law of thermodynamics. An advantage of such an approach to the second law lies in the fact that information-theoretic entropies and divergences can be defined for arbitrary probability distributions and quantum states including out-of-equilibrium ones, and therefore informational quantities can take the place of the Boltzmann entropy defined only for the equilibrium ensembles.

From the physics side, recent progress of experimental technologies of manipulating small-scale systems has led us to a fruitful playground of modern thermodynamics. Small-scale heat engines have been experimentally realized with various systems both in the classical and quantum regimes, such as colloidal particles [21, 22], biomolecules [23], single electrons [24], superconducting qubits [25], and NMR [26], where the connection between the second law and information theory has been investigated in real laboratories.

In contrast to the case of conventional macroscopic systems, thermodynamic quantities of small-scale systems become random variables, because dynamics of small systems exhibit intrinsic stochasticity induced by thermal fluctuations of heat baths. This is a fundamental reason why modern thermodynamics is relevant to information theory based on probability theory.

There are two complementary research streams that deal with theoretical formulations of small-scale thermodynamics in the above-mentioned spirit. One is *stochastic thermodynamics* and the other is *resource theory of thermodynamics*. Stochastic thermodynamics has been developed as a branch of nonequilibrium statistical mechanics [27–30] (see also review articles [31–35]) and has led to modern understanding of thermodynamics of information [36–38], shedding new light on "Maxwell's demon" [39]. On the other hand, resource theory of thermodynamics has been developed more recently as a branch of quantum information theory [40–45] (see also review articles [46–48]). There are several fundamental differences between these two approaches both physically and mathematically. In particular, stochastic thermodynamics allows fluctuations of work, while resource theory of thermodynamics does not, as mentioned below. They do not only correspond to different experimental setups but also are described by different mathematical structures, and as a consequence, lead to different forms of the second law described by different types of entropies. We will see, however, that they lead to the same form of the second law in the macroscopic (asymptotic) limit.

From the perspective of resource theory of thermodynamics, the second law quantifies how much "resource" such as work is required for a desired thermodynamic task. One of the key ideas is that work is supposed to be a deterministic, not random quantity, in order to implement the idea that work is a purely "mechanical" quantity without any entropic contribution. This is referred to as *single-shot* (or one-shot) thermodynamics [41, 42] and is contrastive to the setup of stochastic thermodynamics that allows work fluctuations. We note that there are attempts to connect resource-theoretic results with experimentally relevant setups (e.g., Refs. [49, 50]).

In general, resource theory is an information-theoretic framework to quantify "useful resources" (see also a review article [51]). For example, resource theory of

entanglement is an earliest resource theory in quantum information theory [4, 52], and has a similar mathematical structure to resource theory of thermodynamics at infinite temperature. To formulate a resource theory, we need to identify *free states* and *free operations*, which can be prepared and performed without any cost. It is also important to consider a *monotone*, which is a quantity that monotonically changes (does not increase or does not decrease) under free operations [53]. In particular, there is a concept called a *complete monotone*, which provides a sufficient, not only necessary, condition that an operation is possible. We remark that the concept of single-shot is significant in various resource theories [54].

In the case of thermodynamics, work and nonequilibrium states are regarded as resources, because work can be extracted only from nonequilibrium states and nonequilibrium states can be created only by investing work. On the other hand, Gibbs states are free states and relaxation processes (called Gibbs-preserving maps or thermal operations) are free operations.

Conventional thermodynamics can be regarded as a prototypical resource theory, where entropy is a complete monotone that provides a necessary and sufficient condition of transitions between equilibrium states [2]. In modern nonequilibrium thermodynamics, the divergences (including the KL divergence and the Rényi divergence) are monotones, implying that these divergences serve as thermodynamic potentials (or free energies). However, except for equilibrium transitions, such divergences are not complete monotones in the single-shot scenario. This brings us to a mathematical concept called *majorization* [7, 55], which enables a complete characterization of thermodynamic state transformations. Majorization plays a central role in resource theory of entanglement, while a generalized concept called *thermo-majorization* [41] (and *d-majorization* [56–58]) is crucial for resource theory of thermodynamics at finite temperature.

It is also interesting to take the asymptotic limit in resource theories, where many copies of the system are available. In the asymptotic limit, structure of state convertibility often becomes simpler than standard majorization and can be characterized by a single complete monotone. This is a main focus of the theory of *information spectrum* [59–65], where a quantum version of *asymptotic equipartition properties (AEP)* [3] plays a significant role. To take the asymptotic limit, it is useful to utilize quantities called *smooth* entropy and divergence [66, 67]. From the physics point of view, the asymptotic limit represents the thermodynamic (macroscopic) limit of many-body systems, and a complete monotone is regarded as a complete macroscopic thermodynamic potential like the one discussed by Lieb and Yngvason [2]. We will emphasize that the asymptotic theory of information spectrum can be applied to interacting many-body systems beyond the independent and identically-distributed (i.i.d.) situations.

The organization of this book is as follows.

In Chap. 2, we discuss the properties of classical information-theoretic entropies and divergences: especially, the Shannon entropy, the KL divergence, and the Rényi α-entropies and divergences. We also briefly mention the f-divergence and the Fisher information.

In Chap. 3, we review majorization and thermo-majorization (and d-majorization) for classical stochastic systems. We introduce the Lorenz curve to visualize majorization and show that the Rényi 0- and ∞-divergences provide a useful characterization of state convertibility.

In Chap. 4, we apply the foregoing information-theoretic argument (Chaps. 2 and 3) to classical thermodynamics and show that the second law immediately follows from general information-theoretic inequalities. Specifically, the KL divergence is relevant to the setup of stochastic thermodynamics with fluctuating work, while the Rényi 0- and ∞-divergences are relevant to resource theory of thermodynamics in the single-shot scenario.

In Chap. 5, we move to the quantum case. After a brief overview of quantum states and quantum dynamics, we discuss quantum information-theoretic entropies and divergences. In particular, we will focus on the von Neumann entropy, the quantum KL divergence, and the quantum Rényi 0- and ∞-divergences.

In Chap. 6, we consider the quantum counterpart of majorization. The mathematical structure of ordinary majorization of the quantum case is similar to that of the classical case. The quantum version of thermo-majorization (and d-majorization) is a more subtle concept, while the quantum Rényi 0- and ∞-divergences can still characterize state convertibility.

In Chap. 7, we consider approximate and asymptotic state conversion. We introduce the smooth Rényi 0- and ∞-divergences, and as their asymptotic limit, introduce the spectral divergence rates that play a key role in the theory of information spectrum. We then discuss a quantum version of the AEP under certain assumptions including ergodicity, and show that the AEP implies the existence of a single complete monotone, which is nothing but the KL divergence rate.

In Chap. 8, we apply the foregoing argument (Chaps. 5, 6, and 7) to quantum thermodynamics. We generally formulate Gibbs-preserving maps and thermal operations and derive the second law of quantum thermodynamics. We discuss both of the fluctuating-work formulation and the single-shot scenario. We also consider the asymptotic limit, where a complete macroscopic thermodynamic potential emerges.

In Appendix A, we provide a proof of the monotonicity properties of general quantum divergences, by invoking mathematical techniques of matrix analysis. In particular, we show that operator monotone and operator convex play crucial roles. We also discuss the quantum Fisher information and prove its monotonicity property.

In Appendix B, we briefly overview quantum hypothesis testing, which has a fundamental connection to the asymptotic theory discussed in Chap. 7. We discuss that quantum hypothesis testing provides essentially the same information as the smooth Rényi 0- and ∞-divergences. We also discuss semidefinite programming as a useful tool and focus on the quantum Stein's lemma as another representation of the quantum AEP.

In Appendix C, we discuss the classical AEP by considering classical ergodic processes, and provide a proof of the classical Stein's lemma. This is regarded as the classical counterpart of Chap. 7 and Appendix B.

Finally, we remark on a notation that is used throughout this book. $\beta \geq 0$ represents the inverse temperature of the environment (i.e., the heat bath), that is, $\beta := (k_{\mathrm{B}} T)^{-1}$ with $0 < T \leq \infty$ being the corresponding temperature and k_{B} being the Boltzmann constant. Also, we set the Planck constant to unity.

We will basically restrict ourselves to finite-dimensional systems for both of the classical and quantum cases, unless stated otherwise. Only for a few topics such as continuous-variable majorization and the AEP, we will adopt infinite-dimensional setups.

Chapter 2
Classical Entropy and Divergence

In this chapter, we consider the basic concepts of classical information theory. In Sect. 2.1, we formulate classical states and dynamics as probability distributions and stochastic matrices. We then introduce the Shannon entropy and the Kullback-Leibler (KL) divergence (relative entropy) in Sect. 2.2 and the Rényi α-entropy and α-divergence in Sect. 2.3. In Sect. 2.4, we consider more general divergence-like quantities including the f-divergence, and as special cases, provide proofs of the properties of the KL and the Rényi divergences. In Sect. 2.5, we briefly discuss the classical Fisher information, which is a lower-order expansion of the KL divergence.

2.1 Classical State and Dynamics

As a preliminary, we fix our terminologies and notations for classical probability theory. The concept of *state* of a classical system is represented by a probability distribution $p := (p_1, p_2, \ldots, p_d)^{\mathrm{T}}$ with $\sum_{i=1}^{d} p_i = 1$ and $p_i \geq 0$. Here, p is regarded as a d-dimensional column vector and T represents the transpose. In the following, we do not distinguish *state* and *distribution*. On the other hand, the classical phase-space point corresponds to index i, which is also referred to as a "state" in the physics convention, while we do not use this terminology in this book.

Let $\mathscr{P}_d$ be the set of d-dimensional probability distributions. We denote the uniform distribution as $u := (1/d, 1/d, \ldots, 1/d)^{\mathrm{T}}$. With two independent distributions $p \in \mathscr{P}_d$ and $q \in \mathscr{P}_{d'}$, we write their joint distribution as $p \otimes q \in \mathscr{P}_{dd'}$ whose components are given in the form $p_i q_j$. This "tensor product" notation is consistent with the quantum case, where a classical probability distribution is regarded as the diagonal elements of a density matrix. If there are n samples that are independent and identically distributed (i.i.d.), we write the joint distribution as $p^{\otimes n} \in \mathscr{P}_{d^n}$.

We note that the *support* of p, denoted as supp[p], is defined as the set of indexes $\{i : p_i > 0\}$. The *rank* of p is the number of elements of supp[p], i.e., rank[p] := |supp[p]|. If $p \in \mathscr{P}_d$ does not have a zero component, then p is called to have full rank.

A time evolution of classical probability distributions is represented by a *stochastic matrix*, which maps the input state at the initial time into the output state at the final time. For the sake of simplicity, we assume that the input and the output spaces have the same dimensions in Chaps. 2, 3, and 4 (i.e., for the classical case), which means that any stochastic matrix is a square matrix. The components of a stochastic matrix T must satisfy $\sum_{i=1}^{d} T_{ij} = 1$ and $T_{ij} \geq 0$. The time evolution of a probability vector is then given by $p' = Tp$, or equivalently, $p_i' = \sum_{j=1}^{d} T_{ij} p_j$. Such a map on probability distributions is called a stochastic map or a Markov map, for which we use the same notation T.

If a stochastic matrix T further satisfies $\sum_{j=1}^{d} T_{ij} = 1$, it is called a *doubly stochastic matrix*. By definition, a stochastic matrix T is doubly stochastic if and only if the uniform distribution is its fixed point: $u = Tu$.

Meanwhile, we note that for distributions p, q, the *trace distance* is defined as

$$D(p, q) := \frac{1}{2} \| p - q \|_1 := \frac{1}{2} \sum_{i=1}^{d} |p_i - q_i|, \tag{2.1}$$

where $\| \cdot \|_1$ is called the trace norm. The trace distance does not increase under any stochastic map T, which is called the monotonicity (or the data processing inequality):

$$D(p, q) \geq D(Tp, Tq). \tag{2.2}$$

The proof of this is easy, but we will postpone it to Sect. 2.4 where we give a more general perspective on the monotonicity.

2.2 Shannon Entropy and the KL Divergence

A most basic concept in information theory is the *Shannon entropy*, which is defined for a classical probability distribution $p \in \mathscr{P}_d$ as

$$S_1(p) := - \sum_{i=1}^{d} p_i \ln p_i. \tag{2.3}$$

Here, we added the subscript "1" because the Shannon entropy is the Rényi 1-entropy as discussed later. Obviously, $S_1(p) \geq 0$.

We next consider two distributions $p, q \in \mathscr{P}_d$. The *Kullback-Leibler (KL) divergence* (or the *relative entropy*) is defined as

$$S_1(p\|q) := \sum_{i=1}^{d} p_i \ln \frac{p_i}{q_i}. \tag{2.4}$$

If the support of p is not included in that of q, we define $S_1(p\|q) := +\infty$. In order to avoid too much complexity, however, we always assume that the support of p is included in that of q throughout this book, whenever we consider divergence-like quantities.

The KL divergence is regarded as an asymmetric "distance" between two distributions. It is non-negative:

$$S_1(p\|q) \geq 0, \tag{2.5}$$

where the equality $S_1(p\|q) = 0$ holds if and only if $p = q$. A simplest way to see this is as follows: Since $\ln(x^{-1}) \geq 1 - x$ holds for $x > 0$, where the equality holds if and only if $x = 1$, we have $\sum_i p_i \ln(p_i/q_i) \geq \sum_i p_i(1 - q_i/p_i) = 0$ and the equality condition.

The entropy and the divergence are related as

$$S_1(p) = \ln d - S_1(p\|u), \tag{2.6}$$

where u is the uniform distribution. From this and the non-negativity of the divergence, we obtain $S_1(p) \leq \ln d$.

We note that if $\Delta p := p - q$ is small, i.e., $\varepsilon := \|\Delta p\|_1 \ll 1$, we can expand the KL divergence up to the second order of ε as

$$S_1(p\|p - \Delta p) = \frac{1}{2} \sum_i \frac{(\Delta p_i)^2}{p_i} + O(\varepsilon^3), \tag{2.7}$$

where we used $\sum_i \Delta p_i = 0$ to drop the term of $O(\varepsilon)$. The right-hand side of (2.7) is related to the Fisher information [18, 20].

The KL divergence also satisfies the *monotonicity* (or the *data processing inequality*). Suppose that $p, q, p', q' \in \mathscr{P}_d$ satisfy $p' = Tp$ and $q' = Tq$ for a stochastic matrix T. Then,

$$S_1(p\|q) \geq S_1(p'\|q'), \tag{2.8}$$

which represents that two distributions become harder to be distinguished, if they are "coarse-grained" by a stochastic map. In terms of resource theory, this implies that the KL divergence is a *monotone*, where in general a monotone is a quantity that does not increase (or decrease) under free operations. Here, all stochastic maps are (rather formally) regarded to be free, while in the thermodynamic setup, only Gibbs-preserving maps are supposed to be free (see also Sect. 3.2 and Chap. 4).

A straightforward way to show inequality (2.8) is that

$$S_1(p\|q) - S_1(p'\|q') = \sum_i p_i \ln \frac{p_i}{q_i} - \sum_j p'_j \ln \frac{p'_j}{q'_j} \tag{2.9}$$

$$= \sum_{ij} T_{ji} p_i \ln \frac{T_{ji} p_i}{T_{ji} q_i} - \sum_j p'_j \ln \frac{p'_j}{q'_j} = \sum_j p'_j S_1(\tilde{p}^{(j)}\|\tilde{q}^{(j)}) \geq 0, \tag{2.10}$$

where $S_1(\tilde{p}^{(j)}\|\tilde{q}^{(j)}) := \sum_i \frac{T_{ji} p_i}{p'_j} \ln \frac{T_{ji} p_i / p'_j}{T_{ji} q_i / q'_j}$ is the KL divergence between the conditional distributions $\tilde{p}_i^{(j)} := T_{ji} p_i / p'_j$ and $\tilde{q}_i^{(j)} := T_{ji} q_i / q'_j$. Another proof of (2.8) will be shown in Sect. 2.4 as a special case of the monotonicity of more general divergence-like quantities. We will discuss the significance of the monotonicity in thermodynamics of classical systems in Sect. 4.1.

We note that the converse of the monotonicity (2.8) is not true in general: There are pairs of states (p, q) and (p', q') satisfying inequality (2.8) such that any stochastic map T satisfying $p' = Tp$ and $q' = Tq$ does not exist. In other words, the monotonicity of the KL divergence is only a necessary condition, but not a sufficient condition for the convertibility of pairs of states. In the resource theory terminology, this implies that the KL divergence is a monotone but not a *complete* monotone under stochastic maps, because a complete monotone must provide a sufficient condition for state convertibility [53]. This observation brings us to the concept of majorization (more generally, d-majorization), which will be discussed in the next chapter. In particular, we will show that there exists a *complete set* of monotones, which provides a sufficient condition with a collection of infinitely many divergence-like quantities.

We next consider the case of doubly stochastic matrices. By noting Eq. (2.6) and the fact that a doubly stochastic matrix T does not change u, the monotonicity (2.8) of the KL divergence implies that

$$S_1(p) \leq S_1(Tp), \tag{2.11}$$

which represents that a doubly stochastic matrix makes distributions more "random." In terms of resource theory, this implies that the Shannon entropy is a monotone under doubly stochastic maps (but again, is not a complete monotone).

The Shannon entropy and the KL divergence also have other fundamental properties such as the subadditivity, the strong subadditivity, the concavity of the Shannon entropy, and the joint concavity of the KL divergence. However, we will postpone to discuss them to Chap. 5 for the quantum setup, from which the classical counterpart immediately follows.

We here briefly remark on a quantity called mutual information. We consider two systems A and B. Let p_{AB} be a distribution of AB, and p_A and p_B be its marginal distributions of A and B, respectively. Then, the mutual information between A and B in distribution p_{AB} is defined as

$$I_1(p_{AB})_{A:B} := S_1(p_A) + S_1(p_B) - S_1(p_{AB}) = S_1(p_{AB}\|p_A \otimes p_B) \geq 0. \tag{2.12}$$

This quantifies the correlation between A and B, and $I_1(p_{AB})_{A:B} = 0$ holds if and only if A and B are statistically independent, i.e., $p_{AB} = p_A \otimes p_B$. The monotonicity (2.8) of the KL divergence implies the data processing inequality of the mutual information: $I_1(p_{AB})_{A:B} \geq I_1(T_A \otimes T_B p_{AB})_{A:B}$, where $T_A \otimes T_B$ represents a stochastic map independently acting on A and B.

2.3 Rényi Entropy and Divergence

We next discuss generalized entropies: the *Rényi α-entropies* for $0 \leq \alpha \leq \infty$. For a distribution $p \in \mathscr{P}_d$, the Rényi α-entropy is defined as

$$S_\alpha(p) := \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^{d} p_i^\alpha \right). \tag{2.13}$$

Here, $S_\alpha(p)$ for $\alpha = 0, 1, \infty$ is defined by taking the limit: $S_1(p)$ is indeed the Shannon entropy, and

$$S_0(p) := \ln (\text{rank}[p]), \tag{2.14}$$

$$S_\infty(p) := -\ln \left(\max_i \{p_i\} \right). \tag{2.15}$$

We also denote these quantities by $S_{\min}(p) := S_\infty(p)$ and $S_{\max}(p) := S_0(p)$, which are referred to as the min and the max entropies, respectively. It is obvious that

$$S_\alpha(p) \geq 0. \tag{2.16}$$

We next consider the *Rényi α-divergence* for distributions $p, q \in \mathscr{P}_d$ with $0 \leq \alpha \leq \infty$ [68, 69]. It is defined as

$$S_\alpha(p\|q) := \frac{1}{\alpha-1} \ln \left(\sum_{i=1}^{d} \frac{p_i^\alpha}{q_i^{\alpha-1}} \right). \tag{2.17}$$

For $\alpha = 0, 1, \infty$, we again take the limit: $S_1(p\|q)$ is the KL divergence, and

$$S_0(p\|q) := -\ln \left(\sum_{i:p_i>0} q_i \right), \tag{2.18}$$

$$S_\infty(p\|q) := \ln \left(\max_i \left\{ \frac{p_i}{q_i} \right\} \right). \tag{2.19}$$

We also denote these quantities by $S_{\min}(p\|q) := S_0(p\|q)$ and $S_{\max}(p\|q) := S_\infty(p\|q)$, where on the contrary to the entropy case, the min and the max divergences correspond to $\alpha = 0$ and ∞, respectively. We again note that, in the definition of these divergence-like quantities, we always assume that the support of p is included in that of q.

An important property of the Rényi divergence is non-negativity, while we postpone the proof to Sect. 2.4.

Proposition 2.1 (Non-negativity of the Rényi divergence)

$$S_\alpha(p\|q) \geq 0. \tag{2.20}$$

For $0 < \alpha \leq \infty$, the equality $S_\alpha(p\|q) = 0$ holds if and only if $p = q$. For $\alpha = 0$, the equality holds if and only if the supports of p and q are the same.

The Rényi α-divergence satisfies the monotonicity: it does not increase (and thus is a monotone) under stochastic maps. (We again postpone the proof to Sect. 2.4.)

Proposition 2.2 (Monotonicity of Rényi divergence, Theorem 1 of [69]) *For any stochastic matrix T, the Rényi α-divergence with $0 \leq \alpha \leq \infty$ satisfies*

$$S_\alpha(p\|q) \geq S_\alpha(Tp\|Tq). \tag{2.21}$$

With the uniform distribution u, the α-divergence and the α-entropy are related as

$$S_\alpha(p) = \ln d - S_\alpha(p\|u), \tag{2.22}$$

from which we have

$$S_\alpha(p) \leq \ln d. \tag{2.23}$$

From the monotonicity of the α-divergence, we have, for any doubly stochastic matrix T,

$$S_\alpha(p) \leq S_\alpha(Tp). \tag{2.24}$$

We also note the following property. (Again, we will prove it in Sect. 2.4.)

Proposition 2.3 (Theorem 3 of [68])

$$S_\alpha(p\|q) \leq S_{\alpha'}(p\|q) \ \text{ for } \ \alpha \leq \alpha', \tag{2.25}$$

and thus

$$S_\alpha(p) \geq S_{\alpha'}(p) \ \text{ for } \ \alpha \leq \alpha'. \tag{2.26}$$

In addition, we note that a property called the joint convexity of the Rényi α-divergence is true for $0 \leq \alpha \leq 1$ [69]. See also Corollary A.7 in Appendix A for the quantum case.

We finally remark on the Rényi α-divergence for negative α. It can be defined for $-\infty \leq \alpha < 0$ by

$$S_\alpha(p\|q) := \frac{\text{sgn}(\alpha)}{\alpha - 1} \ln \left(\sum_{i=1}^d \frac{p_i^\alpha}{q_i^{\alpha-1}} \right), \tag{2.27}$$

where $\text{sgn}(\alpha) := 1$ for $\alpha > 0$ and $\text{sgn}(\alpha) := -1$ for $\alpha < 0$. It is straightforward to check that $S_\alpha(p\|q) = \frac{\alpha}{\alpha-1} S_{1-\alpha}(q\|p)$ for $\alpha < 0$. Correspondingly, the Rényi α-entropy for negative α can be defined as

$$S_\alpha(p) := \frac{\text{sgn}(\alpha)}{1 - \alpha} \ln \left(\sum_{i=1}^d p_i^\alpha \right) = \text{sgn}(\alpha) \ln d - S_\alpha(p\|u). \tag{2.28}$$

In the following, however, we only consider non-negative α unless stated otherwise.

2.4 General Classical Divergences

We now give the proofs of the properties of the Rényi α-divergence discussed in Sect. 2.3. Moreover, in order to provide a more general perspective, here we discuss general divergence-like quantities including the f-divergence. See Appendix A for the quantum counterpart. We start with the formal definition of convexity/concavity.

Definition 2.1 (*Convexity and concavity*) Let $I \subset \mathbb{R}$ be an interval. A function $f : I \to \mathbb{R}$ is convex, if for any $x, y \in I$ and any $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{2.29}$$

Moreover, f is strictly convex at $z \in I$, if for any $x, y \in I$ with $x \neq y$ and any $\lambda \in (0, 1)$ such that $\lambda x + (1 - \lambda)y = z$,

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y). \tag{2.30}$$

If $\leq$ and $<$ above are replaced by $\geq$ and $>$, f is concave and strictly concave, respectively.

We now state the main lemma of this section, which holds true for vectors in $\mathbb{R}^d$ that are not necessarily probability distributions.

Lemma 2.1 *Let f be a convex function and let $p, q, p', q' \in \mathbb{R}^d$. Suppose that all the components of q, q' are positive. If $p' = Tp$ and $q' = Tq$ hold for a stochastic matrix T, then*

$$\sum_{i=1}^{d} q_i' f\left(\frac{p_i'}{q_i'}\right) \le \sum_{i=1}^{d} q_i f\left(\frac{p_i}{q_i}\right). \tag{2.31}$$

If f is concave, we have the opposite inequality.

Proof Let f be convex. By noting that

$$\frac{p_j'}{q_j'} = \sum_{i=1}^{d} \frac{T_{ji} q_i}{q_j'} \frac{p_i}{q_i}, \quad \sum_{i=1}^{d} \frac{T_{ji} q_i}{q_j'} = 1, \tag{2.32}$$

we have from the Jensen inequality

$$\sum_{j=1}^{d} q_j' f\left(\frac{p_j'}{q_j'}\right) \le \sum_{j=1}^{d} \sum_{i=1}^{d} q_j' \frac{T_{ji} q_i}{q_j'} f\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{d} q_i f\left(\frac{p_i}{q_i}\right). \tag{2.33}$$

We apply the same proof for the concave case. □

Let $p, q \in \mathscr{P}_d$. If we take $f(x) := x \ln x$ in the above lemma, we have $S_1(p\|q) = \sum_{i=1}^{d} q_i f(p_i/q_i)$. Thus inequality (2.31) implies the monotonicity of the KL divergence (2.8). If we take $f(x) = |x - 1|/2$, the trace distance is written as $D(p, q) = \sum_{i=1}^{d} q_i f(p_i/q_i)$, from which we obtain the monotonicity of the trace distance (2.2).

Corollary 2.1 *Let f be a convex function and let $p, p' \in \mathbb{R}^d$. If $p' = Tp$ holds for a doubly stochastic matrix T,*

$$\sum_{i=1}^{d} f(p_i') \le \sum_{i=1}^{d} f(p_i). \tag{2.34}$$

If f is concave, we have the opposite inequality.

Proof Set $q_i = 1$ in Lemma 2.1. □

We now prove the properties of the Rényi α-divergence, which are stated in Sect. 2.3.

Proof of Proposition 2.1. Let $f_\alpha(x) := x^\alpha$. For $1 < \alpha < \infty$, f_α is convex. From the Jensen inequality, we have

$$\sum_{i=1}^{d} q_i f\left(\frac{p_i}{q_i}\right) \ge f\left(\sum_{i=1}^{d} q_i \frac{p_i}{q_i}\right) = f(1) = 1, \tag{2.35}$$

and then take the logarithm of this. The equaity holds if and only if $p_i/q_i = 1$ for all i, because f_α is strictly convex at $x = 1$.

For $0 < \alpha < 1$, f_α is concave, and thus we have the opposite inequality to the above; then take the logarithm of it, by noting the sign of $\alpha - 1$.

For $\alpha = 0, 1, \infty$, we can take the limit to show the non-negativity, but can also easily show it directly. The equality condition can be confirmed directly for these cases. In particular, for the case of $\alpha = 0$, the equality holds if and only if the support of p includes that of q; but under our assumption that the support of q always includes that of p, this condition implies that the supports of them are the same. $\qquad\square$

Proof of Proposition 2.2. For $0 < \alpha < 1$ and $1 < \alpha < \infty$, we apply Lemma 2.1 to $f_\alpha(x) = x^\alpha$ and take the logarithm of it, by noting the sign of $\alpha - 1$. For $\alpha = 0, 1, \infty$, we can take the limit. For $\alpha = 1$, we can also directly take $f(x) = x \ln x$ as mentioned before. $\qquad\square$

Proof of Proposition 2.3. Let $\alpha < \alpha'$. $f(x) := x^{(\alpha-1)/(\alpha'-1)}$ is concave for $1 < \alpha < \alpha' < \infty$, while is convex for $0 < \alpha < \alpha' < 1$ and $0 < \alpha < 1 < \alpha'$. From the Jensen's inequality, and noting the sign of $\alpha - 1$, we obtain

$$S_\alpha(p\|q) = \frac{1}{\alpha - 1} \ln \left(\sum_i p_i \left(\frac{p_i}{q_i} \right)^{\alpha-1} \right) \tag{2.36}$$

$$= \frac{1}{\alpha - 1} \ln \left(\sum_i p_i \left(\frac{p_i}{q_i} \right)^{(\alpha'-1)\frac{\alpha-1}{\alpha'-1}} \right) \tag{2.37}$$

$$\leq \frac{1}{\alpha' - 1} \ln \left(\sum_i p_i \left(\frac{p_i}{q_i} \right)^{\alpha'-1} \right) \tag{2.38}$$

$$= S_{\alpha'}(p\|q). \tag{2.39}$$

For $\alpha, \alpha' = 0, 1, \infty$, we take the limit. $\qquad\square$

We next discuss the concept called the f-divergence introduced in Refs. [14, 15] (see also Ref. [70]). Let $f : (0, \infty) \to \mathbb{R}$ be a convex function. Suppose that $f(x)$ is strictly convex at $x = 1$ and $f(1) = 0$. Then, the f-divergence is defined for $p, q \in \mathcal{P}_d$ by

$$D_f(p\|q) := \sum_{i=1}^d q_i f \left(\frac{p_i}{q_i} \right). \tag{2.40}$$

The f-divergence is non-negative:

$$D_f(p\|q) \geq 0, \tag{2.41}$$

where the equality $D_f(p\|q) = 0$ holds if and only if $p = q$. This follows from the Jensen inequality (2.35) along with $f(1) = 0$, where the equality condition follows from the assumption that $f(x)$ is strictly convex at $x = 1$. The f-divergence also satisfies the monotonicity, which is nothing but Lemma 2.1.

The KL divergence is the f-divergence with $f(x) = x \ln x$. On the other hand, the Rényi α-divergence with $\alpha \neq 1$ is not in the form of f-divergence. We note that there is another concept also called α-divergence [20], which is the f-divergence with $f_\alpha(x) := \frac{1}{\alpha(\alpha-1)}(x^\alpha - 1)$ for $\alpha \neq 0, 1, \infty$, that is,

$$D_{f_\alpha}(p\|q) := \frac{1}{\alpha(\alpha-1)}\left(\sum_i \frac{p_i^\alpha}{q_i^{\alpha-1}} - 1\right). \tag{2.42}$$

It can be shown that $\lim_{\alpha\to 1} D_{f_\alpha}(p\|q) = S_1(p\|q)$. In this book, however, whenever we simply mention "the α-divergence," it indicates the Rényi α-divergence in the sense of Sect. 2.3.

The trace distance is the f-divergence with $f(x) := |x-1|/2$. The f-divergence is also related to the fidelity defined as $F(p,q) := \sum_i \sqrt{p_i q_i}$. To see this, let $f(x) := 1 - \sqrt{x}$, which is equivalent to above-mentioned f_α with $\alpha = 1/2$ up to normalization. Then we have $D_f(p\|q) = 1 - F(p,q)$, from which the monotonicity of the fidelity follows, i.e., $F(p,q) \leq F(Tp, Tq)$. It can be also rewritten as $D_f(p\|q) = \frac{1}{2}\sum_i (\sqrt{p_i} - \sqrt{q_i})^2$, and $\sqrt{D_f(p\|q)}$ is called the Hellinger distance.

We finally note that, in general, a *divergence* $D(p\|q)$ is defined as a quantity satisfying $D(p\|q) \geq 0$ where the equality holds if and only if $p = q$. The monotonicity under stochastic maps is also often required for the definition of divergence. The f-divergence and the Rényi α-divergence with $0 < \alpha \leq \infty$ are both divergences in this sense, while the Rényi 0-divergence does not satisfy the above equality condition.

2.5 Fisher Information

We here briefly discuss the Fisher information [18, 20] (see also Ref. [3]), which has a fundamental connection to divergences. The quantum counterpart will be discussed in Appendix A.4.

A main practical application of the Fisher information is found in the theory of parameter estimation, where one wants to estimate unknown parameters of probability distributions from observed data of random events. Here, the Fisher information gives a fundamental bound of the accuracy of parameter estimation, which is known as the Cramer-Rao bound. In this section, instead of going into details of estimation theory, we will focus on the general mathematical properties of the Fisher information.

We consider smooth parametrization of probability distributions, written as $p(\theta) \in \mathscr{P}_d$ with parameters $\theta := (\theta^1, \theta^2, \ldots, \theta^m) \in \mathbb{R}^m$, where the domain of θ is an open subset of $\mathbb{R}^m$. We suppose that $p(\theta)$ has the full rank, i.e., $p_i(\theta) > 0$ for any i and θ. We denote $\partial_k := \partial/\partial\theta^k$. We note that the parametrization must satisfy $m \leq d - 1$.

Definition 2.2 (*Fisher information*) Let $p(\theta) \in \mathscr{P}_d$ have full rank and $\theta \in \mathbb{R}^m$ be the parameter. The (classical) Fisher information matrix is an $m \times m$ matrix, whose (k, l)-component is defined as

$$J_{p(\theta),kl} := \sum_{i=1}^d p_i(\theta)\partial_k[\ln p_i(\theta)]\partial_l[\ln p_i(\theta)] = \sum_{i=1}^d \frac{\partial_k p_i(\theta)\partial_l p_i(\theta)}{p_i(\theta)}. \tag{2.43}$$

The Fisher information is obtained as the infinitesimal limit of the f-divergence. Suppose that $f''(1) > 0$ exists and f is sufficiently smooth around 1. Then it is easy to check that

$$D_f(p(\theta)\|p(\theta - \Delta\theta)) = \frac{f''(1)}{2}\sum_{kl} \Delta\theta^k J_{p(\theta),kl}\Delta\theta^l + O(\varepsilon^3), \qquad (2.44)$$

where $\varepsilon := \|\Delta\theta\|$. We note that the term of $O(\varepsilon)$ vanishes because of $\sum_i \partial_k p_i(\theta) = 0$. Equation (2.44) is a generalization of Eq. (2.7); remarkably, for all the f-divergences, we obtain the same Fisher information up to normalization.

The Fisher information satisfies the monotonicity under stochastic maps. While this may be regarded as a trivial consequence of the monotonicity of the f-divergence, we provide a direct proof as follows.

Proposition 2.4 (Monotonicity of the Fisher information) *For any stochastic map T that is independent of θ,*

$$J_{p(\theta)} \geq J_{Tp(\theta)}. \qquad (2.45)$$

Proof We omit the argument θ for simplicity of notations. Let $p' := Tp$. Let $c = (c^1, \ldots, c^m) \in \mathbb{R}^m$ be a column vector and define $\partial := \sum_k c^k \partial_k$. Then,

$$c^{\mathrm{T}} J_p c = \sum_i p_i \left(\frac{\partial p_i}{p_i}\right)^2, \quad c^{\mathrm{T}} J_{p'} c = \sum_i p'_i \left(\frac{\partial p'_i}{p'_i}\right)^2. \qquad (2.46)$$

Then, we apply inequality (2.31) with $f(x) := x^2$. □

An operational meaning of the Fisher information is highlighted by the Cramer-Rao bound, which states that the accuracy of any unbiased estimation of θ is bounded by the Fisher information. Consider a task that one estimates unknown θ from observed data i. Let $\theta_{\mathrm{est}}(i)$ be an estimator and suppose that it satisfies the unbiasedness condition: $\sum_i p_i(\theta)\theta_{\mathrm{est}}(i) = \theta$ for all θ. The accuracy of such unbiased estimation can be characterized by the covariance matrix $\mathrm{Cov}_\theta(\theta_{\mathrm{est}})$, whose (k, l)-component is given by

$$\mathrm{Cov}_\theta^{kl}(\theta_{\mathrm{est}}) := \sum_i p_i(\theta)(\theta_{\mathrm{est}}^k(i) - \theta^k)(\theta_{\mathrm{est}}^l(i) - \theta^l). \qquad (2.47)$$

Then, the Cramer-Rao bound states that

$$\mathrm{Cov}_\theta(\theta_{\mathrm{est}}) \geq J_{p(\theta)}^{-1}. \qquad (2.48)$$

We omit the proof of this, which is not difficult (e.g., Theorem 11.10.1 of Ref. [3]).

As an example, we consider a family of probability distributions called the exponential family. For simplicity, we consider a single parameter $\theta \in \mathbb{R}$ and the parameterized distribution of the form

$$p_i(\theta) := h_i \exp(\theta T_i - A(\theta)), \tag{2.49}$$

where $A(\theta)$ is a smooth function of θ. By straightforward computation, we have

$$\sum_i T_i p_i(\theta) = A'(\theta), \quad \sum_i T_i^2 p_i(\theta) = A''(\theta) + A'(\theta)^2, \tag{2.50}$$

where $A'(\theta) := dA(\theta)/d\theta$. Thus, the Fisher information is given by $J_{p(\theta)} = A''(\theta)$.

In terms of thermodynamics, we can interpret that $p_i(\theta)$ is a Gibbs state, where T_i is the energy, $-\theta$ is the inverse temperature β, and $\theta^{-1}A(\theta)$ is the free energy $F(\beta)$ (we set $h_i = 1$). With this correspondence, Eq. (2.50) are well-known formulas in equilibrium statistical mechanics.

We note that Eq. (2.50) imply that T_i is regarded as an unbiased estimator of $A'(\theta)$ with the variance $A''(\theta)$. Then, we replace the parameter θ with $\theta' := A'(\theta)$. The corresponding Fisher information is given by $J_{p(\theta')} = A''(\theta)^{-1}$ because of $d/d\theta' = (A''(\theta))^{-1}d/d\theta$. Therefore, T_i as an unbiased estimator of θ' attains the Cramer-Rao bound (2.48).

We finally remark that the Fisher information can be regarded as a *metric* on the space of probability distributions (or on the parameter space), which is the perspective of *information geometry* [20] (see Ref. [71] for its relation to stochastic thermodynamics). Here, we introduce the concept called *monotone metric*.

Definition 2.3 (*Monotone metric*) Suppose that $G_p : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is defined for distributions $p \in \mathscr{P}_d$ with full support. We call G_p a monotone metric, if it satisfies the following.

- G_p is bilinear.
- $G_p(a, a) \geq 0$ holds for any p, where the equality is achieved if and only if $a = 0$.
- $p \mapsto G_p(a, a)$ is continuous for any a.
- The monotonicity

$$G_p(a, a) \geq G_{Tp}(Ta, Ta) \tag{2.51}$$

holds for any stochastic matrix T and for any p, a.

In particular, the *Fisher information metric* is defined by

$$G_p^{\mathrm{F}}(a, b) := \sum_{i=1}^{d} \frac{a_i b_i}{p_i}, \tag{2.52}$$

where $a = (a_1, \ldots, a_d)^{\mathrm{T}}$ and $b = (b_1, \ldots, b_d)^{\mathrm{T}}$. This metric is related to the Fisher information matrix as

$$J_{p(\theta),kl} = G^{\mathrm{F}}_{p(\theta)}(\partial_k p(\theta), \partial_l p(\theta)). \tag{2.53}$$

In completely the same manner as the proof of Proposition 2.4, we can show that the Fisher information metric (2.52) satisfies the monotonicity (2.51). Thus, the Fisher information metric is a monotone metric. Conversely, any monotone metric is the Fisher information metric up to an unimportant additional term and normalization; the Fisher information is essentially unique. This is known as the Chentsov's theorem.

Theorem 2.1 (Chentsov's Theorem [72]) *Any monotone metric G_p is given by the Fisher information metric G^{F}_p as*

$$G_p(a, b) = k G^{\mathrm{F}}_p(a, b) + k' \left(\sum_{i=1}^{d} a_i \right) \left(\sum_{i=1}^{d} b_i \right), \tag{2.54}$$

where $k > 0$ and $k' \geq 0$ are some constants.

We note that the second term on the right-hand side of Eq. (2.54) vanishes if the domain of G_p is restricted to the space satisfying $\sum_{i=1}^{d} a_i = \sum_{i=1}^{d} b_i = 0$.

Chapter 3
Classical Majorization

Majorization is a useful tool in various fields of information theory [7, 55]. In the context of thermodynamics, a generalized majorization plays a crucial role for characterizing a necessary and sufficient condition for state conversion by thermodynamically feasible transformations described by Gibbs-preserving maps at finite temperature.

In Sect. 3.1, we start with a simplest case: (ordinary) majorization related to infinite-temperature thermodynamics. In Sect. 3.2, we consider a generalized version of majorization, called thermo-majorization, or more generally, d-majorization (relative majorization). In Sect. 3.3, we briefly mention majorization in the presence of "catalyst." In Sect. 3.4, we roughly discuss the continuous-variable majorization, without going into mathematical details. In Sect. 3.5, we provide rigorous proofs of main theorems of this chapter.

3.1 Majorization

We first consider *majorization* for classical probability distributions, which characterizes state convertibility in a simplest thermodynamic setup where the Gibbs state is just the uniform distribution (i.e., all the energy levels are degenerate or the temperature is infinite, $\beta = 0$).

Let us illustrate a motivation to introduce majorization (see also Sect. 2.2). Remember that if a classical distribution p is mapped to p' by a doubly stochastic map, the Shannon entropy increases (or does not change), $S_1(p) \leq S_1(p')$. Then, the central problem of this section is related to the converse: If the Shannon entropy increases, does there always exist a doubly stochastic map that converts p to p'? The answer is negative; The Shannon entropy does not give a sufficient condition for such state conversion. In other words, the Shannon entropy is a monotone but
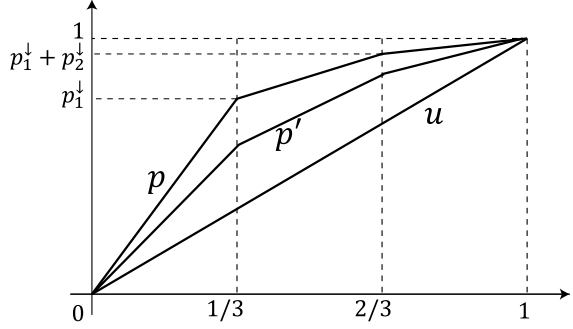
not a complete monotone. Instead, the full characterization of state convertibility in a necessary and sufficient way is given by majorization.

Let p and p' be classical probability distributions. We define $p^{\downarrow}$ by rearranging the components of $p = (p_1, p_2, \ldots, p_d)^{\mathrm{T}}$ in the decreasing order: $p_1^{\downarrow} \geq p_2^{\downarrow} \geq \cdots \geq p_d^{\downarrow}$. We also define $p_i'^{\downarrow}$ in the same manner. Majorization is then defined as follows.

Definition 3.1 (*Majorization*) Let $p, p' \in \mathscr{P}_d$. We say that p majorizes p', written as $p' \prec p$, if for all $k = 1, 2, \ldots, d$,

$$\sum_{i=1}^{k} p_i'^{\downarrow} \leq \sum_{i=1}^{k} p_i^{\downarrow}. \tag{3.1}$$

This definition implies that p' is "more random" or "more uniformly distributed" than p. To visualize this, we can rephrase Definition 3.1 by the *Lorenz curve* as follows. As shown in Fig. 3.1, we plot $1/d, 2/d, \ldots, 1$ (at equal spaces) on the horizontal axis, and plot $p_1^{\downarrow}, p_1^{\downarrow} + p_2^{\downarrow}, \ldots, p_1^{\downarrow} + \cdots + p_d^{\downarrow}(= 1)$ on the vertical axis. We obtain a concave polyline by connecting these points, which is called the Lorenz curve of p. It is obvious from the definition that $p' \prec p$ holds if and only if the Lorenz curve of p lies above that of p'.

The Lorenz curve of the uniform distribution u is the diagonal line between $(0, 0)$ and $(1, 1)$. Thus, $u \prec p$ holds for all p. We remark that $\prec$ is not a total order but a preorder. In fact, $p \prec p$ holds for any p, and if $p'' \prec p'$ and $p' \prec p$ then $p'' \prec p$ holds. However, there is a pair of distributions p, p' such that neither $p' \prec p$ nor $p \prec p'$ holds; this happens when the Lorenz curves of p and p' cross with each other.

The definition of majorization can be equivalently rephrased in several ways as follows (see also, e.g., Ref. [7]):

Theorem 3.1 *Let $p, p' \in \mathscr{P}_d$. The following are equivalent.*

(i) $p' \prec p$.
(ii) *For all $t \in \mathbb{R}$,*

$$\sum_{i=1}^{d} |p_i' - t| \leq \sum_{i=1}^{d} |p_i - t|. \tag{3.2}$$

(iii) For all convex functions f,

$$\sum_{i=1}^{d} f(p_i') \leq \sum_{i=1}^{d} f(p_i). \tag{3.3}$$

(iv) There exists a doubly stochastic matrix T such that $p' = Tp$.

We will present the proof of this theorem in Sect. 3.5. Here we only remark on the following point. Given (iv) above, (iii) is just the monotonicity in the form of Corollary 2.1. In particular, by taking $f(x) := x \ln x$, inequality (3.3) implies that the Shannon entropy does not decrease under doubly stochastic maps. Thus, Theorem 3.1 implies that the Shannon entropy $S_1(p)$ does not provide a sufficient condition for state convertibility under doubly stochastic maps; we need to take into account all convex functions f to obtain a sufficient condition. In terms of resource theory, $\sum_{i=1}^{d} f(p_i)$ with all convex functions constitute a complete set of monotones for doubly stochastic maps.

We note an explicit example that $S_1(p) \leq S_1(p')$ is not sufficient for the existence of a doubly stochastic matrix T such that $p' = Tp$ [46]. Let $p = (2/3, 1/6, 1/6)^{\mathrm{T}}$ and $p' = (1/2, 1/2, 0)^{\mathrm{T}}$. In this case, it is easy to check that $S_1(p) < S_1(p')$ holds, while $p' \prec p$ does not.

We note that the uniform distribution u is the fixed point of doubly stochastic matrices: $u = Tu$. In terms of thermodynamics, u is regarded as the Gibbs state of a system with all the energy levels being degenerate or at infinite temperature $\beta = 0$; a doubly stochastic map represents a thermodynamic process that does not change such a special Gibbs state u.

We remark the following theorem as a fundamental characterization of doubly stochastic matrices in the classical case.

Theorem 3.2 (Birkhoff's theorem) *Every extreme point of the set of doubly stochastic matrices is a permutation matrix. That is, the following are equivalent.*

(i) T is a doubly stochastic matrix.
(ii) T can be written as a convex combination of permutation matrices: There exist permutation matrices P_k and coefficients $r_k \geq 0$ with $\sum_k r_k = 1$ such that $T = \sum_k r_k P_k$.

Proof (ii) $\Rightarrow$ (i) is trivial. (i) $\Rightarrow$ (ii) is Theorem II.2.3 of Ref. [7], whose proof is not very easy. $\qquad\square$

Here, we note that the set of doubly stochastic matrices is compact and convex, and any element of it can be written as a convex combination of extreme points (cf. the Krein-Milman theorem).

So far, we have considered majorization of probability distributions in $\mathscr{P}_d$. More generally, we can define majorization of vectors in $\mathbb{R}^d$: For $p, p' \in \mathbb{R}^d$, we write $p' \prec p$ if inequality (3.1) and $\sum_{i=1}^d p_i = \sum_{i=1}^d p_i'$ are satisfied. Theorem 3.1 still holds under this definition, where the proof goes in completely the same manner. From this viewpoint, we remark the following proposition for characterization of doubly stochastic matrices:

Proposition 3.1 (Theorem II.1.9 of [7]) *A matrix T is doubly stochastic if and only if $Tp \prec p$ for all $p \in \mathbb{R}^d$.*

Proof The "only if" part is (iv) $\Rightarrow$ (i) of Theorem 3.3. We can prove the "if" part by choosing p to be the uniform distribution u and the distributions of the form $(0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}$. $\qquad\qquad\square$

We finally remark on the concept called *Schur-convexity*. A function $F : \mathbb{R}^d \to \mathbb{R}$ is called a Schur-convex function, if $p' \prec p$ implies $F(p') \leq F(p)$ for any $p, p' \in \mathbb{R}^d$. In other words, a Schur-convex function is a monotone of majorization. For example, $F(p) := \sum_{i=1}^d f(p_i)$ with f being a convex function is Schur-convex from Theorem 3.1 (iii). The following proposition is known as a characterization of Schur-convex functions.

Proposition 3.2 (Theorem II.3.14 of *[7]*) *A differentiable function $F : \mathbb{R}^d \to \mathbb{R}$ is Schur-convex, if and only if F is permutation invariant (i.e., $F(Pp) = F(p)$ for all permutation P) and for all $p \in \mathbb{R}^d$ and i, j,*

$$(p_i - p_j)\left(\frac{\partial F}{\partial p_i} - \frac{\partial F}{\partial p_j}\right) \geq 0. \tag{3.4}$$

3.2   d-Majorization and Thermo-Majorization

We next consider a generalization of majorization, called *d-majorization* (or *relative majorization*) for classical distributions (see also, e.g., Refs. [43, 55, 73]). This concept was originally introduced as a generalization of majorization in the context of statistical comparison [56–58]. While ordinary majorization concerns convertibility from a distribution to another distribution, d-majorization concerns convertibility between pairs of distributions. A special class of d-majorization is called *thermo-majorization* [41], which characterizes a necessary and sufficient condition for state convertibility by thermodynamic processes at finite temperature $\beta > 0$.

We first define d-majorization by using the (generalized or relative) Lorenz curve. We consider classical probability distributions $p = (p_1, \ldots, p_d)^{\mathrm{T}}$ and $q = (q_1, \ldots, q_d)^{\mathrm{T}}$. We define p^* and q^* by rearranging their components such that $p_1^*/q_1^* \geq p_2^*/q_2^* \geq \cdots \geq p_d^*/q_d^*$ holds, where the ways of rearranging the components are the same for p and q. As shown in Fig. 3.2a, we plot $q_1^*, q_1^* + q_2^*, \ldots, q_1^* + \cdots + q_d^*(= 1)$ on the horizontal axis and plot $p_1^*, p_1^* + p_2^*, \ldots, p_1^* + \cdots + p_d^*(= 1)$
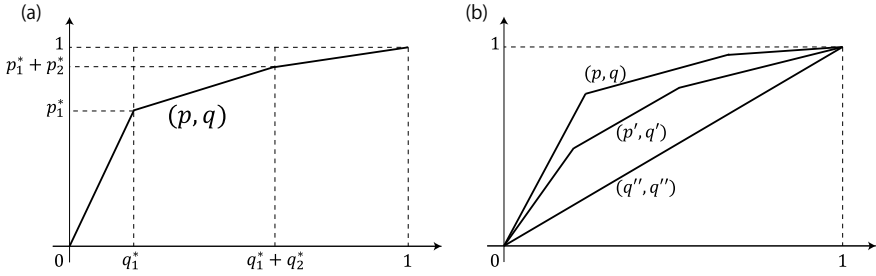
Fig. 3.2 Lorenz curves for d-majorization ($d = 3$). **a** An example of the Lorenz curve of (p, q). **b** An example of a situation where $(p', q') \prec (p, q)$. The diagonal line describes (q'', q'')

on the vertical axis. By connecting these points, we obtain a concave polyline, which is the (generalized) Lorenz curve of the pair (p, q).

Definition 3.2 Let $p, q, p', q' \in \mathscr{P}_d$. We say that (p, q) d-majorizes (p', q'), written as $(p', q') \prec (p, q)$, if the Lorenz curve of (p, q) lies above that of (p', q').

Figure 3.2b shows an example of the Lorenz curves with $(p', q') \prec (p, q)$. The diagonal line describes the Lorenz curve of (q'', q'') for any q'', which is d-majorized by all other (p, q).

As is the case for ordinary majorization, the definition of d-majorization can be equivalently rephrased in several ways as follows, which is referred to as the Blackwell's theorem [56] (see also Ref. [57]):

Theorem 3.3 (Blackwell's theorem) *Let $p, q, p', q' \in \mathscr{P}_d$ and suppose that q, q' have full rank. Then, the following are equivalent.*

(i) *$(p', q') \prec (p, q)$.*
(ii) *For all $t \in \mathbb{R}$,*

$$\sum_{i=1}^{d} |p_i' - tq_i'| \leq \sum_{i=1}^{d} |p_i - tq_i|. \tag{3.5}$$

(iii) *For all convex functions f,*

$$\sum_{i=1}^{d} q_i' f\left(\frac{p_i'}{q_i'}\right) \leq \sum_{i=1}^{d} q_i f\left(\frac{p_i}{q_i}\right). \tag{3.6}$$

(iv) *There exists a stochastic matrix T such that $p' = Tp$ and $q' = Tq$.*

Clearly, Theorem 3.1 for majorization is regarded as a special case of Theorem 3.3 for d-majorization, by letting $q = q' = u$. While we will postpone the proof of Theorem 3.3 to Sect. 3.5, we here make some remarks on the proof.

The most nontrivial part of the above theorem is (i) $\Rightarrow$ (iv). A way to obtain an intuition about it is to consider majorization for continuous variables as discussed in Ref. [58]. In fact, by applying a variable transformation to the Lorenz curve of a continuous variable, we can see that d-majorization is a special case of continuous (ordinary) majorization. We will discuss this idea in Sect. 3.4 in detail, by which our rigorous proof in Sect. 3.5 is inspired. We also note that there is an alternative direct and "graphical" proof of (i) $\Rightarrow$ (iv) of Theorem 3.3, where one does not even need to invoke (i) $\Rightarrow$ (iv) of Theorem 3.1, as shown in Ref. [74].

Given (iv) above, (iii) is just the monotonicity in the form of Lemma 2.1. The KL divergence with $f(x) = x \ln x$ does not provide a sufficient condition for convertibility of distributions; instead, we need all convex functions f. In terms of resource theory, $\sum_{i=1}^d q_i f(p_i/q_i)$ with all convex functions constitute a complete set of monotones for stochastic maps. We note that $\sum_{i=1}^d q_i f(p_i/q_i)$ is the f-divergence defined in Eq. (2.40), if $f(1) = 0$ is satisfied (this is always possible by adding a constant) and $f(x)$ is strictly convex at $x = 1$.

If $q = q'$, d-majorization is called thermo-majorization, where the condition (iv) of Theorem 3.3 reduces to $p' = Tp$ and $q = Tq$, that is, q is a fixed point of T.

Definition 3.3 (*Thermo-majorization*) Let $p, q, p' \in \mathscr{P}_d$. We say that p thermo-majorizes p' with respect to q, if $(p', q) \prec (p, q)$.

Because the Lorenz curve of (q, q) is the diagonal line, q is thermo-majorized by any other distributions with respect to q itself: $(q, q) \prec (p, q)$ for all p. Therefore, the thermo-majorization relation characterizes how "close" a distribution p is to q.

By the naming of thermo-majorization, we have in mind that q is a Gibbs state of a Hamiltonian, written as $q = p^G$. We note that any distribution q of full rank is regarded as a Gibbs state of some Hamiltonian. In this case, $q = Tq$ implies that T does not change the Gibbs state, which is called a Gibbs-preserving map (see also Chap. 4). In terms of resource theory of thermodynamics, Gibbs-preserving maps are specified as free operations (and Gibbs states are free states). Note that at infinite temperature, the Gibbs state is u and a Gibbs-preserving map is a doubly-stochastic map.

We next remark that the min and the max divergences, $S_0(p\|q)$ and $S_\infty(p\|q)$, can be visualized by the Lorenz curve of (p, q), as illustrated in Fig. 3.3. The following theorem straightforwardly follows from this graphical representation.

Theorem 3.4 (Conditions for state conversion)

(a) *Necessary conditions: If $(p', q') \prec (p, q)$, then*

$$S_0(p\|q) \geq S_0(p'\|q'), \quad S_\infty(p\|q) \geq S_\infty(p'\|q'). \tag{3.7}$$

(b) *Sufficient condition: $(p', q') \prec (p, q)$ holds, if (but not only if)*

$$S_\infty(p'\|q') \leq S_0(p\|q). \tag{3.8}$$

Fig. 3.3 Graphical representation of the relationship between the Lorenz curve of (p, q) and the divergences $S_0(p\|q)$ and $S_\infty(p\|q)$, where $e^{-S_0(p\|q)} = \sum_{i:p_i>0} q_i$ and $e^{-S_\infty(p\|q)} = (\max_i \{p_i/q_i\})^{-1}$
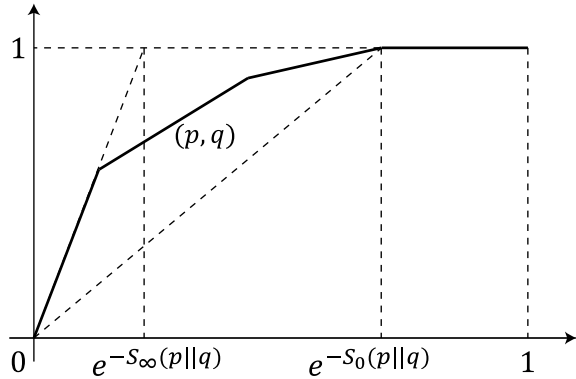
Proof These are obvious from Fig. 3.3. We note that (a) is nothing but the monotonicity of $S_0(p\|q)$ and $S_\infty(p\|q)$, given Theorem 3.3. □

In the above theorem, the necessary condition and the sufficient condition are distinct, suggesting that there is no single necessary and sufficient condition for d-majorization in terms of the Rényi divergences (see also Fig. 6.1 in Sect. 6.3). However, if we take the asymptotic limit, the Rényi 0- and ∞-divergences can collapse to a single value for a broad class of distributions, giving a complete monotone for d-majorization. This is the main topic of Sect. 7.2. We also note that the sufficient condition (b) is of no use when p and q have the same support, while it will turn out useful in Chap. 7 by considering smoothing of S_0 and taking the asymptotic limit.

Finally, we note that, as is the case for ordinary majorization, the concept of d-majorization can be extended to general $p, q, p', q' \in \mathbb{R}^d$, where $q_i > 0, q'_i > 0$ for all i and $\sum_{i=1}^d q_i = \sum_{i=1}^d q'_i$ are always assumed. Then, we write $(p', q') \prec (p, q)$ if the Lorenz curve of (p, q) lies above that of (p', q') and $\sum_{i=1}^d p_i = \sum_{i=1}^d p'_i$ is satisfied. Under this definition, Theorem 3.3 still holds.

3.3 Catalytic Majorization

It is useful to introduce a *catalyst* for state transformations, where a catalyst means an auxiliary system whose states are the same before and after the transformation. Here we briefly introduce some important results about majorization in the presence of catalyst, without going into the proofs. This section is logically independent of the subsequent sections.

We first define catalytic majorization, where, as mentioned later, the condition that the initial and final states of the catalyst are exactly the same is crucial.

Definition 3.4 (*Catalytic majorization*) Let $p, p' \in \mathscr{P}_d$. We say that p majorizes p' with the aid of a catalyst, or p can be *trumped* into p', if there exists a finite-dimensional distribution $r \in \mathscr{P}_N$ such that $p' \otimes r \prec p \otimes r$.

An example that $p' \prec p$ does not hold but p can be trumped into p' is given by $p = (4/10, 4/10, 1/10, 1/10)^{\mathrm{T}}$ and $p' = (1/2, 1/4, 1/4, 0)^{\mathrm{T}}$, along with $r = (6/10, 4/10)^{\mathrm{T}}$ [75].

A necessary and sufficient condition for catalytic majorization was proved in Refs. [76, 77], which is stated as follows.

Theorem 3.5 (Theorem 1 of [76]; Theorem 2 of [77]) *Let* $p, p' \in \mathscr{P}_d$ *have full rank and suppose that* $p^{\downarrow} \neq p'^{\downarrow}$. *Then,* p *can be trumped into* p', *if and only if* $f_\alpha(p') < f_\alpha(p)$ *for all* $\alpha \in (-\infty, \infty)$, *where*

$$
f_\alpha(p) := \begin{cases}
\ln \sum_i p_i^\alpha & (\alpha > 1), \\
\sum_i p_i \ln p_i & (\alpha = 1), \\
-\ln \sum_i p_i^\alpha & (0 < \alpha < 1), \\
-\sum_i \ln p_i & (\alpha = 0), \\
\ln \sum_i p_i^\alpha & (\alpha < 0).
\end{cases}
\tag{3.9}
$$

In Theorem 3.5, f_α is the same as the Rényi α-entropy (2.28) (including negative α) up to a negative coefficient, except for $\alpha = 0$. By allowing an infinitesimal error in the final state, the $\alpha = 0$ case turns into the Rényi 0-entropy as follows.

Lemma 3.1 (Proposition 4 of [43]) *Let* $p, p' \in \mathscr{P}_d$. *The following are equivalent.*

(i) *For any* $\varepsilon > 0$, *there exists a distribution* p'_ε *such that* p *can be trumped into* p'_ε *and* $D(p', p'_\varepsilon) \leq \varepsilon$.
(ii) $S_\alpha(p) \leq S_\alpha(p')$ *holds for all* $\alpha \in (-\infty, \infty)$.

A slightly different definition of approximate trumping from the above has been considered in Ref. [78], where the necessary and sufficient condition is given by $S_\alpha(p) \leq S_\alpha(p')$ only with $\alpha \in (1, \infty)$ (Theorem 1 of Ref. [78]; see also discussion in Ref. [43]).

We next consider the catalytic d-majorization.

Definition 3.5 (*Catalytic d-majorization*) Let $p, q, p', q' \in \mathscr{P}_d$. We say that (p, q) d-majorizes (p', q') with the aid of catalysts, or (p, q) can be *d-trumped* into (p', q'), if there exist finite-dimensional distributions $r, s \in \mathscr{P}_N$ such that $(p' \otimes r, q' \otimes s) \prec (p \otimes r, q \otimes s)$.

In the above definition, it is often possible to take s as the uniform distribution u. In the case of thermo-majorization, this implies that the Hamiltonian of the catalyst can be taken trivial. In fact, Lemma 3.1 can be further generalized to the following form.

Theorem 3.6 (Theorem 17 of [43]) *Let* $p, q, p', q' \in \mathscr{P}_d$ *have full rank. Then, the following are equivalent.*

(i) *For any* $\varepsilon > 0$, *there exists a distribution* $p'_\varepsilon \in \mathscr{P}_d$ *such that* (p, q) *can be d-trumped into* (p'_ε, q') *and* $D(p', p'_\varepsilon) \leq \varepsilon$.

(ii) $S_\alpha(p\|q) \geq S_\alpha(p'\|q')$ *for all $\alpha \in (-\infty, \infty)$.*

Moreover, the catalyst distribution s of Definition 3.5 for the d-trumping of (i) can be taken as the uniform distribution u.

The above theorem implies that the Rényi α-divergences with $-\infty < \alpha < \infty$ constitute a complete set of monotones in the presence of catalyst.

Once we remove the requirement that the catalyst state exactly goes back to the initial state without any correlation with the system, the characterization of state convertibility becomes drastically different. In fact, any state conversion becomes possible in the absence of this requirement [79], even if the error in the final state is arbitrarily small, as long as it is independent of the sizes of the system and the catalyst. In other words, the majorization structure becomes trivial in the presence of such non-exact catalyst, which is called the embezzling phenomenon. In the context of thermodynamics [43], this implies that one can extract more work than the standard thermodynamic bound by "embezzling" an auxiliary system whose operation is not exactly cyclic.

Theorem 3.7 (Main result of [79]) *For any $\varepsilon > 0$ and for any p, $p' \in \mathscr{P}_d$, there exist a distribution $r \in \mathscr{P}_N$ of a catalyst and a distribution $r' \in \mathscr{P}_{dN}$ such that $r' \prec p \otimes r$ and $D(r', p' \otimes r) < \varepsilon$.*

On the other hand, it is known that there are several ways to make the non-exact majorization structure nontrivial by considering slightly different setups from the case of the embezzling phenomenon; We now suppose that the change of the catalyst state is "modestly" non-exact. Remarkably, the KL divergence often serves as a single "complete monotone" for such setups, as described below. Therefore, in the presence of such modestly non-exact catalyst, thermodynamic transformation can be completely characterized by the standard nonequilbrium free energy (4.9) introduced in Sect. 4.1 based on the KL divergence.

First, in Ref. [43], it has been shown that if the approximation error scales modestly in the dimension of the catalyst, then the majorization condition is only given by the Shannon entropy up to small correction terms.

Theorem 3.8 (Theorem 23 of [43]) *Let p, $p' \in \mathscr{P}_d$.*

(a) Let $\varepsilon \geq 0$. If there exist a distribution $r \in \mathscr{P}_N$ of a catalyst and a distribution $r' \in \mathscr{P}_{dN}$ such that $r' \prec p \otimes r$ and $D(r', p' \otimes r) \leq \varepsilon/(2 \ln N)$, then

$$S_1(p) \leq S_1(p') - \varepsilon - \varepsilon \frac{\ln d}{\ln N} - h\left(\frac{\varepsilon}{\ln N}\right), \tag{3.10}$$

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

(b) If $S_1(p) < S_1(p')$, then for any sufficiently large N, there exist a distribution $r \in \mathscr{P}_N$ of a catalyst and a distribution $r' \in \mathscr{P}_{dN}$ such that $r' \prec p \otimes r$ and $D(r', p' \otimes r) \leq \exp(-c\sqrt{\ln N})$ for some constant $c > 0$.

We next consider the situation where several subsystems in the catalyst can form correlations in the final state [80, 81]. This implies that independence of subsystems in the catalyst is a resource of state transformation. The case of majorization is given by the following theorem (see Theorem 1 of Ref. [81] for the thermo-majorization case).

Theorem 3.9 (Theorem 1 of [80]) *Let $p, p' \in \mathscr{P}_d$ with $p^\downarrow \neq p'^\downarrow$. The following are equivalent.*

(i) *There exist distributions $r_1, r_2, \ldots, r_k \in \mathscr{P}_N$ of a k-partite catalyst and a distribution $r \in \mathscr{P}_{kN}$ with marginals $r_1, r_2, \ldots, r_k$, such that $p' \otimes r \prec p \otimes r_1 \otimes r_2 \otimes \cdots \otimes r_k$.*

(ii) *$S_0(p) \leq S_0(p')$ and $S_1(p) < S_1(p')$ hold.*

Moreover, we can choose $k = 3$.

Finally, we consider the case that a small amount of correlation between the system and the catalyst is allowed in the final state, while the marginal state of the catalyst should be exactly the same as the initial state. This is referred to as correlated-catalytic transformation [82, 83].

Theorem 3.10 (Main theorem of [84]) *Let $p, p' \in \mathscr{P}_d$ with $p^\downarrow \neq p'^\downarrow$. The following are equivalent.*

(i) *There exist a distribution $r \in \mathscr{P}_N$ of a catalyst and a distribution $r' \in \mathscr{P}_{dN}$, such that $r' \prec p \otimes r$ and the marginal distributions of r' equal p' and r.*

(ii) *$S_0(p) \leq S_0(p')$ and $S_1(p) < S_1(p')$ hold.*

Moreover, for any $\delta > 0$, we can take r' with which the mutual information between the system and the catalyst is smaller than δ.

The above theorem can be generalized to d-majorization (Theorem 1 of Ref. [85]). Furthermore, it can be generalized to approximate d-majorization, as proved in Theorem 7 of Ref. [84] and Theorem 2 of Ref. [85], which is characterized only by the KL divergence.

Theorem 3.11 *Let $p, p', q, q' \in \mathscr{P}_d$ and suppose that q, q' have full rank. Then, the following are equivalent.*

(i) *For any $\varepsilon > 0$, there exist a distribution $p'_\varepsilon \in \mathscr{P}_d$, distributions $r_\varepsilon, s \in \mathscr{P}_N$ of a catalyst, and a distribution $r'_\varepsilon \in \mathscr{P}_{dN}$, such that $(r'_\varepsilon, q' \otimes s) \prec (p \otimes r_\varepsilon, q \otimes s)$, the marginals of r'_ε equal p'_ε and r_ε, and $D(p', p'_\varepsilon) \leq \varepsilon$.*

(ii) *$S_1(p\|q) \geq S_1(p'\|q')$ holds.*

Moreover, in (i), we can take s as the uniform distribution u. Also, for any $\delta > 0$, we can take r'_ε (for a given ε) with which the mutual information between the system and the catalyst is smaller than δ.

Theorem 3.11 was originally raised as a conjecture for the quantum setup in Ref. [83] (see also Ref. [86]). We note that Theorem 3.10 straightforwardly applies to the quantum case (i.e., quantum majorization defined in Sect. 6.1) as discussed in Ref. [84]. The quantum version of Theorem 3.11 has been proved in Ref. [87], except that the catalyst state (written as s in the classical case above) can be taken to be the uniform distribution. The proof in Ref. [87] is based on the asymptotic theory discussed in Sect. 7.3, especially Theorem 7.7, where the KL divergence appears as a single complete monotone as well.

3.4 Continuous Variable Case

We next consider majorization and d-majorization for continuous variables, i.e., for infinite-dimensional spaces (see also Refs. [88, 89] for mathematical details). In this case, random variable x takes continuous values, and correspondingly, we write the probability *density* as $p(x)$.

For simplicity, let $x \in [0, 1]$. We denote the set of L^1-functions on $[0, 1]$ just by L^1. Let μ be the Lebesgue measure on $[0, 1]$, and the following arguments should be read as "almost everywhere" with respect to μ, if needed. We also abbreviate $\int_{[0,1]} \mu(dx)$ as $\int_0^1 dx$. Any probability density $p(x)$ is a L^1-function (i.e., $p \in L^1$) that satisfies $\int_0^1 p(x)dx = 1$ and $p(x) \geq 0$. We note that the finite-dimensional setup is regarded as a special case of this continuous setup by letting $p(x) := p_i d$ for $(i - 1)/d \leq x < i/d$.

We consider the way of "rearranging" of $p \in L^1$ in order to define $p^\downarrow \in L^1$. Let $m_p(y) := \mu[x : p(x) > y]$. It is straightforward to see that

$$p^\downarrow(x) := \sup\{y : m_p(y) > x\} \tag{3.11}$$

gives a proper generalization of $p_i^\downarrow$ to the continuous case. We note that $p^\downarrow \in L^1$ and $\int_0^1 dx p(x) = \int_0^1 dx p^\downarrow(x)$. The Lorenz curve in the continuous case is then given by the graph of $l_p(x) := \int_0^x p^\downarrow(x')dx'$, which can be a smooth curve, while the Lorenz curve was a polyline in the discrete case. Given the definition of majorization for general vectors in $\mathbb{R}^d$ (see the second last paragraph of Sect. 3.1), we define majorization in the continuous case:

Definition 3.6 (*Majorization for continuous variables*) Let $p, p' \in L^1$. We say that p majorizes p', written as $p' \prec p$, if

$$\int_0^y dx p'^\downarrow(x) \leq \int_0^y dx p^\downarrow(x), \quad \forall y \in [0, 1) \tag{3.12}$$

and $\int_0^1 dx p'(x) = \int_0^1 dx p(x)$ are satisfied.

We next define doubly stochastic maps of the continuous case as follows, given Proposition 3.1 of the discrete case.

Definition 3.7 (*Doubly stochastic maps*) A linear map $T : L^1 \to L^1$ is called doubly stochastic, if $Tp \prec p$ for all $p \in L^1$.

The following theorem guarantees that the above definition is indeed reasonable (see also Ref. [89] for mathematical details).

Theorem 3.12 $T : L^1 \to L^1$ *is a doubly stochastic map, if and only if*

$$(Tp)(x') = \frac{d}{dx'} \int_0^1 K(x', x) p(x) dx, \tag{3.13}$$

where $K(x', x)$ is monotonically increasing in x', and satisfies $K(0, x) = 0$, $K(1, x)$ $= 1$, and $\int_0^1 K(x', x) dx = x'$. Note that there are also additional technical conditions on $K(x', x)$: the essential supremum of V_x is finite where V_x is the total variation of $K(x', x)$ as a function of x', and $\int_0^1 K(x', x) p(x) dx$ is absolutely continuous with respect to x' for any $p \in L^1$.

Here, the reason why d/dx' is put outside of the integral in Eq. (3.13) is to avoid the explicit appearance of generalized functions (i.e., Schwartz's distributions) such as the delta function. However, it is intuitively easier to look at Eq. (3.13) by rewriting it as

$$(Tp)(x') = \int_0^1 K'(x', x) p(x) dx, \tag{3.14}$$

where $K'(x', x) := \partial K(x', x)/\partial x'$ is a generalized function and satisfies

$$K'(x', x) \geq 0, \quad \int_0^1 K'(x', x) dx' = \int_0^1 K'(x', x) dx = 1. \tag{3.15}$$

Equality (3.14) is regarded as a "matrix" representation of a continuous map, where $K'(x', x)$ is the integral kernel. For example, if T is the identity map, we have $K(x', x) = \theta(x' - x)$ (the step function) and $K'(x', x) = \delta(x' - x)$ (the delta function).

Under these definitions, the following theorem holds, which is the continuous version of (i) $\Leftrightarrow$ (iv) of Theorem 3.1.

Theorem 3.13 (Theorem 3 of [89]) *For p, $p' \in L^1$, $p' \prec p$ holds if and only if there exists a doubly stochastic map T such that $p' = Tp$.*

The "if" part is now trivial by definition of doubly stochastic maps. The essential idea of the proof of the "only if" part is the same as the proof of Theorem 3.1 (i) $\Rightarrow$ (iv) presented in Sect. 3.5, while we now need functional analysis for the rigorous proof; the hyperplane separation theorem should be replaced by the Hahn-Banach separation theorem.
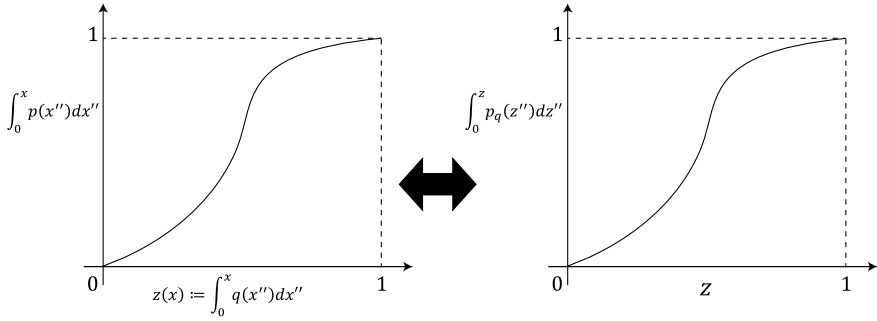
Fig. 3.4 Schematic of the variable transformation (3.16) and (3.17). We obtain the concave Lorenz curve by $l_{(p,q)}(z) := \int_0^z p_q^{\downarrow}(z'')dz''$ with Eq. (3.11)

We next consider d-majorization for continuous variables in line with Ref. [58] in a non-rigorous manner. As mentioned before, we can intuitively derive Theorem 3.3 for d-majorization from Theorem 3.1 for majorization, by going through majorization of continuous variables and using a variable transformation. Especially, we focus on (i) $\Leftrightarrow$ (iv) of Theorem 3.3 in the following.

Let $p, q \in L^1$ be probability densities. For simplicity, we assume that $q(x) > 0$ for all $x \in [0, 1]$, while the generalization to the case with $q(x) = 0$ for some intervals of x is straightforward. We consider the following variable transformation. Let $dz := q(x)dx$, or equivalently,

$$z(x) := \int_0^x q(x'')dx''. \tag{3.16}$$

We then define $x(z)$ as the inverse function of $z(x)$. Since $\int_0^1 q(x'')dx'' = 1$, we have $z \in [0, 1]$. With this new variable z, we define a probability density

$$p_q(z) := \frac{p(x(z))}{q(x(z))}, \tag{3.17}$$

which satisfies $\int_0^1 p_q(z)dz = \int_0^1 p(x)dx = 1$. Figure 3.4 illustrates this transformation.

The Lorenz curve of continuous d-majorization is obtained by rearranging $p(x)/q(x)$ in the decreasing order. That is, we can define $p_q^{\downarrow}(z)$ with Eq. (3.11), and obtain the Lorenz curve $l_{(p,q)}(z) := \int_0^z p_q^{\downarrow}(z'')dz''$. In other words, d-majorization $(p', q') \prec (p, q)$ is equivalent to majorization $p'_{q'} \prec p_q$, which, from Theorem 3.13, is further equivalent to that there exists a doubly stochastic map $\bar{T}$ such that $p'_{q'} = \bar{T} p_q$. This doubly stochastic map can be written as, with the notation of Eq. (3.14),

$$p'_{q'}(z') = \int_0^1 \bar{K}'(z', z) p_q(z)dz, \tag{3.18}$$

where $\int_0^1 \bar{K}'(z', z)dz = \int_0^1 \bar{K}'(z', z)dz' = 1$. We then restore the variable to the original one by defining

$$K'(x', x) := q'(x')\bar{K}'(z'(x'), z(x)), \tag{3.19}$$

where $z'(x')$ is defined by Eq. (3.16) with q'. By noting that $dz' = q'(x')dx'$ and $dz = q(x)dx$, we obtain $\int_0^1 K'(x', x)dx' = 1$, $p'(x') = \int_0^1 K'(x', x)p(x)dx$, and $q'(x') = \int_0^1 K'(x', x)q(x)dx$. We finally obtain the desired stochastic map T, which is defined by the form of Eq. (3.14) with the integral kernel $K'(x', x)$ constructed above.

To summarize, $(p', q') \prec (p, q)$ holds if and only if there exists a stochastic map T such that $p' = Tp$ and $q' = Tq$. As a special case of this argument, we obtain (i) $\Leftrightarrow$ (iv) of Theorem 3.3 of the finite-dimensional case.

The above (non-rigorous) argument for continuous variables suggests a rigorous proof for the finite-dimensional case. The benefit of the above argument is that we can perform the variable transformation of the horizontal axis of the Lorenz curve, which is in general impossible if the variable is discrete. However, if all of the components of q and q' are rational numbers, we can make the same argument for the finite-dimensional case in a rigorous manner without referring to the continuous variable case. We can then take a limit if there are irrational components. We will prove Theorem 3.3 in this line in Sect. 3.5.

3.5 Proofs

We here present the proofs of the two main theorems of this chapter (Theorems 3.1 and 3.3).

Proof of Theorem 3.1. We prove this theorem for general p, $p' \in \mathbb{R}^d$.

(iv) $\Rightarrow$ (iii) follows from Corollary 2.1.

(iii) $\Rightarrow$ (ii) is trivial because $|x|$ is convex.

(i) $\Leftrightarrow$ (ii) is Theorem II.1.3 of Ref. [7]. Suppose (i). Let $p_{k+1}'^\downarrow \le t \le p_k'^\downarrow$ with $p_0'^\downarrow := \infty$ and $p_{d+1}'^\downarrow := 0$. By noting that $\sum_{i=1}^d p_i' = \sum_{i=1}^d p_i$,

$$\sum_{i=1}^d |p_i' - t| = \sum_{i=1}^k (p_i'^\downarrow - t) - \sum_{i=k+1}^d (p_i'^\downarrow - t) \le \sum_{i=1}^k (p_i^\downarrow - t) - \sum_{i=k+1}^d (p_i^\downarrow - t) \le \sum_{i=1}^d |p_i - t|, \tag{3.20}$$

which implies (ii). Next, suppose (ii). By choosing t to be sufficiently large and sufficiently small, we obtain $\sum_{i=1}^d p_i' = \sum_{i=1}^d p_i =: C$. Let $t = p_k^\downarrow$. Then,

$$\sum_{i=1}^d |p_i - t| = \sum_{i=1}^k (p_i^\downarrow - t) - \sum_{i=k+1}^d (p_i^\downarrow - t) = 2\sum_{i=1}^k p_i^\downarrow + (d - 2k)t - C. \tag{3.21}$$

On the other hand,

$$\sum_{i=1}^{d} |p_i' - t| \geq \sum_{i=1}^{k} (p_i'^{\downarrow} - t) - \sum_{i=k+1}^{d} (p_i'^{\downarrow} - t) = 2 \sum_{i=1}^{k} p_i'^{\downarrow} + (d - 2k)t - C.$$

(3.22)

We thus obtain $\sum_{i=1}^{k} p_i'^{\downarrow} \leq \sum_{i=1}^{k} p_i^{\downarrow}$, which is nothing but (i).

(i) $\Rightarrow$ (iv) has several proofs. Here, we prove it by using the hyperplane separation theorem in line with Ref. [88]. Let $\mathscr{D}$ be the set of doubly stochastic matrices and define $\mathscr{D}(p) := \{Tp : T \in \mathscr{D}\}$. We consider a proof by contradiction; Suppose that $p' \prec p$ but $p' \notin \mathscr{D}(p)$. Since $\mathscr{D}(p)$ is a closed convex set, if $p' \notin \mathscr{D}(p)$, then p' and $\mathscr{D}(p)$ can be separated by a hyperplane. That is, there exists a vector $r \in \mathbb{R}^d$ such that for all $T \in \mathscr{D}$,

$$(Tp, r) < (p', r)$$

(3.23)

holds, where $(\cdot, \cdot)$ describes the ordinary inner product of $\mathbb{R}^d$. Since $\sum_{i=1}^{d} (Tp)_i = \sum_{i=1}^{d} p_i$, inequality (3.23) does not change by adding a constant vector $(c, c, \ldots, c)^{\mathrm{T}}$ to r. Therefore, without loss of generality we can assume that the components of r are all non-negative. Let P, R be permutation matrices such that $p = Pp^{\downarrow}, r = Rr^{\downarrow}$. Then we have $(p^{\downarrow}, r^{\downarrow}) = (RP^{\dagger}p, r)$. Let $T := RP^{\dagger}$, which is also a permutation matrix, and thus $T \in \mathscr{D}$. From inequality (3.23), we have

$$(p^{\downarrow}, r^{\downarrow}) = (Tp, r) < (p', r) = (R^{\dagger}p', r^{\downarrow}).$$

(3.24)

We note that $R^{\dagger}p' \prec p$. On the other hand, in general, if $v', v \in \mathbb{R}^d$ satisfy $v' \prec v$, we have for any $w \in \mathbb{R}^d$ with non-negative components

$$(v^{\downarrow}, w^{\downarrow}) \geq (v', w^{\downarrow}).$$

(3.25)

In fact, $0 \leq \sum_{k=1}^{d} (w_k^{\downarrow} - w_{k+1}^{\downarrow}) \sum_{i=1}^{k} (v_i^{\downarrow} - v_i') = \sum_{k=1}^{d} w_k^{\downarrow} (v_k^{\downarrow} - v_k')$ with $w_{d+1}^{\downarrow} := 0$. Then, inequalities (3.25) and (3.24) contradict. Therefore, $p' \in \mathscr{D}(p)$. $\square$

Proof of Theorem 3.3.

The following proof is parallel to that of Theorem 3.1, except that (ii) $\Rightarrow$ (iv) is only a nontrivial part. For simplicity, we here assume that $p, q, p', q' \in \mathscr{P}_d$, while the generalization to the more general case is straightforward.

(iv) $\Rightarrow$ (iii) follows from Lemma 2.1.

(iii) $\Rightarrow$ (ii) is trivial.

(i) $\Leftrightarrow$ (ii) is also proved in essentially the same manner as (i) $\Leftrightarrow$ (ii) of Theorem 3.1. In general, $[0, q_1^*], [q_1^*, q_1^* + q_2^*), \ldots, [q_1^* + \cdots + q_{d-1}^*, 1]$ and $[0, q_1'^*], [q_1'^*, q_1'^* + q_2'^*), \ldots, [q_1'^* + \cdots + q_{d-1}'^*, 1]$ give different partitions of the interval $[0, 1]$. Then, take $\tilde{q} = (\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_{d'})$ such that $[0, \tilde{q}_1^*], [\tilde{q}_1^*, \tilde{q}_1^* + \tilde{q}_2^*), \ldots, [\tilde{q}_1^* + \cdots + \tilde{q}_{d'-1}^*, 1]$ gives a refinement of both of the partitions. Correspondingly, we can take $\tilde{p}$ and $\tilde{p}'$ such that the Lorenz curves of (p, q) and (p', q') are respectively the same as those of $(\tilde{p}, \tilde{q})$ and $(\tilde{p}', \tilde{q})$, where the ways of rearranging $\tilde{q}$ to $\tilde{q}^*$ are the same for these

Lorenz curves. Therefore, it is sufficient to prove (i) ⇔ (ii) for $(\tilde{p}, \tilde{q})$ and $(\tilde{p}', \tilde{q}')$, but this can be proved in the same manner as (i) ⇔ (ii) of Theorem 3.1.

We now prove the nontrivial part, (ii) ⇒ (iv). We first suppose that all the components of q, q' are rational numbers, and let $q_i = m_i/M$, $q_i' = m_i'/M$ (m_i, m_i', $M \in \mathbb{N}$) with $\sum_i m_i = \sum_i m_i' = M$. We denote the m_i-dimensional uniform distribution by $u_{m_i} := (1/m_i, 1/m_i, \ldots, 1/m_i)$. We then define the following M-dimensional probability vectors:

$$p_q := (p_1 u_{m_1}, p_2 u_{m_2}, \ldots, p_d u_{m_d})^{\mathrm{T}}, \ \ p_{q'}' := (p_1' u_{m_1'}, p_2' u_{m_2'}, \ldots, p_d' u_{m_d'})^{\mathrm{T}}. \quad (3.26)$$

Then, by replacing t/M by t, inequality (3.5) is written as

$$\sum_{i=1}^{M} |p_{q',i}' - t| \leq \sum_{i=1}^{M} |p_{q,i} - t|. \quad (3.27)$$

Therefore, from Theorem 3.1, there exists an $M \times M$ doubly stochastic matrix $\bar{T}$ such that $p_{q'}' = \bar{T} p_q$. We partition $\bar{T}$ into several blocks, where each block is an $m_i' \times m_j$ matrix. We sum up all the elements of each block, and obtain a $d \times d$ matrix whose elements are those sums. We denote this $d \times d$ matrix by $\bar{T}'$. We then define a matrix T, which is obtained by dividing all the elements of the jth column of $\bar{T}'$ by m_j. Since the sum of the jth column of $\bar{T}'$ is m_j, T is a stochastic matrix. Furthermore, since $\bar{T}$ is a doubly stochastic matrix, the sum of all the elements of the ith row of $\bar{T}'$ is m_i'. Therefore, we obtain

$$\begin{pmatrix} m_1' \\ m_2' \\ \vdots \\ m_d' \end{pmatrix} = T \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_d \end{pmatrix}, \quad (3.28)$$

or equivalently, $q' = Tq$. Also, $p' = Tp$ holds by construction. Therefore, (ii) ⇒ (iv) is proved for the case that all the components of q, q' are rational.

Finally, we prove (ii) ⇒ (iv) for the case where there are irrational components in q, q' by approximating them. Let $\{q^{(n)}\}_{n \in \mathbb{N}}$, $\{q'^{(n)}\}_{n \in \mathbb{N}}$ be sequences of probability vectors that converges to q, q', respectively. We suppose that all the components of these sequences are rational and that $q_i^{*(n)} \leq q_i^*$, $q_i'^{*(n)} \geq q_i'^*$ for $i = 1, \ldots, d-1$ and $q_d^{*(n)} \geq q_d^*$, $q_d'^{*(n)} \leq q_d'^*$. Here, $q^{(n)}$, $q'^{(n)}$ should be taken sufficiently close to q, q' such that the orders of rearranging to define the Lorenz curves do not change; If there is i such that $p_i'^*/q_i'^* = p_{i+1}'^*/q_{i+1}'^*$, we need to approximate p' as well in order to keep the order of rearranging.

Then, the Lorenz curve of (p, q) is shifted to the left and that of (p', q') is shifted to the right, while fixing $(0, 0)$ and $(1, 1)$. Correspondingly, we can show that

$$\sum_{i=1}^{d} |p'_i - tq'^{(n)}_i| \le \sum_{i=1}^{d} |p'_i - tq'_i| \le \sum_{i=1}^{d} |p_i - tq_i| \le \sum_{i=1}^{d} |p_i - tq^{(n)}_i|. \quad (3.29)$$

Therefore, $(p', q') \prec (p, q)$ implies $(p', q'^{(n)}) \prec (p, q^{(n)})$. Therefore, there exists a stochastic matrix $T^{(n)}$ such that $p' = T^{(n)}p$, $q'^{(n)} = T^{(n)}q^{(n)}$. Since the set of stochastic matrices is sequential compact, $\{T^{(n)}\}_{n \in \mathbb{N}}$ has a convergent subsequence. Let T be its limit. Then, we have $p' = Tp$ and $q' = Tq$, which proves (ii) $\Rightarrow$ (iv). $\square$

As discussed in Ref. [41], the above proof of (ii) $\Rightarrow$ (iv) has a clear thermodynamic interpretation in the case of Gibbs-preserving maps. We here briefly discuss this point of view (without caring about rational vs irrational etc.). First, in addition to the present system (named S), we consider a heat bath B. Suppose that for any energy level E_i of S, there exists an energy level $E^{\mathrm{B}}_i := E - E_i$ of B, where E is a constant. Also suppose that the energy level E^{B}_i of B is d_i-fold degenerate with d_i being proportional to $e^{-\beta E_i}$. Then, as in the standard argument in statistical mechanics, the Gibbs state of S is regarded as the marginal distribution of the microcanonical distribution of SB with the total energy E. Thus, a Gibbs-preserving map of S corresponds to a doubly stochastic map of SB (that preserves the microcanonical distribution of SB), which is equivalent to the construction of $\bar{T}$ in the above proof.

Chapter 4
Classical Thermodynamics

We discuss the thermodynamic implications of classical information theory developed in Chaps. 2 and 3. (See also Chap. 8 for a more comprehensive formulation of the quantum case.)

In Sect. 4.1, we show that the second law of thermodynamics is derived as a straightforward consequence of the monotonicity (2.8) of the KL divergence. Here, work is a fluctuating quantity and satisfies the second law only at the level of the ensemble average, which is a perspective of stochastic thermodynamics.

In Sect. 4.2, we focus on the resource-theory approach to thermodynamics, where the work does not fluctuate even when dynamics of the system (and thus heat) are stochastic. This is referred to as single-shot (or one-shot) thermodynamics [41, 42]. In this setup, any entropic contribution from the work storage can be excluded, and thus work can be regarded as a purely "mechanical" quantity. Specifically, we show that the work bounds are given by the Rényi 0- and ∞-divergences in the single-shot situations on the basis of thermo-majorization.

4.1 Second Law and the KL Divergence

First, we formulate thermodynamic processes in terms of classical probability theory, and discuss the relationship between thermodynamics and the KL divergence in the spirit of stochastic thermodynamics. In particular, we will show that the second law of thermodynamics at the level of the ensemble average is an immediate consequence of the monotonicity (2.8) of the KL divergence.

Let E_i be the energy of level i of the system; the classical Hamiltonian H can be identified with $(E_1, E_2, \ldots, E_d) \in \mathbb{R}^d$. The Gibbs state $p^G \in \mathscr{P}_d$ is defined as $p_i^G := e^{-\beta E_i}/Z$, where $Z := \sum_i e^{-\beta E_i}$ is the partition function. The equilibrium free energy is defined as $F := -\beta^{-1} \ln Z$.

A stochastic map T is called a *Gibbs-preserving map*, if it does not change the Gibbs state, i.e.,

$$Tp^{\mathrm{G}} = p^{\mathrm{G}}. \tag{4.1}$$

We note that if the detailed balance condition

$$T_{ji}e^{-\beta E_i} = T_{ij}e^{-\beta E_j} \tag{4.2}$$

is satisfied, then T is a Gibbs-preserving map. (The converse is not true in general.) We emphasize that $\beta \geq 0$ is interpreted as the inverse temperature of the heat bath.

By substituting $q = p^{\mathrm{G}}$ to the monotonicity (2.8) of the KL divergence, we obtain for any Gibbs-preserving map T and for any $p \in \mathscr{P}_d$

$$S_1(p\|p^{\mathrm{G}}) \geq S_1(Tp\|p^{\mathrm{G}}). \tag{4.3}$$

This can be rewritten as $S_1(Tp) - S_1(p) \geq \beta \left(\sum_i E_i(Tp)_i - \sum_i E_i p_i \right)$, where the right-hand side represents the average energy change in the system by T. Because we do not consider the time-dependence of the Hamiltonian (and we do not explicitly consider the work storage) at present, no external work is performed on or extracted from the system, and thus the energy change equals the heat absorption from the heat bath. Thus, we can identify $Q := \sum_i E_i(Tp)_i - \sum_i E_i p_i$ to the average heat absorption. We note that the stochastic heat is given by $Q_{ji} := E_j - E_i$ during transition from i to j, whose ensemble average equals Q, i.e., $Q = \sum_{ij} T_{ji} p_i Q_{ji}$.

Let $\Delta S_1 := S_1(Tp) - S_1(p)$ be the change in the Shannon entropy of the system. Inequality (4.3) is now rewritten as

$$\Delta S_1 \geq \beta Q, \tag{4.4}$$

which is the second law of thermodynamics in the present setup [29, 30, 90]. This is in the same form as the conventional Clausius inequality of equilibrium thermodynamics [1], while inequality (4.4) includes an informational entropy (i.e., the Shannon entropy) instead of the Boltzmann entropy defined only for equilibrium states. We note that we will derive the same inequality for the quantum case in Sect. 5.2 (see also Sect. 8.4).

In stochastic thermodynamics,

$$\Sigma := S_1(p\|p^{\mathrm{G}}) - S_1(Tp\|p^{\mathrm{G}}) = \Delta S_1 - \beta Q \geq 0 \tag{4.5}$$

is often called the (average) *entropy production* [32, 34]. Here, $-\beta Q$ is interpreted as the entropy increase in the heat bath [91], which can be justified if the state of the bath is sufficiently close to the Gibbs state because of Eq. (2.7). We note that if T does not preserve the Gibbs state but preserves a nonequilibrium state q (i.e., a nonequilibrium steady state), $S_1(p\|q) - S_1(Tp\|q)$ is called the excess (or non-adiabatic) entropy production [92, 93] (see also Ref. [94, 95] for the quantum extension).

Inequality (4.4) is also a starting point of thermodynamics of information [37]. For example, inequality (4.4) can be regarded as a generalized Landauer's principle [96]. The original form of the Landauer's principle states that the erasure of one bit ($= \ln 2$) of information must be accompanied by at least $\beta^{-1} \ln 2$ of heat emission to the environment. Correspondingly, if $\Delta S_1 = -\ln 2$ on the left-hand side of (4.4), the heat emission is bounded as $-Q \geq \beta^{-1} \ln 2$. Inequality (4.4) also sets a fundamental upper bound of the work extraction by utilizing information, as is the case for the thought experiments of the Szilard engine and Maxwell's demon [36–38, 97, 98].

We note that a doubly stochastic map is regarded as a Gibbs-preserving map (with any Hamiltonian) at infinite temperature or a Gibbs-preserving map (at any temperature) of the trivial Hamiltonian whose energy levels are all degenerate; in these cases, the Gibbs state is given by the uniform distribution $p^G = u$, where the Shannon entropy does not decrease as represented by inequality (2.11). This may be related to the second law for adiabatic processes, stating that the entropy does not decrease if the system does not exchange the heat with the environment.

We next consider the second law in terms of the work and the free energy. We now suppose that the Hamiltonian of the system is time-dependent, and the work is performed on the system through the time-dependence of the Hamiltonian. Here, we do not explicitly take into account the work storage and suppose that the Hamiltonian of the system is driven by an external agent. This is contrastive to another formulation discussed in Sect. 4.2 and Chap. 8, where the entire system is supposed to be autonomous by including the work storage and the "clock" degrees of freedom as a part of the total system [41].

First, as a simplest situation, we suppose that the work is induced by a quench (i.e., an instantaneous change) of the Hamiltonian from H to $H' = (E'_1, \ldots, E'_d)$. Let $E := \sum_i E_i p_i$ and $E' := \sum_i E'_i p_i$ be the average energies immediately before and after the quench, respectively. Because the quench is supposed to be very quick and the heat exchange between the system and the bath is ignored during the quench, the average work performed on the system by the quench is given by

$$W := E' - E. \tag{4.6}$$

We emphasize that, in this setup, the work is fluctuating; If the system is in i immediately before the quench, the stochastic work is given by $w_i := E'_i - E_i$, which is a random variable. Then, the average work W is given by the ensemble average of w_i, that is, $W = \sum_i w_i p_i$.

We next consider more general thermodynamic processes. Suppose that the entire process consists of multiple quench steps and relaxation processes between the quenches (see Fig. 4.1 for a schematic), which we refer to as a "quench-and-relax" process. The Hamiltonian is fixed during the period between the quenches. We assume that each relaxation process is stochastically independent of other relaxation processes and is described by a single Gibbs-preserving map with respect to the Hamiltonian of that period. This assumption means that stochastic dynamics is essentially Markovian and no memory effect is present.
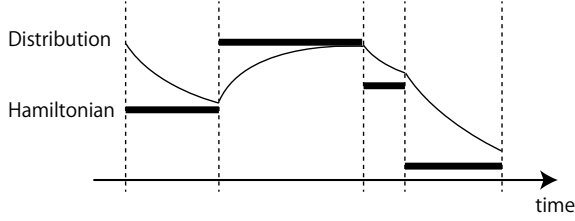
Fig. 4.1 A schematic of the quench-and-relax protocol. The bold horizontal lines symbolically represent the Hamiltonians, and the curves symbolically represent the time evolution of the probability distribution. The Hamiltonian is quenched at the times of the dashed vertical lines

We denote the initial and final distributions of the entire process by p, p'. Let H, H' be the initial and final Hamiltonians, p^G, $p^{G'}$ be the corresponding Gibbs states, and F, F' be the corresponding equilibrium free energies. We consider the average work W and the average heat Q during the above-mentioned entire process. Let $\Delta E := E' - E$ be the change in the average energy, where $E := \sum_i E_i p_i$ and $E' := \sum_i E_i' p_i'$. These energetic quantities satisfy the first law of thermodynamics, i.e., the energy conservation:

$$\Delta E = W + Q. \tag{4.7}$$

Again because a quench is instantaneous, the probability distribution does not change and the heat exchange is zero during it. Thus, the second law of the form (4.4) is unaffected, i.e., $\Delta S_1 \geq \beta Q$ holds with $\Delta S_1 := S_1(p') - S_1(p)$, even in the presence of the quenches. From the first law (4.7), we can rewrite the second law as

$$W \geq \Delta E - \beta^{-1} \Delta S_1. \tag{4.8}$$

Now we define the nonequilibrium (Rényi 1-) free energy by

$$F_1(p; H) := E - \beta^{-1} S_1(p) = \beta^{-1} S_1(p \| p^G) + F. \tag{4.9}$$

This quantity equals the equilibrium free energy, i,e, $F_1(p; H) = F$ holds, if and only if $p = p^G$. See also Sect. 8.1 for the α-free energy of the quantum case. By using the nonequilibrium free energy, the second law (4.8) is rewritten as

$$W \geq F_1(p', H') - F_1(p; H), \tag{4.10}$$

which is the work bound for the case where we allow work fluctuations.

We consider the following special cases of the above work bound. First, we consider the necessary work for the formation of a nonequilibrium state p' starting from the Gibbs state p^G by fixing the Hamiltonian $H = H'$. Then, inequality (4.10) reduces to $W \geq F_1(p'; H) - F = \beta^{-1} S_1(p' \| p^G)$.

Second, we consider the work extraction from a nonequilibrium state p again by fixing the Hamiltonian $H = H'$. Suppose that $p' = p^{G'}(= p^G)$. Then, inequality (4.10) reduces to $-W \leq F_1(p; H) - F = \beta^{-1}S_1(p \| p^G)$, where $-W$ is the extracted work.

Finally, for a transition between equilibrium states, i.e., if $p = p^G$ and $p' = p^{G'}$ (and in general $H \neq H'$), inequality (4.10) reduces to the conventional work bound:

$$W \geq \Delta F, \tag{4.11}$$

where $\Delta F := F' - F$ is the change in the equilibrium free energy.

We note that in the single-shot setup where the work does not fluctuate, the above inequalities are replaced by inequalities (4.15), (4.18), (4.22) in Sect. 4.2, respectively.

We can construct quasi-static processes by taking the limit of the foregoing quench-and-relax processes, which achieves the equality of (4.10). Let H and H' be the initial and final Hamiltonians, respectively. We divide the total process into N steps of quench-and-relax with the same time intervals. Let $H^{(n)} = (E_1^{(n)}, \ldots, E_d^{(n)})$ be the Hamiltonian after the nth quench ($n = 0, 1, 2, \ldots, N$), where $H^{(0)} = H$ and $H^{(N)} = H'$. We suppose that the energy change in each quench is of the order of $\varepsilon := 1/N$, i.e., $|E_i^{(n)} - E_i^{(n+1)}| = O(\varepsilon)$ for all i and n. Let $p^{G,(n)}$ be the Gibbs state of $H^{(n)}$. Suppose that the initial state is the Gibbs state $p^{G,(0)}$ the system reaches the Gibbs state $p^{G,(n)}$ before the $(n + 1)$th quench. This can be achieved by letting each relaxation process long enough compared to the relaxation time, where in the continuous-time picture described by the master equation, the relaxation time is determined by the rate matrix. The quasi-static limit is then given by $N \to \infty$ or equivalently $\varepsilon \to 0$.

In this setup, $\Delta S_1 - \beta Q$ in the nth relaxation process ($n = 1, 2, \ldots, N$) is given by $S_1(p^{G,(n)} \| p^{G,(n+1)})$. Since $D(p^{G,(n)}, p^{G,(n+1)}) = O(\varepsilon)$, we have $S_1(p^{G,(n)} \| p^{G,(n+1)}) = O(\varepsilon^2)$ from Eq. (2.7). By summing up these terms over $n = 1, 2, \ldots, N$, we obtain for the total process

$$\Delta S_1 - \beta Q = O(N\varepsilon^2) = O(\varepsilon). \tag{4.12}$$

Thus, the equality in the second law (4.10) is achieved, or equivalently the average entropy production becomes zero, in the quasi-static limit $\varepsilon \to 0$. From the law of large numbers, we can also show that the fluctuation of the work during the total process vanishes, which is consistent with the single-shot scenario discussed in Sect. 4.2.

The quasi-static process discussed above requires that the system is always close to the Gibbs state of the Hamiltonian at that moment. If the initial and final states are not the Gibbs states in general, the equality in the second law (4.10) can be achieved by the following protocol (see Fig. 4.2) [37, 38, 42]. Let p and p' be the initial and final states, and again let H and H' be the initial and final Hamiltonians. Suppose that p and p' have full rank.
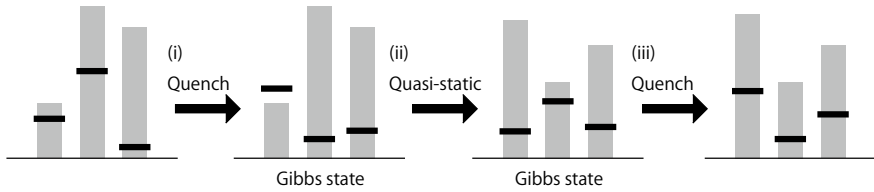
Fig. 4.2 A schematic of the protocol that achieves the equality of the second law (4.10). The bold horizontal lines represent the energy levels, and the shaded vertical bars represent the probability distributions

(i) At the initial time, we instantaneously quench the Hamiltonian from H to $\tilde{H} = (\tilde{E}_1, \ldots, \tilde{E}_d)$ such that p is the Gibbs state of $\tilde{H}$. Explicitly, we take $\tilde{E}_i := -\beta^{-1} \ln p_i$ up to constant.

(ii) We next change the Hamiltonian from $\tilde{H}$ to $\tilde{H}'$ quasi-statically. Here, $\tilde{H}' = (\tilde{E}'_1, \ldots, \tilde{E}'_d)$ is the Hamiltonian such that p' is the Gibbs state of $\tilde{H}'$, i.e., $\tilde{E}'_i := -\beta^{-1} \ln p'_i$ up to constant.

(iii) Finally, we instantaneously quench the Hamiltonian $\tilde{H}'$ to the final one H'.

In (i) and (iii), the state does not change and thus the energy change (work) equals the free energy change, leading to the equality of (4.10). In (ii), the equality is achieved because the process is quasi-static. Therefore, the equality of (4.10) is achieved in the entire process. We note that this protocol can be straightforwardly generalized to the quantum case [99].

The fact that the equality of (4.10) is achievable implies that the KL divergence provides a necessary and sufficient condition of state conversion in the present setup. That is, for given initial and final Hamiltonians H and H', the initial distribution p can be converted into the final distribution p' by some (time-dependent) protocol with average work cost W, if and only if inequality (4.10) is satisfied. This might seem different from the argument in Sect. 3.2, where we emphasized that the KL divergence does not provide a necessary and sufficient condition of state conversion. However, of course, this is not a contradiction. A crucial point here is that the work inevitably fluctuates in the quench steps (in contrast to the single-shot case). This would be reminiscent of the case of "modestly non-exact" catalytic majorization discussed in Sect. 3.3; To elaborate this point of view, we need to consider the work storage as a "catalyst," which we did not explicitly take into account in this section. See Sect. 8.4 for a related argument.

4.2   Single-Shot Work Bound

We next consider the single-shot situations where work does not fluctuate, and derive the work bounds as an important application of thermo-majorization. Especially, we focus on the necessary work for creating a nonequilibrium state and the extractable

work from a given nonequilibrium state [41, 42]. In this section, we graphically derive the work bounds for these setups by using the Lorenz curves.

A key idea of the present setup, in contrast to the setup of the previous section, is that we include a work storage (or a mechanical "weight") as a part of the entire system, and also include a "clock" [41] that can effectively simulate the time-dependence of the Hamiltonian of the system. By doing so, the Hamiltonian of the entire system becomes time-independent, and thus the Gibbs-preserving map on the entire system becomes relevant, even when the effective Hamiltonian of the system is time-dependent. The entire system now consists of the work storage W, the clock C, and the system S, and obeys a Gibbs-preserving map of the total Hamiltonian of SCW. See also Sect. 8.3 for more details of the clock and the work storage in the quantum case.

Let p^G be the Gibbs state of S with $p_i^G := e^{-\beta E_i}/Z$, where E_i is the energy of level i and Z is the partition function, as in Sect. 4.1. Let p be an arbitrary distribution of S. Suppose that W has only two energy levels, 0 and w. We write the Gibbs state of W as $r^G := (1/(1 + e^{-\beta w}), e^{-\beta w}/(1 + e^{-\beta w}))^T$ and write an arbitrary distribution of W as $r = (r_0, r_w)^T$ with $r_0 + r_w = 1$. We assume that the initial and final energies of W are given by 0 or w, and the energy change of W is always given by w or $-w$ (depending on the situation that we are considering) with unit probability, which is the characteristic of the single-shot scenario.

State formation. First, we consider the minimum work that is needed for creating nonequilibrium distribution p of S from the Gibbs state p^G. We here suppose that the initial and final Hamiltonians of S are the same, and do not consider C explicitly. Let $w > 0$.

The initial distribution of S is p^G and that of W is $r^{up} := (0, 1)^T$, and the initial distribution of SW is given by $p^G \otimes r^{up}$. The Lorenz curve of it (more precisely, the (relative) Lorenz curve of $(p^G \otimes r^{up}, p^G \otimes r^G)$) consists of a straight line and a horizontal line (see Fig. 4.3). Then, the final distribution of S is p and that of W is $r^{down} := (1, 0)^T$. The final distribution of SW is given by $p \otimes r^{down}$, where $w > 0$ is the work performed on S by W.

Figure 4.3 shows the Lorenz curves of the initial and the final distributions of SW, along with the relevant ∞-divergences. Then we see that the state conversion is possible, if and only if

$$e^{-S_\infty(p^G \otimes r^{up} \| p^G \otimes r^G)} \leq e^{-S_\infty(p \otimes r^{down} \| p^G \otimes r^G)}, \tag{4.13}$$

which reduces to

$$\frac{e^{-\beta w}}{1 + e^{-\beta w}} \leq \frac{1}{1 + e^{-\beta w}} e^{-S_\infty(p \| p^G)}. \tag{4.14}$$

Thus, we obtain the necessary and sufficient condition for the state conversion as

$$w \geq \beta^{-1} S_\infty(p \| p^G), \tag{4.15}$$

which gives the lower bound of the necessary work [41].

Fig. 4.3 Lorenz curves of
the state conversion from the
Gibbs state p^G of S to a
nonequilibrium state p with
work cost $w > 0$. The state
conversion is possible, if and
only if $p^G \otimes r^{up}$
thermo-majorizes $p \otimes r^{down}$.
We note that the shape of the
Lorenz curve of $p \otimes r^{down}$ is
chosen to be just a simple,
not generic one
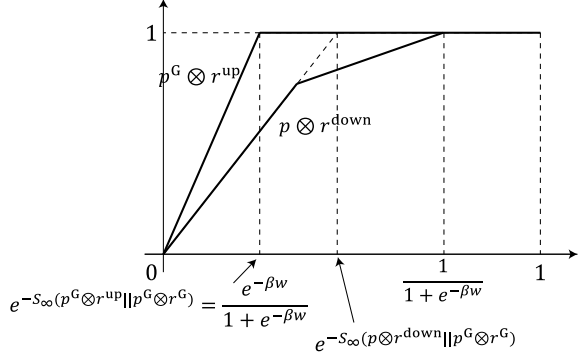
Work extraction. We next consider the maximum work that is extractable from an
initial nonequilibrium distribution p of S. We here again suppose that the initial and
final Hamiltonians of S are the same, and do not consider C explicitly. Let $w < 0$.

The initial state of W is $r^{down} := (0, 1)^T$ and that of SW is $p \otimes r^{down}$, where "up"
and "down" are exchanged from the previous setup. The final distribution of S is p^G
and that of W is $r^{up} := (1, 0)^T$, and the final distribution of SW is $p^G \otimes r^{up}$. In this
setup, $-w > 0$ is the work that is extracted from S and finally stored in W.

Figure 4.4 shows the Lorenz curves of the initial and the final distributions of SW,
along with the relevant 0-divergences. From this, we see that the state conversion is
possible, if and only if

$$e^{-S_0(p \otimes r^{down} \| p^G \otimes r^G)} \leq e^{-S_0(p^G \otimes r^{up} \| p^G \otimes r^G)}, \tag{4.16}$$

which reduces to

$$\frac{e^{-\beta w}}{1 + e^{-\beta w}} e^{-S_0(p \| p^G)} \leq \frac{1}{1 + e^{-\beta w}}. \tag{4.17}$$

Therefore, we obtain the necessary and sufficient condition for the state conversion

$$-w \leq \beta^{-1} S_0(p \| p^G), \tag{4.18}$$

which gives the upper bound of the extracted work $-w$ [41, 42].

The equality in (4.18) can be achieved by the following protocol proposed in
Ref. [42], which is illustrated in Fig. 4.5. Here, we do not take W and C into account
explicitly.

 (i) We instantaneously push the energy levels of i with $p_i = 0$ up to $+\infty$, by keeping
     the other energy levels fixed. During this quench, no work is performed on S.
(ii) After waiting for the relaxation of S, we quasi-statically restore the pushed-up
     energy levels to the original values.

The work extraction during the quasi-static process equals the change in the equilib-
rium free energies, which is given by
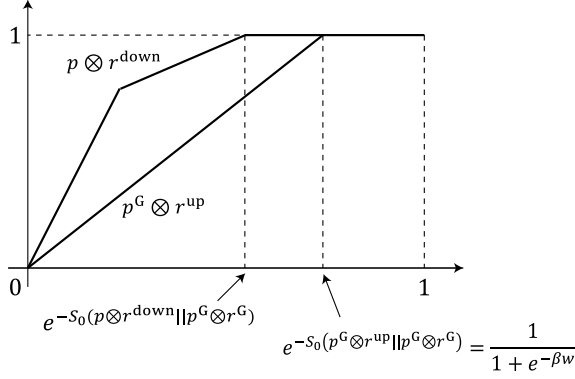
Fig. 4.4 Lorenz curves of the state conversion from an arbitrary initial state p of S to its Gibbs state p^G with work extraction $-w > 0$. The state conversion is possible, if and only if $p \otimes r^{\text{down}}$ thermo-majorizes $p^G \otimes r^{\text{up}}$. We note that the shape of the Lorenz curve of $p \otimes r^{\text{down}}$ is chosen to be just a simple, not generic one
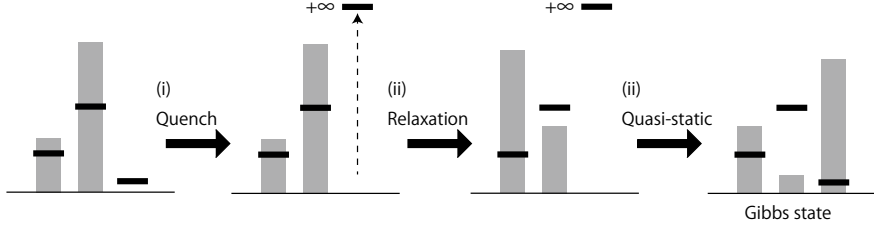
Fig. 4.5 A schematic of the protocol that achieves the equality in (4.18). The bold horizontal lines represent the energy levels, and the shaded vertical bars represent the probability distributions

$$- w = -\beta^{-1} \ln \frac{Z_0}{Z} = -\beta^{-1} \ln \left(\sum_{i:p_i>0} p_i^G \right) = \beta^{-1} S_0(p \| p^G), \qquad (4.19)$$

where $Z_0 := \sum_{i:p_i>0} e^{-\beta E_i}$. The work does not fluctuate in the entire process, and thus it is a single-shot protocol.

Equilibrium transition. Finally, we consider transitions between Gibbs states of S, where the Hamiltonian of S is changed from the initial one to the final one by using the clock C. We will see that this setup reproduces the conventional form of the second law (4.11), which is characterized by the equilibrium free energies.

We suppose that Hamiltonian of S is given by the energy levels E_i if the clock indicates "0", and by E_i' if "1". The initial and the final distributions of S, written as p and p', are given by the Gibbs states of the initial and final Hamiltonians: $p_i := e^{-\beta E_i}/Z$ and $p_i' := e^{-\beta E_i'}/Z'$ with Z, Z' being the partition functions (here we dropped the superscript "G" for simplicity). The corresponding free energies are given by $F := -\beta^{-1} \ln Z$ and $F' := -\beta^{-1} \ln Z'$.

Fig. 4.6 Lorenz curves for a transition between Gibbs states of S. The state conversion is possible, if and only if $p \otimes c \otimes r^{\text{down}}$ thermo-majorizes $p' \otimes c' \otimes r^{\text{up}}$
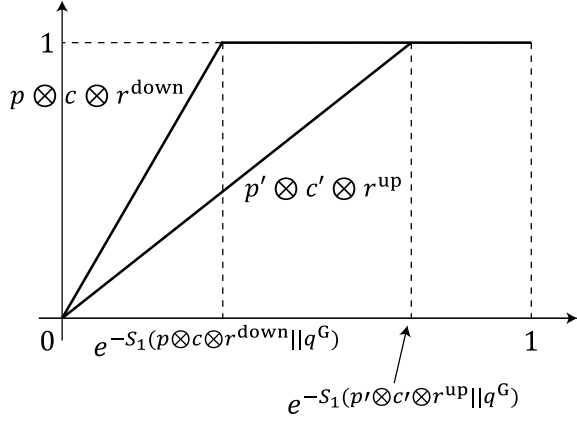
For simplicity, we suppose that C only has these two states 0 and 1. Let $(c_0, c_1)^{\text{T}}$ be the probability distribution of C in general. Specifically, we suppose that the initial distribution of C is $c = (1, 0)^{\text{T}}$ and the final one is $c' = (0, 1)^{\text{T}}$. By noting that C is coupled to S such that C induces the change of the Hamiltonian of S, the total Gibbs state of SC should be given by

$$p_{\text{SC}}^{\text{G}} = \frac{Z}{Z + Z'} p \otimes c + \frac{Z'}{Z + Z'} p' \otimes c'. \tag{4.20}$$

See Eqs. (8.16) and (8.19) of Sect. 8.5 for a more explicit argument.

Suppose that the initial and final states of W are respectively given by $r^{\text{down}} :=$ $(0, 1)^{\text{T}}$ and $r^{\text{up}} = (1, 0)^{\text{T}}$. Here, we supposed that $-w > 0$ (i.e., the case of work extraction), while this is not actually necessary for the following argument.

Then, the Gibbs state of SCW is given by $q^{\text{G}} := p_{\text{SC}}^{\text{G}} \otimes r^{\text{G}}$. The entire system SCW obeys a Gibbs-preserving map with respect to q^{G}. In this setup, each of the initial and the final Lorenz curves consists of a straight line and a horizontal line (see Fig. 4.6). Thus, $S_\alpha(p \otimes c \otimes r^{\text{down}} \| q^{\text{G}})$ collapses to a single value for all $0 \leq \alpha \leq \infty$. In particular, the KL divergence ($\alpha = 1$) coincides with all the other Rényi divergences. Thus, from the Lorenz curves shown in Fig. 4.6, the state conversion is possible if and only if

$$e^{-S_1(p \otimes c \otimes r^{\text{down}} \| q^{\text{G}})} \leq e^{-S_1(p' \otimes c' \otimes r^{\text{up}} \| q^{\text{G}})}. \tag{4.21}$$

We also have $e^{-S_1(p \otimes c \otimes r^{\text{down}} \| q^{\text{G}})} = Z/(Z + Z') \cdot e^{-\beta w}/(1 + e^{-\beta w})$, $e^{-S_1(p' \otimes c' \otimes r^{\text{up}} \| q^{\text{G}})} = Z'/(Z + Z') \cdot 1/(1 + e^{-\beta w})$. Thus, we obtain the necessary and sufficient condition for the state conversion as

$$-w \leq -(F' - F), \tag{4.22}$$

which is nothing but the second law (4.11) in the conventional form. This is also consistent with the observation in Sect. 4.1 that the work fluctuation vanishes in the quasi-static limit, which implies that equilibrium transitions can be realized by single-shot protocols.

Chapter 5
Quantum Entropy and Divergence

In this chapter, we consider the basic concepts of quantum information theory. In Sect. 5.1, we briefly summarize the basic concepts of quantum states and quantum dynamics. We then consider the von Neumann entropy and the quantum KL divergence (quantum relative entropy) in Sect. 5.2, and the quantum Rényi 0- and ∞-divergences (min and max divergences) in Sect. 5.3. We postpone the proof of the monotonicity of these divergences to Appendix A.

5.1 Quantum State and Dynamics

As a preliminary, we overview the basic concepts of quantum states and quantum dynamics very briefly. See also, for example, Ref. [4] for details.

A quantum system is described by a Hilbert space $\mathscr{H}$. Throughout this book, we assume that the Hilbert space is finite-dimensional, unless stated otherwise. We write the identity on any Hilbert space by $\hat{I}$. We denote the set of linear operators acting on $\mathscr{H}$ by $\mathscr{L}(\mathscr{H})$.

Let d be the dimension of the Hilbert space $\mathscr{H}$. A pure state of the system is represented by a normalized vector $|\varphi\rangle \in \mathscr{H}$ with $\langle\varphi|\varphi\rangle = 1$. In general, a *quantum state* is represented by a density operator $\hat{\rho}$ acting on $\mathscr{H}$, which is normalized, $\text{tr}[\hat{\rho}] = 1$, and is positive, $\hat{\rho} \geq 0$. The trace is defined as $\text{tr}[\hat{\rho}] := \sum_{i=1}^{d} \langle\varphi_i|\hat{\rho}|\varphi_i\rangle$ with $\{|\varphi_i\rangle\}_{i=1}^{d}$ being an orthonormal basis of $\mathscr{H}$.

Here, an operator $\hat{\rho} \in \mathscr{L}(\mathscr{H})$ is *positive* (or more precisely, positive semidefinite), if $\langle\varphi|\hat{\rho}|\varphi\rangle \geq 0$ for all $|\varphi\rangle \in \mathscr{H}$. In addition, an operator $\hat{\rho}$ is *positive definite*, if $\langle\varphi|\hat{\rho}|\varphi\rangle > 0$ for all nonzero $|\varphi\rangle \in \mathscr{H}$. We denote $\hat{\rho} \geq 0$ and $\hat{\rho} > 0$ if $\hat{\rho}$ is positive and positive definite, respectively. Also, $\hat{A} \geq \hat{B}$ and $\hat{A} > \hat{B}$ represent $\hat{A} - \hat{B} \geq 0$ and $\hat{A} - \hat{B} > 0$, respectively.

We denote the set of (normalized) quantum states on $\mathcal{H}$ by $\mathcal{S}(\mathcal{H})$. This set is convex; a state is pure if and only if it is an extreme point of $\mathcal{S}(\mathcal{H})$. (In this book, we do not consider a superselection rule that makes the definition of pure states more subtle.) We note that if an operator $\hat{\rho}$ is positive and satisfies $\mathrm{tr}[\hat{\rho}] \leq 1$, it is referred to as a *subnormalized state*. An operator that is only positive is referred to an *unnormalized state*. In this book, states are always normalized unless otherwise stated.

Since any positive operator is Hermitian, a quantum state $\hat{\rho}$ has the spectrum decomposition of the form

$$\hat{\rho} = \sum_{i=1}^{d} p_i |\varphi_i\rangle\langle\varphi_i|, \tag{5.1}$$

where $|\varphi_i\rangle\langle\varphi_i|$ is a projection. We note that $p_i \geq 0$ and $\sum_{i=1}^{d} p_i = 1$. Here, $p := (p_1, p_2, \ldots, p_d)^{\mathrm{T}}$ is regarded as a classical probability distribution, which is referred to as the *diagonal distribution* of $\hat{\rho}$. If a state is pure, the density operator is given by $\hat{\rho} = |\varphi\rangle\langle\varphi|$.

The *support* of $\hat{\rho}$ is the subspace spanned by the eigenvectors with nonzero eigenvalues of $\hat{\rho}$, which is denoted as $\mathrm{supp}[\hat{\rho}] \subset \mathcal{H}$. The rank of $\hat{\rho}$, denoted as $\mathrm{rank}[\hat{\rho}]$, is the dimension of $\mathrm{supp}[\hat{\rho}]$. The support of any positive-definite operator equals the entire space $\mathcal{H}$ (i.e., it has full rank).

We next consider a composite system AB that consists of subsystems A and B. Let $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ be the Hilbert spaces of subsystems A and B with dimensions d and d', respectively. Then, the Hilbert space of AB is given by their tensor product $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ with dimension dd'.

If the partial states of A and B are independent and written as $\hat{\rho}_{\mathrm{A}} \in \mathcal{S}(\mathcal{H}_{\mathrm{A}})$ and $\hat{\rho}_{\mathrm{B}} \in \mathcal{S}(\mathcal{H}_{\mathrm{B}})$, the corresponding state of AB is given by $\hat{\rho}_{\mathrm{A}} \otimes \hat{\rho}_{\mathrm{B}} \in \mathcal{S}(\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}})$, which is called a product state. If the partial states are both pure and written as $|\varphi_{\mathrm{A}}\rangle \in \mathcal{H}_{\mathrm{A}}, |\varphi_{\mathrm{B}}\rangle \in \mathcal{H}_{\mathrm{B}}$, the state vector of AB is given by $|\varphi_{\mathrm{A}}\rangle \otimes |\varphi_{\mathrm{B}}\rangle \in \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$. In the following, we abbreviate $|\varphi_{\mathrm{A}}\rangle \otimes |\varphi_{\mathrm{B}}\rangle$ as $|\varphi_{\mathrm{A}}\rangle|\varphi_{\mathrm{B}}\rangle$. In general, a state of AB is a density operator $\hat{\rho} \in \mathcal{S}(\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}})$, whose partial states are given by the partial trace; for example, the partial state of A is given by

$$\hat{\rho}_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}}[\hat{\rho}] := \sum_{i=1}^{d'} \langle\varphi_{\mathrm{B},i}|\hat{\rho}|\varphi_{\mathrm{B},i}\rangle \in \mathcal{S}(\mathcal{H}_{\mathrm{A}}), \tag{5.2}$$

where $\{|\varphi_{\mathrm{B},i}\rangle\}_{i=1}^{d'}$ is an orthonormal basis of $\mathcal{H}_{\mathrm{B}}$. Here, $\langle\varphi_{\mathrm{B},i}|\hat{\rho}|\varphi_{\mathrm{B},i}\rangle$ should be understood as an operator acting on $\mathcal{H}_{\mathrm{A}}$; It is given by $\langle\varphi_{\mathrm{B},i}|\hat{\rho}_{\mathrm{B}}|\varphi_{\mathrm{B},i}\rangle\hat{\rho}_{\mathrm{A}}$ for $\hat{\rho} = \hat{\rho}_{\mathrm{A}} \otimes \hat{\rho}_{\mathrm{B}}$, and by a linear combination of such operators for the general case.

We note that a pure state of AB is called *entangled*, if it is not in the form of a product state $|\varphi_{\mathrm{A}}\rangle|\varphi_{\mathrm{B}}\rangle$. A mixed state $\hat{\rho}_{\mathrm{AB}}$ is called a *product* state if it is written as $\hat{\rho}_{\mathrm{AB}} = \hat{\rho}_{\mathrm{A}} \otimes \hat{\rho}_{\mathrm{B}}$, and it is called a *separable* state if it can be written as $\hat{\rho}_{\mathrm{AB}} = \sum_k r_k \hat{\rho}_{\mathrm{A}}^{(k)} \otimes \hat{\rho}_{\mathrm{B}}^{(k)}$ with some classical probability distribution $r = (r_1, r_2, \ldots, r_K)^{\mathrm{T}}$ with $K < \infty$. Any separable state has only classical correlations between A and B. If $\hat{\rho}_{\mathrm{AB}}$ is not separable, it is called entangled.

We next consider quantum dynamics, which is represented by a linear map $\mathscr{E} : \mathscr{L}(\mathscr{H}) \to \mathscr{L}(\mathscr{H}')$. Such a linear map acting on an operator space $\mathscr{L}(\mathscr{H})$ is sometimes called a superoperator. In general, we allow different Hilbert spaces $\mathscr{H}$, $\mathscr{H}'$ for the input and output systems.

The simplest case is unitary dynamics that is given by, with a unitary operator $\hat{U} \in \mathscr{L}(\mathscr{H})$,

$$\mathscr{E}(\hat{\rho}) = \hat{U} \hat{\rho} \hat{U}^{\dagger}, \tag{5.3}$$

which maps a pure state to a pure state, i.e., $|\varphi\rangle \mapsto \hat{U}|\varphi\rangle$. In general, any physically-realizable dynamics must be completely-positive (CP) as defined as follows:

Definition 5.1 (*Completely positive maps*) Let $\mathscr{E} : \mathscr{L}(\mathscr{H}) \to \mathscr{L}(\mathscr{H}')$ be a linear map.

- $\mathscr{E}$ is positive, if any positive operator is mapped to a positive operator, that is, $\hat{X} \geq 0$ implies $\mathscr{E}(\hat{X}) \geq 0$.
- $\mathscr{E}$ is n-positive if $\mathscr{E} \otimes \mathscr{I}_n$ is positive, where $\mathscr{I}_n$ is the identity superoperator on $\mathscr{L}(\mathbb{C}^n)$.
- $\mathscr{E}$ is completely positive (CP) if it is n-positive for all $n \in \mathbb{N}$.

An example of a map that is positem but not CP is the transpose of matrices with a given basis. For example, if we transpose the density matrix of a partial state of an entangled state, the total density operator can have negative eigenvalues. Based on this property, the transpose can be used for quantifying entanglement [100, 101].

We also define the concept of trace-preserving (TP), which represents the conservation of probability; If quantum dynamics is TP, the output state is normalized for any normalized input state.

Definition 5.2 (*Trace-preserving*) A linear map $\mathscr{E}$ is called trace-preserving (TP), if it satisfies

$$\mathrm{tr}[\mathscr{E}(\hat{X})] = \mathrm{tr}[\hat{X}] \tag{5.4}$$

for all $\hat{X}$. If $\mathrm{tr}[\mathscr{E}(\hat{X})] \leq \mathrm{tr}[\hat{X}]$ holds for all $\hat{X} \geq 0$, $\mathscr{E}$ is called trace-nonincreasing.

To summarize, any physically realizable map that conserves probability must be completely-positive and trace-preserving (CPTP). It is known that a superoperator $\mathscr{E}$ is CP if and only if it is written as

$$\mathscr{E}(\hat{\rho}) = \sum_k \hat{M}_k \hat{\rho} \hat{M}_k^{\dagger}, \tag{5.5}$$

which is called the Kraus representation ($\hat{M}_k$'s are called the Kraus operators). We omit the proof of this (see, e.g., Theorem 8.1 of [4], Theorem 1 of [102]). If $\mathscr{E}$ is CPTP, the Kraus operators satisfy $\sum_k \hat{M}_k^{\dagger} \hat{M}_k = \hat{I}$, which guarantees the conservation of the trace. If $\mathscr{E}$ is CP and trace-nonincreasing, we have $\sum_k \hat{M}_k^{\dagger} \hat{M}_k \leq \hat{I}$.

We further remark that any CPTP map $\mathscr{E}$ on $\mathscr{H}$ can be represented as a unitary map on an extended space (see Ref. [4] for details). If $\mathscr{E}$ is CPTP, there exists a quantum

state $\hat{\sigma} \in \mathscr{S}(\mathscr{H}_A)$ of an auxiliary system A and a unitary operator on $\mathscr{H} \otimes \mathscr{H}_A$ such that

$$\mathscr{E}(\hat{\rho}) = \mathrm{tr}_{A'}[\hat{U}\hat{\rho} \otimes \hat{\sigma}\hat{U}^{\dagger}], \tag{5.6}$$

where $\mathrm{tr}_{A'}$ is the partial trace over an output auxiliary system A' such that $\mathscr{H} \otimes \mathscr{H}_A \simeq \mathscr{H}' \otimes \mathscr{H}_{A'}$. Conversely, any map of the form (5.6) is CPTP. This implies that any quantum dynamics can be regarded as unitary dynamics of a larger system including the environment.

We here remark some examples of CP maps. First, unitary dynamics (5.3) is CPTP with a single Kraus operator $\hat{U}$. Second, the partial trace is CPTP: for $\mathscr{H} \simeq \mathscr{H}' \otimes \mathscr{H}_{A'}$, $\mathscr{E}(\hat{\rho}) := \mathrm{tr}_{A'}[\hat{\rho}]$ is CPTP. In particular, the trace is CPTP, where the output space is the one-dimensional Hilbert space $\mathbb{C}$.

A quantum measurement with a particular (post-selected) outcome is an example of CP and trace-nonincreasing maps. Let k be a measurement outcome. For the measured state $\hat{\rho}$, the post-measurement state with outcome k is given by $\mathscr{E}_k(\hat{\rho})$, where the probability of outcome k is given by $p_k = \mathrm{tr}[\mathscr{E}_k(\hat{\rho})]$. $\mathscr{E}_k$ must be CP, while it does not preserve the trace if $p_k < 1$. The normalized post-measurement state is given by $\mathscr{E}_k(\hat{\rho})/p_k$. If we average the post-measurement states over all outcomes, the averaged state is given by $\mathscr{E}(\hat{\rho}) := \sum_k \mathscr{E}_k(\hat{\rho})$. Now $\mathscr{E}$ is CPTP, because of $\sum_k p_k = 1$. We note that the map $\hat{\rho} \mapsto p_k$ is also CP where the output Hilbert space is one dimensional. Correspondingly, we can embed $\{p_k\}$ to the diagonal elements of a density operator with some fixed basis, as $\hat{p} := \sum_k p_k |k\rangle\langle k|$. Then, $\hat{\rho} \mapsto \hat{p}$ is a CPTP map.

Meanwhile, a map $\mathscr{E} : \mathscr{L}(\mathscr{H}) \to \mathscr{L}(\mathscr{H})$ is called *unital*, if it preserves the identity of $\mathscr{H}$, i.e., $\mathscr{E}(\hat{I}) = \hat{I}$. Unital should not be confused with unitary, but any unitary map is unital. The concept of CPTP unital is a quantum analogue of doubly stochastic, as seen in Sect. 6.1.

We next consider an inner product of the operator space, called the Hilbert-Schmidt inner product defined as

$$\langle \hat{Y}, \hat{X} \rangle_{\mathrm{HS}} := \mathrm{tr}[\hat{Y}^{\dagger}\hat{X}]. \tag{5.7}$$

We can define the Hermitian conjugate of $\mathscr{E} : \mathscr{L}(\mathscr{H}) \to \mathscr{L}(\mathscr{H}')$ with respect to the Hilbert-Schmidt inner product, which is a superoperator $\mathscr{E}^{\dagger} : \mathscr{L}(\mathscr{H}') \to \mathscr{L}(\mathscr{H})$ satisfying

$$\langle \hat{Y}, \mathscr{E}(\hat{X}) \rangle_{\mathrm{HS}} = \langle \mathscr{E}^{\dagger}(\hat{Y}), \hat{X} \rangle_{\mathrm{HS}} \tag{5.8}$$

for all $\hat{X}, \hat{Y}$. Clearly, $\mathscr{E}^{\dagger\dagger} = \mathscr{E}$.

$\mathscr{E}$ is TP if and only if $\mathscr{E}^{\dagger}$ is unital. In fact, by taking $\hat{Y} = \hat{I}$ in the above relation, we have $\mathrm{tr}[\mathscr{E}(\hat{X})] = \langle \hat{I}, \mathscr{E}(\hat{X}) \rangle_{\mathrm{HS}} = \langle \mathscr{E}^{\dagger}(\hat{I}), \hat{X} \rangle_{\mathrm{HS}} = \mathrm{tr}[\mathscr{E}^{\dagger}(\hat{I})\hat{X}]$. Thus, $\mathrm{tr}[\mathscr{E}(\hat{X})] = \mathrm{tr}[\hat{X}]$ holds for all $\hat{X}$, if and only if $\mathscr{E}^{\dagger}(\hat{I}) = \hat{I}$.

We note that $\mathscr{E}$ is positive if and only if $\mathscr{E}^{\dagger}$ is positive. To see this, let $\hat{Y}$ be positive and take $\hat{X} = |\varphi\rangle\langle\varphi|$. Suppose that $\mathscr{E}$ is positive. Then $\langle\varphi|\mathscr{E}^{\dagger}(\hat{Y})|\varphi\rangle = \mathrm{tr}[\hat{Y}\mathscr{E}(\hat{X})] = \mathrm{tr}[\mathscr{E}(\hat{X})^{1/2}\hat{Y}\mathscr{E}(\hat{X})^{1/2}] \geq 0$. Moreover, $\mathscr{E}$ is n-positive (resp. CP) if and only if $\mathscr{E}^{\dagger}$ is

n-positive (resp. CP). In fact, $\mathscr{E} \otimes \mathscr{I}_n$ is positive if and only if $\mathscr{E}^\dagger \otimes \mathscr{I}_n$ is positive, because $\mathscr{I}_n^\dagger = \mathscr{I}_n$.

We finally remark on norms of operators. Let $\hat{X}$ be an arbitrary operator acting on $\mathscr{H}$. First, the trace norm is defined as

$$\|\hat{X}\|_1 := \mathrm{tr}[|\hat{X}|], \tag{5.9}$$

where $|\hat{X}| := \sqrt{\hat{X}^\dagger \hat{X}}$. This is indeed a norm in the mathematics sense, because it satisfies the triangle inequality $\|\hat{X} + \hat{X}'\|_1 \leq \|\hat{X}\|_1 + \|\hat{X}'\|_1$ and $\|\hat{X}\|_1 = 0$ holds if and only if $\hat{X} = 0$. The trace distance is then defined as

$$D(\hat{\rho}, \hat{\sigma}) := \frac{1}{2}\|\hat{\rho} - \hat{\sigma}\|_1, \tag{5.10}$$

which is commonly used for measuring a distance between quantum states $\hat{\rho}, \hat{\sigma}$, and satisfies the monotonicity under CPTP map $\mathscr{E}$ (Theorem 9.2 of [4]):

$$D(\hat{\rho}, \hat{\sigma}) \geq D(\mathscr{E}(\hat{\rho}), \mathscr{E}(\hat{\sigma})). \tag{5.11}$$

Another important norm is the operator norm defined by

$$\|\hat{X}\|_\infty := \max_{\||\varphi\rangle\|=1} \sqrt{\langle\varphi|\hat{X}^\dagger \hat{X}|\varphi\rangle}, \tag{5.12}$$

which is nothing but the largest singular value of $\hat{X}$.

5.2   von Neumann Entropy and the Quantum KL Divergence

The quantum analogue of the Shannon entropy is the *von Neumann entropy*, which is defined for a density operator $\hat{\rho}$ of dimension d by

$$S_1(\hat{\rho}) := -\mathrm{tr}[\hat{\rho} \ln \hat{\rho}]. \tag{5.13}$$

Let p be the diagonal distribution of $\hat{\rho}$. Then $S_1(p) = S_1(\hat{\rho})$ holds, implying that the von Neumann entropy characterizes "classical randomness" of quantum states. We thus have $0 \leq S_1(\hat{\rho}) \leq \ln d$. It is obvious that the von Neumann entropy is invariant under unitary transformation: $S_1(\hat{U} \hat{\rho} \hat{U}^\dagger) = S_1(\hat{\rho})$ holds for any unitary $\hat{U}$.

The KL divergence is also generalized to the quantum case, which is referred to as the quantum KL divergence or the quantum relative entropy [4]. It is defined for two quantum states $\hat{\rho}$ and $\hat{\sigma}$ of dimension d as

$$S_1(\hat{\rho}\|\hat{\sigma}) := \mathrm{tr}[\hat{\rho}\ln\hat{\rho} - \hat{\rho}\ln\hat{\sigma}]. \tag{5.14}$$

If the support of $\hat{\rho}$ is not included in that of $\hat{\sigma}$, we define $S_1(\hat{\rho}\|\hat{\sigma}) = +\infty$. As in the classical case, however, we assume that the support of $\hat{\rho}$ is included in that of $\hat{\sigma}$ throughout this book, whenever we consider quantum divergence-like quantities (including those in Appendix A). We note that

$$S_1(\hat{\rho}) = \ln d - S_1(\hat{\rho}\|\hat{I}/d), \tag{5.15}$$

where $\hat{I}/d$ is the maximally mixed state (i.e., the uniform distribution). If $\hat{\rho}$ and $\hat{\sigma}$ are simultaneously diagonalizable such that $\hat{\rho} = \sum_{i=1}^{d} p_i|\varphi_i\rangle\langle\varphi_i|$ and $\hat{\sigma} = \sum_{i=1}^{d} q_i|\varphi_i\rangle\langle\varphi_i|$ with the same basis, $S_1(\hat{\rho}\|\hat{\sigma})$ reduces to the classical KL divergence of the diagonal distributions of $\hat{\rho}$ and $\hat{\sigma}$, i.e., $S_1(p\|q) = S_1(\hat{\rho}\|\hat{\sigma})$. We note that $S_1(\hat{U}\hat{\rho}\hat{U}^\dagger\|\hat{U}\hat{\sigma}\hat{U}^\dagger) = S_1(\hat{\rho}\|\hat{\sigma})$ holds for any unitary $\hat{U}$.

As in the classical case, the quantum KL divergence is an asymmetric "distance" between two quantum states. In fact, the quantum KL divergence is non-negative:

$$S_1(\hat{\rho}\|\hat{\sigma}) \geq 0, \tag{5.16}$$

where the equality $S_1(\hat{\rho}\|\hat{\sigma}) = 0$ holds if and only if $\hat{\rho} = \hat{\sigma}$. The proof of this is straightforward (e.g., Theorem 11.7 of Ref [4]), but we will prove it as Corollary A.2 in Appendix A as a special case of a more general statement. We here note that if two states are unnormalized, the non-negativity (5.16) is generalized as $S_1(\hat{\rho}\|\hat{\sigma}) \geq \mathrm{tr}[\hat{\rho} - \hat{\sigma}]$, which is called the Klein's inequality.

Another fundamental property of the quantum KL divergence is the monotonicity (or the data processing inequality):

Theorem 5.1 (Monotonicity) *For any CPTP map $\mathscr{E}$,*

$$S_1(\hat{\rho}\|\hat{\sigma}) \geq S_1(\mathscr{E}(\hat{\rho})\|\mathscr{E}(\hat{\sigma})). \tag{5.17}$$

This implies that, as in the classical case discussed in Sect. 2.2, the KL divergence is a monotone under CPTP maps. Note that, however, the KL divergence is not a complete monotone, as it does not provide a sufficient condition for state convertibility, again as in the classical case discussed in Chap. 3.

The proof of the monotonicity in the quantum case is nontrivial, in contrast to the classical counterpart (2.8) whose proof was easy. Historically, Lieb and Ruskai [103, 104] proved the strong subadditivity of the von Neumann entropy (inequality (5.21)) on the basis of the Lieb's theorem [105] (see Ref. [106] for a self-contained review), which was then rephrased as the monotonicity of the quantum KL divergence [107–109]. Later, Petz gave a simple proof of the monotonicity [17] (see also Refs. [110–112]), on which we focus in this book. In Appendix A, we will prove the monotonicity of general divergence-like quantities called the Petz's quasi-entropies [16, 17]; Theorem 5.1 above is an immediate consequence of Corollary A.4.

We consider the change in the von Neumann entropy by CPTP unital maps. Let $\mathscr{E}$ be CPTP unital. By noting Eq. (5.15) and the monotonicity (5.17), we have

$$S_1(\hat{\rho}) \leq S_1(\mathscr{E}(\hat{\rho})), \tag{5.18}$$

which implies that a CPTP unital map makes quantum states more "random," as in the classical counterpart (2.11) with doubly-stochastic maps. In other words, the von Neumann entropy is a monotone under CPTP unital maps.

We next discuss some important properties of the von Neumann entropy and the quantum KL divergence. We consider two systems A and B and their composite system AB. Let $\hat{\rho}_{AB}$ be a density operator of AB and $\hat{\rho}_A$, $\hat{\rho}_B$ be its partial states. Then, the *subadditivity* of the von Neumann entropy is expressed as

$$S_1(\hat{\rho}_{AB}) \leq S_1(\hat{\rho}_A) + S_1(\hat{\rho}_B), \tag{5.19}$$

where the equality holds if and only if $\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B$. This is a straightforward consequence of the non-negativity of the quantum KL divergence: $S_1(\hat{\rho}_A) + S_1(\hat{\rho}_B) - S_1(\hat{\rho}_{AB}) = S_1(\hat{\rho}_{AB} \| \hat{\rho}_A \otimes \hat{\rho}_B) \geq 0$. The left-hand side above is called the mutual information between A and B, written as

$$I_1(\hat{\rho}_{AB})_{A:B} := S_1(\hat{\rho}_A) + S_1(\hat{\rho}_B) - S_1(\hat{\rho}_{AB}) \geq 0. \tag{5.20}$$

Note that for any CPTP map of the form $\mathscr{E}_A \otimes \mathscr{E}_B$ acting independently on A and B, the monotonicity of the KL divergence (5.17) implies the data processing inequality of mutual information: $I_1(\hat{\rho}_{AB})_{A:B} \geq I_1(\mathscr{E}_A \otimes \mathscr{E}_B(\hat{\rho}_{AB}))_{A:B}$.

A much stronger property than the subadditivity is the strong subadditivity. We consider three subsystems A, B, C. Then,

$$S_1(\hat{\rho}_{ABC}) + S_1(\hat{\rho}_B) \leq S_1(\hat{\rho}_{AB}) + S_1(\hat{\rho}_{BC}). \tag{5.21}$$

This can be proved from the monotonicity (5.17) of the quantum KL divergence. Let $\hat{\sigma}_A := \hat{I}_A/d_A$, where $\hat{I}_A$ is the identity of $\mathscr{H}_A$ and d_A is its dimension. We then have $[S_1(\hat{\rho}_{AB}) + S_1(\hat{\rho}_{BC})] - [S_1(\hat{\rho}_{ABC}) + S_1(\hat{\rho}_B)] = [S_1(\hat{\rho}_{BC}) - S_1(\hat{\rho}_{ABC})] - [S_1(\hat{\rho}_B) - S_1(\hat{\rho}_{AB})] = S_1(\hat{\rho}_{ABC} \| \hat{\sigma}_A \otimes \hat{\rho}_{BC}) - S_1(\hat{\rho}_{AB} \| \hat{\sigma}_A \otimes \hat{\rho}_B) \geq 0$, where we used the monotonicity (5.17) for the partial trace $\mathscr{E}(\hat{\rho}_{ABC}) := \mathrm{tr}_C[\hat{\rho}_{ABC}] = \hat{\rho}_{AB}$ that is CPTP as mentioned before.

As seen in the above proof, the strong subadditivity (5.21) is equivalent to the monotonicity of the partial trace. Conversely, the monotonicity of any CPTP map follows from the monotonicity of the partial trace because of Eq. (5.6). Thus, the strong subadditivity (5.21) is essentially equivalent to the monotonicity (5.17) of CPTP maps.

We next discuss the convex/concave properties. The quantum KL divergence satisfies the following property called the joint convexity.

Theorem 5.2 (Joint convexity) *Let $\hat{\rho} = \sum_k p_k \hat{\rho}_k$ and $\hat{\sigma} = \sum_k p_k \hat{\sigma}_k$, where $\hat{\rho}_k$ and $\hat{\sigma}_k$ are quantum states and p_k's represent a classical distribution with $p_k > 0$. Then,*

$$S_1(\hat{\rho}\|\hat{\sigma}) \le \sum_k p_k S_1(\hat{\rho}_k\|\hat{\sigma}_k). \tag{5.22}$$

The equality holds if $\mathscr{P}_k$'s are orthogonal to each other, where $\mathscr{P}_k$ is the subspace spanned by the supports of $\hat{\rho}_k$ and $\hat{\sigma}_k$.

Proof We start with the case of the equality. By assumption, $\hat{\rho}$ and $\hat{\sigma}$ are simultaneously block-diagonal with subspaces $\mathscr{P}_k$'s, and thus we have

$$S_1(\hat{\rho}\|\hat{\sigma}) = \mathrm{tr}\left[\left(\sum_k p_k \hat{\rho}_k\right)\ln\left(\sum_k p_k \hat{\rho}_k\right) - \left(\sum_k p_k \hat{\rho}_k\right)\ln\left(\sum_k p_k \hat{\sigma}_k\right)\right] \tag{5.23}$$

$$= \mathrm{tr}\left[\sum_k p_k \hat{\rho}_k \ln(p_k \hat{\rho}_k) - \sum_k p_k \hat{\rho}_k \ln(p_k \hat{\sigma}_k)\right] \tag{5.24}$$

$$= \sum_k p_k S_1(\hat{\rho}_k\|\hat{\sigma}_k). \tag{5.25}$$

In general, let $\{|k\rangle\}$ be an orthonormal basis of an auxiliary system A, and define $\hat{\rho}' := \sum_k p_k \hat{\rho}_k \otimes |k\rangle\langle k|$, $\hat{\sigma}' := \sum_k p_k \hat{\sigma}_k \otimes |k\rangle\langle k|$. From the equality case, we have

$$S_1(\hat{\rho}'\|\hat{\sigma}') = \sum_k p_k S_1(\hat{\rho}_k \otimes |k\rangle\langle k|\|\hat{\sigma}_k \otimes |k\rangle\langle k|) = \sum_k p_k S_1(\hat{\rho}_k\|\hat{\sigma}_k). \tag{5.26}$$

Note that $\hat{\rho} = \mathrm{tr}_A[\hat{\rho}']$, $\hat{\sigma} = \mathrm{tr}_A[\hat{\sigma}']$. Because the partial trace is CPTP, the monotonicity of the quantum KL divergence implies that $S_1(\hat{\rho}\|\hat{\sigma}) \le S_1(\hat{\rho}'\|\hat{\sigma}') = \sum_k p_k S_1(\hat{\rho}_k\|\hat{\sigma}_k)$. $\square$

By letting $\hat{\sigma}_k = \hat{I}/d$ for all k in the above theorem, we obtain the concavity of the von Neumann entropy:

$$\sum_k p_k S_1(\hat{\rho}_k) \le S_1(\hat{\rho}). \tag{5.27}$$

The equality is achieved if we have a single k, i.e., $p_1 = 1$ and $\hat{\rho} = \hat{\rho}_1$. We note that this itself is easily provable, without invoking the joint convexity of the quantum KL divergence: Consider $\hat{\rho}' := \sum_k p_k \hat{\rho}_k \otimes |k\rangle\langle k|$ and apply the subadditivity (5.19):

$$S_1(\hat{\rho}) + S_1(p) \ge S_1(\hat{\rho}') = \sum_k p_k S_1(\hat{\rho}_k) + S_1(p), \tag{5.28}$$

where, to obtain the right equality, we used that $\hat{\rho}_k \otimes |k\rangle\langle k|$'s are mutually orthogonal and $S_1(\hat{\rho}_k \otimes |k\rangle\langle k|) = S_1(\hat{\rho}_k)$ holds.

Related to this, an upper bound of the von Neumann entropy is also known:

$$S_1(\hat{\rho}) \leq \sum_k p_k S_1(\hat{\rho}_k) + S_1(p), \tag{5.29}$$

where the equality holds if and only if the supports of $\hat{\rho}_k$'s are mutually orthogonal. We omit the proof of this (see, e.g., Theorem 11.10 of [4]).

The relation between the quantum KL divergence and the second law of thermodynamics is completely parallel to the classical case discussed in Sect. 4.1. We here briefly discuss the second law in the quantum case, while a more detailed argument will be provided in Chap. 8.

Let $\hat{H}$ be the Hamiltonian of the system. The corresponding Gibbs state is defined as $\hat{\rho}^G := e^{-\beta\hat{H}}/Z$, where $Z := \text{tr}[e^{-\beta\hat{H}}]$ is the partition function. The equilibrium free energy is given by $F := -\beta^{-1}\ln Z$. As in the classical case (4.1), a CPTP map $\mathscr{E}$ is called a *Gibbs-preserving map*, if it satisfies

$$\mathscr{E}(\hat{\rho}^G) = \hat{\rho}^G. \tag{5.30}$$

From the monotonicity (5.17) of the quantum KL divergence, we have $S_1(\hat{\rho}\|\hat{\rho}^G) \geq S_1(\mathscr{E}(\hat{\rho})\|\hat{\rho}^G)$ for any Gibbs-preserving map $\mathscr{E}$. In parallel to inequality (4.4) of the classical case, we can rewrite the above inequality as

$$\Delta S_1 \geq \beta Q, \tag{5.31}$$

where $\Delta S_1 := S_1(\mathscr{E}(\hat{\rho})) - S_1(\hat{\rho})$ is the change in the von Neumann entropy and $Q := \text{tr}[(\mathscr{E}(\hat{\rho}) - \hat{\rho})\hat{H}]$ is the heat absorption. This is a Clausius-type representation of the second law or a generalized Landauer principle in the quantum case [34]. If the Gibbs state is given by $\hat{\rho}^G = \hat{I}/d$ (in the case of $\beta = 0$ or $\hat{H} \propto \hat{I}$), the Gibbs-preserving map is unital and the second law (5.31) reduces to inequality (5.18).

In the quench-and-relax processes discussed in Sect. 4.1 where the Hamiltonian is driven by an external agent, we can also derive the work bound in the same form as the classical counterpart (4.10). In Sect. 8.4, we will derive the second law of this form, where we adopt the setup where the total Hamiltonian is time-independent by including the "clock" degrees of freedom.

5.3 Quantum Rényi Entropy and Divergence

We next consider quantum analogues of the Rényi entropies and divergences. The Rényi α-entropy can be straightforwardly defined in the same manner as the classical case: For $0 < \alpha < \infty$,

$$S_\alpha(\hat{\rho}) := \frac{1}{1-\alpha} \ln\left(\text{tr}[\hat{\rho}^\alpha]\right), \tag{5.32}$$

where $\hat{\rho}$ is a density operator of dimension d. In particular, $S_1(\hat{\rho})$ is the von Neumann entropy. We also define

$$S_0(\hat{\rho}) := \ln(\text{rank}[\hat{\rho}]), \tag{5.33}$$

$$S_\infty(\hat{\rho}) := -\ln \|\hat{\rho}\|_\infty, \tag{5.34}$$

which are the max and the min Rényi entropies.

The quantum Rényi entropy equals the classical one of the diagonal distribution: $S_\alpha(\hat{\rho}) = S_\alpha(p)$ with p being the diagonal distribution of $\hat{\rho}$. We thus have $0 \leq S_\alpha(\hat{\rho}) \leq \ln d$, and $S_\alpha(\hat{\rho}) \geq S_{\alpha'}(\hat{\rho})$ for $\alpha \leq \alpha'$. It is also obvious that $S_\alpha(\hat{U}\hat{\rho}\hat{U}^\dagger) = S_\alpha(\hat{\rho})$ holds for any unitary $\hat{U}$.

On the other hand, the quantum versions of the quantum Rényi divergence are not unique due to the non-commutability of density operators $\hat{\rho}, \hat{\sigma}$ [112, 114–118] (see also Appendix A). Here, we mainly focus on the following two limiting cases with $\alpha = 0, \infty$ [65]:

$$S_0(\hat{\rho}\|\hat{\sigma}) := -\ln\left(\text{tr}[\hat{P}_{\hat{\rho}}\hat{\sigma}]\right), \tag{5.35}$$

where $\hat{P}_{\hat{\rho}}$ is the projection onto the support of $\hat{\rho}$, and

$$S_\infty(\hat{\rho}\|\hat{\sigma}) := \ln\left(\min\{\lambda : \hat{\rho} \leq \lambda\hat{\sigma}\}\right) = \ln\|\hat{\sigma}^{-1/2}\hat{\rho}\hat{\sigma}^{-1/2}\|_\infty. \tag{5.36}$$

We again note that we always assume that the support of $\hat{\rho}$ is included in that of $\hat{\sigma}$ for these divergence quantities. We remark on some properties: $S_\alpha(\hat{\rho}) = \ln d - S_\alpha(\hat{\rho}\|\hat{I}/d)$, $S_\alpha(\hat{U}\hat{\rho}\hat{U}^\dagger\|\hat{U}\hat{\sigma}\hat{U}^\dagger) = S_\alpha(\hat{\rho}\|\hat{\sigma})$ for any unitary $\hat{U}$, and

$$S_\alpha(\hat{\rho} \otimes \hat{\rho}'\|\hat{\sigma} \otimes \hat{\sigma}') = S_\alpha(\hat{\rho}\|\hat{\sigma}) + S_\alpha(\hat{\rho}'\|\hat{\sigma}'). \tag{5.37}$$

Also, these divergences are non-negative and satisfy the following:

Theorem 5.3 (Lemma 5 and Lemma 6 of [65])

$$0 \leq S_0(\hat{\rho}\|\hat{\sigma}) \leq S_\infty(\hat{\rho}\|\hat{\sigma}). \tag{5.38}$$

Here, $S_0(\hat{\rho}\|\hat{\sigma}) = 0$ holds if and only if the supports of $\hat{\rho}$ and $\hat{\sigma}$ are the same, and $S_\infty(\hat{\rho}\|\hat{\sigma}) = 0$ holds if and only if $\hat{\rho} = \hat{\sigma}$.

Proof $0 \leq S_0(\hat{\rho}\|\hat{\sigma})$ is obvious from the definition. $S_0(\hat{\rho}\|\hat{\sigma}) \leq S_\infty(\hat{\rho}\|\hat{\sigma})$ can be shown as follows. Let $\lambda := e^{S_\infty(\hat{\rho}\|\hat{\sigma})}$. Then by definition, $0 \leq \lambda\hat{\sigma} - \hat{\rho}$ holds. Thus $0 \leq \text{tr}[(\lambda\hat{\sigma} - \hat{\rho})\hat{P}_{\hat{\rho}}] = \lambda e^{-S_0(\hat{\rho}\|\hat{\sigma})} - 1$, which implies $S_0(\hat{\rho}\|\hat{\sigma}) \leq S_\infty(\hat{\rho}\|\hat{\sigma})$.

The equality $S_0(\hat{\rho}\|\hat{\sigma}) = 0$ holds if and only if the support of $\hat{\rho}$ includes that of $\hat{\sigma}$, but in our setup, the latter always includes the former.

Suppose that $S_\infty(\hat{\rho}\|\hat{\sigma}) = 0$ holds. By definition, $\hat{\Delta} := \hat{\sigma} - \hat{\rho}$ is positive. If $\hat{\Delta} \neq 0$, there exists a positive eigenvalue and any other eigenvalues are non-negative, and thus $\text{tr}[\hat{\Delta}] > 0$. But this contradicts the normalization of $\hat{\rho}$ and $\hat{\sigma}$. $\qquad\square$

The following relation is also known.

Proposition 5.1

$$S_0(\hat{\rho}\|\hat{\sigma}) \leq S_1(\hat{\rho}\|\hat{\sigma}) \leq S_\infty(\hat{\rho}\|\hat{\sigma}). \tag{5.39}$$

Proof The following proof is based on [119].

We first prove the right inequality. By definition, $e^{S_\infty(\hat{\rho}\|\hat{\sigma})}\hat{\sigma} \geq \hat{\rho}$ holds, and thus $S_\infty(\hat{\rho}\|\hat{\sigma}) - S_1(\hat{\rho}\|\hat{\sigma}) = \mathrm{tr}[\hat{\rho}(\ln(e^{S_\infty(\hat{\rho}\|\hat{\sigma})}\hat{\sigma}) - \ln \hat{\rho})] \geq 0$. Here, to obtain the last inequality, we used the fact that $\ln x$ is operator monotone (see Appendix A.2).

To prove the left inequality, let $\hat{\rho} = \sum_i p_i |\varphi_i\rangle\langle\varphi_i|$ and $\hat{\sigma} = \sum_j q_j |\psi_j\rangle\langle\psi_j|$. Define $\hat{\tau} := \sum_{i:p_i>0} e^{\langle\varphi_i|\ln\hat{\sigma}|\varphi_i\rangle}|\varphi_i\rangle\langle\varphi_i|/Z$, where $Z := \sum_{i:p_i>0} e^{\langle\varphi_i|\ln\hat{\sigma}|\varphi_i\rangle}$. Note that $Z = \sum_{i:p_i>0} e^{\sum_j |\langle\varphi_i|\psi_j\rangle|^2 \ln q_j} \leq \sum_{i:p_i>0} \sum_j |\langle\varphi_i|\psi_j\rangle|^2 q_j = \mathrm{tr}[\hat{P}_{\hat{\rho}}\hat{\sigma}]$, where we used the Jensen inequality for the probability distribution $\{|\langle\varphi_i|\psi_j\rangle|^2\}_{j=1}^d$ for each i. We then have $0 \leq S_1(\hat{\rho}\|\hat{\tau}) = S_1(\hat{\rho}\|\hat{\sigma}) + \ln Z \leq S_1(\hat{\rho}\|\hat{\sigma}) + \ln \mathrm{tr}[\hat{P}_{\hat{\rho}}\hat{\sigma}] = S_1(\hat{\rho}\|\hat{\sigma}) - S_0(\hat{\rho}\|\hat{\sigma})$. □

An alternative way to understand (5.39) is to consider two kinds of Rényi α-divergences, both of which are defined for $0 < \alpha < 1$ and $1 < \alpha < \infty$ (see Appendix A for details). One is the simple Rényi divergence $\tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma})$, defined in Eq. (A.35) [112, 114]. The other is the sandwiched Rényi divergence $S_\alpha(\hat{\rho}\|\hat{\sigma})$, defined in Eq. (A.68) [115–118]. Both are non-decreasing functions in α as in the classical case (Proposition 2.3), and satisfy $\lim_{\alpha\to 1} \tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) = \lim_{\alpha\to 1} S_\alpha(\hat{\rho}\|\hat{\sigma}) = S_1(\hat{\rho}\|\hat{\sigma})$. It is also known that $\lim_{\alpha\to +0} \tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) = S_0(\hat{\rho}\|\hat{\sigma})$ and $\lim_{\alpha\to\infty} S_\alpha(\hat{\rho}\|\hat{\sigma}) = S_\infty(\hat{\rho}\|\hat{\sigma})$. Thus we have inequalities (5.39).

It is also known that these divergences satisfy the monotonicity (and thus are monotones) under CPTP maps, as in the case for $\alpha = 1$:

Theorem 5.4 (Monotonicity of the quantum Rényi divergence, Lemma 7 of [65]) *For $\alpha = 0, \infty$, if $\mathcal{E}$ is CPTP,*

$$S_\alpha(\hat{\rho}\|\hat{\sigma}) \geq S_\alpha(\mathcal{E}(\hat{\rho})\|\mathcal{E}(\hat{\sigma})). \tag{5.40}$$

Proof We first prove the case of $\alpha = \infty$ [65]. As shown below, the positivity of $\mathcal{E}$ is enough to show the monotonicity. From the positivity of $\mathcal{E}$, $\lambda\hat{\sigma} - \hat{\rho} \geq 0$ implies $\lambda\mathcal{E}(\hat{\sigma}) - \mathcal{E}(\hat{\rho}) \geq 0$, and therefore $\{\lambda : \lambda\hat{\sigma} - \hat{\rho} \geq 0\} \subset \{\lambda : \lambda\mathcal{E}(\hat{\sigma}) - \mathcal{E}(\hat{\rho}) \geq 0\}$. We thus have $\min\{\lambda : \lambda\hat{\sigma} - \hat{\rho} \geq 0\} \geq \min\{\lambda : \lambda\mathcal{E}(\hat{\sigma}) - \mathcal{E}(\hat{\rho}) \geq 0\}$, which implies $S_\infty(\hat{\rho}\|\hat{\sigma}) \geq S_\infty(\mathcal{E}(\hat{\rho})\|\mathcal{E}(\hat{\sigma}))$.

We next prove the case of $\alpha = 0$. As shown below, it is enough to assume that $\mathcal{E}$ is positive and trace-nonincreasing. We start with a useful expression

$$\mathrm{tr}[\hat{P}_{\hat{\rho}}\hat{\sigma}] = \min_{0\leq\hat{Q}\leq\hat{I}, \mathrm{tr}[\hat{\rho}\hat{Q}]=1} \mathrm{tr}[\hat{\sigma}\hat{Q}], \tag{5.41}$$

where $0 \leq \hat{Q} \leq \hat{I}$ and $\mathrm{tr}[\hat{\rho}\hat{Q}] = 1$ imply that $\hat{Q}$ is the direct sum of the identity on the support of $\hat{\rho}$ and a positive operator on its orthogonal complement (see Lemma A.2).

We note that the adjoint of a trace-nonincreasing map $\mathscr{E}$ is sub-unital, i.e., $\mathscr{E}^\dagger(\hat{I}) \leq \hat{I}$. Then, we have

$$\min_{0 \leq \hat{Q} \leq \hat{I}, \mathrm{tr}[\mathscr{E}(\hat{\rho})\hat{Q}]=1} \mathrm{tr}[\mathscr{E}(\hat{\sigma})\hat{Q}] = \min_{0 \leq \hat{Q} \leq \hat{I}, \mathrm{tr}[\hat{\rho}\mathscr{E}^\dagger(\hat{Q})]=1} \mathrm{tr}[\hat{\sigma}\mathscr{E}^\dagger(\hat{Q})] \tag{5.42}$$

$$\geq \min_{0 \leq \hat{Q}' \leq \hat{I}, \mathrm{tr}[\hat{\rho}\hat{Q}']=1} \mathrm{tr}[\hat{\sigma}\hat{Q}'], \tag{5.43}$$

where we used that $0 \leq \hat{Q} \leq \hat{I}$ implies $0 \leq \mathscr{E}^\dagger(\hat{Q}) \leq \mathscr{E}^\dagger(\hat{I}) \leq \hat{I}$ for positive and sub-unital $\mathscr{E}^\dagger$ to obtain the second line. $\qquad\square$

The expression (5.41) in the above proof implies that $S_0(\hat{\rho}\|\hat{\sigma})$ is equivalent to the hypothesis testing divergence $S_\mathrm{H}^\eta(\hat{\rho}\|\hat{\sigma})$, introduced in Eq. (B.1) in Appendix B.1, with $\eta = 1$. See also Proposition B.2 for the monotonicity with $0 < \eta < 1$.

We next consider the joint convexity for $\alpha = 0$.

Theorem 5.5 (Joint convexity of Rényi 0-divergence) *We use the same notation as Theorem 5.2 with $p_k > 0$. Then,*

$$S_0(\hat{\rho}\|\hat{\sigma}) \leq \sum_k p_k S_0(\hat{\rho}_k\|\hat{\sigma}_k). \tag{5.44}$$

The equality holds if the same equality condition as in Theorem 5.2 is satisfied and $S_0(\hat{\rho}_k\|\hat{\sigma}_k)$'s are the same for all k.

Proof From $\hat{P}_{\hat{\rho}} \geq \hat{P}_{\hat{\rho}_k}$, we have

$$-\ln\left(\mathrm{tr}[\hat{P}_{\hat{\rho}}\hat{\sigma}]\right) = -\ln\left(\mathrm{tr}\left[\hat{P}_{\hat{\rho}}\sum_k p_k\hat{\sigma}_k\right]\right) \leq -\ln\left(\mathrm{tr}\left[\sum_k p_k\hat{P}_{\hat{\rho}_k}\hat{\sigma}_k\right]\right). \tag{5.45}$$

By noting the convexity of $-\ln x$,

$$-\ln\left(\mathrm{tr}\left[\sum_k p_k\hat{P}_{\hat{\rho}_k}\hat{\sigma}_k\right]\right) \leq -\sum_k p_k\ln\left(\mathrm{tr}[\hat{P}_{\hat{\rho}_k}\hat{\sigma}_k]\right), \tag{5.46}$$

and thus we obtain inequality (5.44). The equality in (5.44) holds if the equality conditions for the above two inequalities are satisfied. $\qquad\square$

The above proof is essentially the same as the proof of Theorem 11 of Ref. [69] for the classical case. We remark that the joint convexity does not hold for $\alpha = \infty$ (see also Corollary A.7 in Appendix A).

Meanwhile, for $\alpha = 0, 1, \infty$, we can define $S_\alpha(\hat{\rho}\|\hat{\sigma})$ for unnormalized $\hat{\sigma} \geq 0$, which is useful to define the nonequilibrium free energy (see also Chap. 8). For $Z > 0$, it is easy to check the scaling property

$$S_\alpha(\hat{\rho}\|\hat{\sigma}/Z) = S_\alpha(\hat{\rho}\|\hat{\sigma}) + \ln Z. \tag{5.47}$$

In general, if $\hat{\sigma} \leq \hat{\sigma}'$, then

$$S_\alpha(\hat{\rho}\|\hat{\sigma}') \leq S_\alpha(\hat{\rho}\|\hat{\sigma}). \tag{5.48}$$

Again it is easy to check this for $\alpha = 0, \infty$. For $\alpha = 1$, we use the fact that $\ln x$ is operator monotone (see Appendix A). The monotonicity of the α-divergence with $\alpha = 0, 1, \infty$ under CPTP maps applies to unnormalized states (see Theorem A.3 that is true for unnormalized states).

Chapter 6
Quantum Majorization

In this chapter, we consider what majorization means in quantum systems. The ordinary majorization can be directly generalized to the quantum case, as discussed in Sect. 6.1. In Sect. 6.2, we discuss a subclass of CPTP unital maps, called noisy operations. On the other hand, the quantum analogue of d-majorization is not very straightforward, while we still have a simple characterization of state conversion by the Rényi 0- and ∞-divergences as discussed in Sect. 6.3. In Sect. 6.4, we briefly remark on resource theory of entanglement and discuss its relation to infinite-temperature thermodynamics.

In this chapter, Chaps. 7, and 8, we assume that the dimensions of the input and output Hilbert spaces of CPTP maps are the same, i.e., $\mathscr{E} : \mathscr{L}(\mathscr{H}) \to \mathscr{L}(\mathscr{H})$, unless stated otherwise.

6.1 Quantum Majorization

We first consider the quantum version of ordinary majorization, which can be formulated in a parallel manner to the classical case (see also Ref. [46]). Let $\hat{\rho}$ and $\hat{\rho}'$ be quantum states of dimension d, whose spectral decompositions are given by

$$\hat{\rho} = \sum_{i=1}^{d} p_i |\varphi_i\rangle\langle\varphi_i|, \ \hat{\rho}' = \sum_{i=1}^{d} p_i' |\varphi_i'\rangle\langle\varphi_i'|, \tag{6.1}$$

where their bases are not necessarily the same. $p := (p_1, \ldots, p_d)^{\mathrm{T}}$ and $p' := (p_1', \ldots, p_d')^{\mathrm{T}}$ are their diagonal distributions. Then, we define quantum majorization as follows.

Definition 6.1 (*Quantum majorization*) We say that a quantum state $\hat{\rho}$ majorizes another state $\hat{\rho}'$, written as $\hat{\rho}' \prec \hat{\rho}$, if their diagonal distributions satisfy $p' \prec p$.

Based on this definition, we have the following theorem.

Theorem 6.1 *Let $\hat{\rho}$, $\hat{\rho}'$ be quantum states. The following are equivalent.*

(i) $\hat{\rho}' \prec \hat{\rho}$.

(ii) *There exists a CPTP unital map $\mathscr{E}$ such that $\hat{\rho}' = \mathscr{E}(\hat{\rho})$.*

(iii) *There exists a CPTP map that is a mixture of unitaries $\mathscr{E}$ such that $\hat{\rho}' = \mathscr{E}(\hat{\rho})$. Here, a mixture of unitaries is given in the form*

$$\mathscr{E}(\hat{\rho}) = \sum_k r_k \hat{U}_k \hat{\rho} \hat{U}_k^\dagger, \tag{6.2}$$

where $\hat{U}_k$ is unitary and $\sum_k r_k = 1$, $r_k \geq 0$.

Proof In the following proof, we use the notations of Eq. (6.1).

We first show (i) $\Rightarrow$ (ii). Since $\hat{\rho}' \prec \hat{\rho}$ implies $p' \prec p$ by definition, Theorem 3.1 for classical majorization implies that there exists a doubly stochastic matrix T such that $p' = Tp$. We then define a map

$$\mathscr{E}(\hat{\rho}) := \sum_{ji} T_{ji} |\varphi_j'\rangle \langle \varphi_i | \hat{\rho} | \varphi_i \rangle \langle \varphi_j' |. \tag{6.3}$$

Since $\sum_{ji} \hat{M}_{ji}^\dagger \hat{M}_{ji} = \hat{I}$ with $\hat{M}_{ji} := \sqrt{T_{ji}} |\varphi_j'\rangle \langle \varphi_i |$, $\mathscr{E}$ is CPTP. Also, since T is doubly stochastic, $\mathscr{E}(\hat{I}) = \hat{I}$ holds and thus $\mathscr{E}$ is unital. Finally, it is obvious that $\hat{\rho}' = \mathscr{E}(\hat{\rho})$.

We next show (ii) $\Rightarrow$ (i). Suppose that $\mathscr{E}$ is CPTP unital, and define

$$T_{ji} := \langle \varphi_j' | \mathscr{E}(|\varphi_i\rangle \langle \varphi_i |) | \varphi_j' \rangle. \tag{6.4}$$

Then, since $\sum_i T_{ji} = \sum_j T_{ji} = 1$ and $T_{ji} \geq 0$, T_{ji} is doubly stochastic. Also, it is obvious that $p_j' = \sum_i T_{ji} p_i$. Therefore, $p' \prec p$ holds, and thus $\hat{\rho}' \prec \hat{\rho}$.

(i) $\Leftrightarrow$ (iii) (or (ii) $\Leftrightarrow$ (iii)) is called the Uhlmann's theorem (Theorem 12.13 of Ref. [4]). Since (iii) $\Rightarrow$ (ii) is obvious, here we only show (i) $\Rightarrow$ (iii). Suppose that $\hat{\rho}' \prec \hat{\rho}$. Then, there exists a doubly stochastic matrix T such that $p' = Tp$ for the diagonal distributions. From the Birkhoff's theorem (Theorem 3.2), T can be written as $T = \sum_k r_k P_k$ with permutation matrices P_k. Let $\hat{P}_k$ be the operator that acts as P_k in the basis $\{|\varphi_k\rangle\}$, and let $\hat{V}$ be the unitary operator that converts the basis $\{|\varphi_k\rangle\}$ into $\{|\varphi_k'\rangle\}$. Then, we have $\hat{\rho}' = \sum_k r_k \hat{V} \hat{P}_k \hat{\rho} \hat{P}_k^\dagger \hat{V}^\dagger$. By noting that $\hat{U}_k := \hat{V} \hat{P}_k$ is unitary, we have $\hat{\rho}' = \sum_k r_k \hat{U}_k \hat{\rho} \hat{U}_k^\dagger$. □

While (ii) $\Leftrightarrow$ (iii) is true in Theorem 6.1, the following fact is known: The Birkhoff's theorem (Theorem 3.2) is not true in the quantum case.

Theorem 6.2   *For $d \geq 3$ with d being the dimension of the system, the set of mixtures of unitaries is a strict subset of CPTP unital maps. That is, there are CPTP unital maps that cannot be written as any mixtures of unitaries (6.2). In other words, an extreme point of CPTP unital maps is not necessarily unitary.*

Proof   We show an example of an operator that is CPTP unital but not a mixture of unitaries [120]. Let $\hat{S}_k$ $(k = x, y, z)$ be the spin-l operator (that is a representation of $\mathfrak{su}(2)$) with $l = 1/2, 1, 3/2, \ldots$. We define a CPTP map

$$\mathscr{E}(\hat{\rho}) := \frac{1}{l(l+1)} \sum_{k=x,y,z} \hat{S}_k \hat{\rho} \hat{S}_k, \tag{6.5}$$

which is shown to be unital. For $l \geq 1$ (or equivalently $d \geq 3$ because of $d = 2l + 1$), $\hat{S}_k$ is not unitary. On the other hand, $\mathscr{E}$ is an extreme point of CPTP unital maps, because of the Choi's theorem (Theorem 5 of Ref. [102]). Therefore, $\mathscr{E}$ in Eq. (6.5) cannot be written as any mixture of unitaries.

We note that Choi's theorem is stated as follows. Let $\mathscr{E}$ be a CPTP unital map that has a Kraus representation $\mathscr{E}(\hat{\rho}) := \sum_k \hat{M}_k \hat{\rho} \hat{M}_k^\dagger$ with $\sum_k \hat{M}_k \hat{M}_k^\dagger = \sum_k \hat{M}_k^\dagger \hat{M}_k = \hat{I}$. Then, $\mathscr{E}$ is an extreme point of CPTP unital maps, if and only if $\hat{M}_k \hat{M}_l^\dagger$'s are linearly independent in the operator space.   □

6.2   Noisy Operations

We next consider a subclass of CPTP unital maps, called *noisy operations*. A noisy operation is also a special case of thermal operations introduced in Sect. 8.2.

Definition 6.2   (*Noisy operations*) A CPTP map $\mathscr{E}$ is an exact noisy operation, if there exists an auxiliary system B with Hilbert space dimension $d_B < \infty$ and exists a unitary operator $\hat{U}$ acting on the composite system, such that

$$\mathscr{E}(\hat{\rho}) = \mathrm{tr}_B[\hat{U} \hat{\rho} \otimes (\hat{I}_B/d_B) \hat{U}^\dagger], \tag{6.6}$$

where $\hat{I}_B$ is the identity operator of B. Furthermore, a CPTP map $\mathscr{E}$ is a noisy operation, if there exists a sequence of exact noisy operations $\{\mathscr{E}_n\}_{n=1}^\infty$ such that $\mathscr{E}_n$ converges to $\mathscr{E}$ in $n \to \infty$.

In Ref. [121], exact noisy operations are called exactly factorizable maps, and noisy operations are called strongly factorizable maps. There is also a concept called factorizable maps [122], for which infinite-dimensional auxiliary systems are allowed with the use of von Neumann algebras. We note that the set of factorizable maps coincide with the set of strongly factorizable maps if and only if Connes' embedding conjecture is true [121] (see Ref. [124] for a recent work on the conjecture).

We next discuss the relationship between unital maps and noisy operations.

Proposition 6.1 (Lemma 5 of [46]) *Any noisy operation is a unital map. Also, any mixture of unitaries is a noisy operation.*

Proof The former is obvious from Eq. (6.6). To show the latter, let $\mathscr{E}(\hat{\rho}) = \sum_k r_k \hat{U}_k \hat{\rho} \hat{U}_k^\dagger$ be the mixture of unitaries with $r_k > 0$. We first consider the case that all of r_k's are rational numbers written as l_k/m with positive integers l_k and m. Then we take an auxiliary Hilbert space in dimension $d_B := m$, and divide it into subspaces labeled by k with dimensions l_k. Then we can take $\hat{U}$ in Eq. (6.6) as a controlled unitary, which acts on the system as $\hat{U}_k$ if the auxiliary state is in the kth subspace. Finally, if some of r_k's are irrational, we can approximate them by rational numbers and then take the limit of $m \to \infty$. $\square$

The above inclusions are strict.

Proposition 6.2 *There are CPTP unital maps that cannot be written as any noisy operations. Also, there are noisy operations that cannot be written as any mixtures of unitaries.*

Proof An example of the former is shown in Example 3.1 of Ref. [122], which is the same as Eq. (6.5) (with $l = 1$). An example of the latter is discussed in Ref. [121] based on Ref. [123]. $\square$

Although the above inclusions are strict, we have the following because any mixture of unitaries is a noisy operation.

Corollary 6.1 *The conditions of Theorem 6.1 is further equivalent to:*
(iv) There exists a noisy operation $\mathscr{E}$ such that $\hat{\rho}' = \mathscr{E}(\hat{\rho})$.

We finally remark on the classical case, where the set of noisy operations and the set of doubly stochastic matrices are equivalent.

Definition 6.3 (*Classical noisy operations*) A stochastic matrix T is an exact classical noisy operation, if there exists an auxiliary system B in dimension $d_B < \infty$ and exists a permutation matrix P acting on the composite system such that

$$(Tp)_i = \sum_{jkl} P_{ij;kl} p_k u_l, \tag{6.7}$$

where $u := (1/d_B, \ldots, 1/d_B)^T$ is the uniform distribution of B. Furthermore, a stochastic matrix T is a classical noisy operation, if there exists a sequence of exact classical noisy operations $\{T_n\}_{n=1}^\infty$ such that T_n converges to T in $n \to \infty$.

Proposition 6.3 (Lemma 6 of [46]) *For any classical stochastic map T, the following are equivalent.*
(i) T is a doubly stochastic map.
(ii) T is a noisy operation.

Proof Because (ii) $\Rightarrow$ (i) is obvious, we only show (i) $\Rightarrow$ (ii). The proof is parallel to the latter part of Proposition 6.1. Let $T = \sum_k r_k P_k$ be the decomposition of the Birkhoff's theorem (Theorem 3.2). If all of r_k's are rational, we divide the sample space (i.e., the set of labels of components of probability vectors) into subspaces labeled by k with dimensions proportional to r_k. Then, apply the controlled permutation acting on the system as P_k if the auxiliary state is in the kth subspace. For general r_k's, take the limit. $\qquad\square$

6.3 Quantum d-Majorization

We now consider quantum d-majorization (or quantum relative majorization). It is not very straightforward to formulate the quantum counterpart of the classical case discussed in Sect. 3.2, especially the "Lorenz curve-like" characterization of the necessary and sufficient condition for the existence of a CPTP map $\mathscr{E}$ satisfying $\hat{\rho}' = \mathscr{E}(\hat{\rho})$ and $\hat{\sigma}' = \mathscr{E}(\hat{\sigma})$ for given pairs of quantum states $(\hat{\rho}, \hat{\sigma})$, $(\hat{\rho}', \hat{\sigma}')$ (i.e., the full quantum version of Theorem. 3.3 or the d-majorization version of Theorem 6.1). A difficulty lies in the fact that $\hat{\rho}$ and $\hat{\sigma}$ are not necessarily commutable (and thus not simultaneously diagonalizable) in the quantum case.

There are several interesting approaches to formulate full quantum d-majorization on the basis of, e.g., a generalized Lorenz curve [125] and matrix majorization [126]. We will mention the former later in this section. Here, however, we adopt a most naive approach to quantum d-majorization: we simply *define* quantum d-majorization in the following form, given (iv) of Theorem 3.3 of the classical case.

Definition 6.4 (*Quantum d-majorization*) Let $\hat{\rho}, \hat{\sigma}, \hat{\rho}', \hat{\sigma}'$ be quantum states. We say that a pair $(\hat{\rho}, \hat{\sigma})$ d-majorizes another pair $(\hat{\rho}', \hat{\sigma}')$, written as $(\hat{\rho}', \hat{\sigma}') \prec (\hat{\rho}, \hat{\sigma})$, if there exists a CPTP map $\mathscr{E}$ such that

$$\hat{\rho}' = \mathscr{E}(\hat{\rho}), \quad \hat{\sigma}' = \mathscr{E}(\hat{\sigma}). \tag{6.8}$$

In particular, we say that $\hat{\rho}$ thermo-majorizes $\hat{\rho}'$ with respect to $\hat{\sigma}$, if $(\hat{\rho}', \hat{\sigma}) \prec (\hat{\rho}, \hat{\sigma})$.

Under this definition, the characterization of state convertibility by the Rényi 0- and ∞-divergences still works, as described by the following theorem. This is the quantum counterpart of Theorem 3.4 of the classical case. While Theorem 3.4 was obvious if one looks at the classical Lorenz curves, the following theorem is proved without invoking such graphical representation.

Theorem 6.3 (Conditions for state conversion)

(a) Necessary conditions: If $(\hat{\rho}', \hat{\sigma}') \prec (\hat{\rho}, \hat{\sigma})$ holds, then

$$S_0(\hat{\rho}'\|\hat{\sigma}') \leq S_0(\hat{\rho}\|\hat{\sigma}), \quad S_\infty(\hat{\rho}'\|\hat{\sigma}') \leq S_\infty(\hat{\rho}\|\hat{\sigma}). \tag{6.9}$$

(b) Sufficient condition: $(\hat{\rho}', \hat{\sigma}') \prec (\hat{\rho}, \hat{\sigma})$ *holds, if (but not only if)*

$$S_\infty(\hat{\rho}'\|\hat{\sigma}') \leq S_0(\hat{\rho}\|\hat{\sigma}). \tag{6.10}$$

Proof (a) This is the monotonicity of $S_0(\hat{\rho}\|\hat{\sigma})$ and $S_\infty(\hat{\rho}\|\hat{\sigma})$ (Theorem 5.4).

(b) The following proof is based on Refs. [45, 119]. We explicitly construct a CPTP map $\mathscr{E}$ that transforms $(\hat{\rho}, \hat{\sigma})$ to $(\hat{\rho}', \hat{\sigma}')$ by a "measure-and-prepare" method.

First, we perform a projection measurement $\{\hat{P}_{\hat{\rho}}, \hat{I} - \hat{P}_{\hat{\rho}}\}$, where $\hat{P}_{\hat{\rho}}$ is the projection onto the support of $\hat{\rho}$. We label the measurement outcome corresponding to $\hat{P}_{\hat{\rho}}$ and $\hat{I} - \hat{P}_{\hat{\rho}}$ by "0" and "1", respectively. Then, the conditional probabilities of getting these outcomes are given by

$$P(0|\hat{\rho}) = 1, \ P(1|\hat{\rho}) = 0, \ P(0|\hat{\sigma}) = c, \ P(1|\hat{\sigma}) = 1 - c, \tag{6.11}$$

where $c := e^{-S_0(\hat{\rho}\|\hat{\sigma})}$. We note that this measurement corresponds to the hypothesis testing with $\eta = 1$ (see Eq. (5.41) and Appendix B).

Next, we prepare a state depending on the outcome. Suppose that $c \neq 1$. When the outcome is "0", we prepare $\hat{\rho}'$. When the outcome is "1", we prepare

$$\hat{\sigma}'' := \frac{\hat{\sigma}' - c\hat{\rho}'}{1 - c}. \tag{6.12}$$

Here, if $S_\infty(\hat{\rho}'\|\hat{\sigma}') \leq S_0(\hat{\rho}\|\hat{\sigma})$, then $\hat{\sigma}'' \geq (\hat{\sigma}' - e^{-S_\infty(\hat{\rho}'\|\hat{\sigma}')}\hat{\rho}')/(1 - c) \geq 0$, and thus $\hat{\sigma}''$ is a (normalized) quantum state. By this state preparation, we obtain the final state $c\hat{\rho}' + (1 - c)\hat{\sigma}'' = \hat{\sigma}'$ from the initial state $\hat{\sigma}$. If $c = 1$, $S_\infty(\hat{\rho}'\|\hat{\sigma}') = 0$ or equivalently $\hat{\rho}' = \hat{\sigma}'$ holds, and thus the state preparation is trivial.

In summary, the constructed CPTP map is given by (for $c \neq 1$)

$$\mathscr{E}(\hat{\tau}) := \mathrm{tr}[\hat{P}_{\hat{\rho}}\hat{\tau}]\hat{\rho}' + \left(1 - \mathrm{tr}[\hat{P}_{\hat{\rho}}\hat{\tau}]\right)\frac{\hat{\sigma}' - c\hat{\rho}'}{1 - c}. \tag{6.13}$$

$\square$

Figure 6.1 schematically illustrates the sufficient condition of Theorem 6.3. In the case of Fig. 6.1a, the 0- and ∞-divergences of $(\hat{\rho}, \hat{\sigma})$ and those of $(\hat{\rho}', \hat{\sigma}')$ are completely separated, and the sufficient condition of Theorem 6.3 is satisfied. On the other hand, in the case of Fig. 6.1b, we cannot judge whether state transformation is possible from $(\hat{\rho}, \hat{\sigma})$ to $(\hat{\rho}', \hat{\sigma}')$, because their divergences are not separated. This illustrates that a necessary and sufficient condition cannot be given only by the 0- and ∞-divergences in general. In Chap. 7, however, we will see that the 0- and ∞-divergences can approximately collapse to a single value under certain conditions, if we take the asymptotic limit; In such a case, a complete monotone emerges and provides a necessary and sufficient condition for state conversion by d-majorization. We also note that the sufficient condition (b) is of no use when $\hat{\rho}$ and $\hat{\sigma}$ have the same support, while it will turn out useful when smoothing S_0.

Fig. 6.1 Schematic of situations where **a** the sufficient condition $S_0(\hat{\rho}\|\hat{\sigma}) \geq S_\infty(\hat{\rho}'\|\hat{\sigma}')$ works and **b** does not work
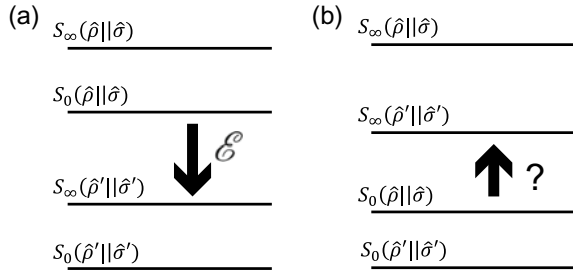
We next briefly discuss a quantum analogue of the Lorenz curve in line with Ref. [125], which, however, cannot characterize state convertibility in general.

Definition 6.5 (*Quantum Lorenz curve [125]*) Let $\hat{\rho}$, $\hat{\sigma}$ be quantum states. The quantum Lorenz curve of $(\hat{\rho}, \hat{\sigma})$ is defined as the upper boundary of the set $\{(x, y) : x = \mathrm{tr}[\hat{Q}\hat{\rho}], y = \mathrm{tr}[\hat{Q}\hat{\sigma}], 0 \leq \hat{Q} \leq \hat{I}\} \subset \mathbb{R}^2$. Then, we write $(\hat{\rho}', \hat{\sigma}') \prec_{\mathrm{q}} (\hat{\rho}, \hat{\sigma})$, if the quantum Lorenz curve of $(\hat{\rho}, \hat{\sigma})$ lies above that of $(\hat{\rho}', \hat{\sigma}')$.

If $[\hat{\rho}, \hat{\sigma}] = 0$ and $[\hat{\rho}', \hat{\sigma}'] = 0$, the above definition reduces to the classical case. In the quantum case, however, $(\hat{\rho}', \hat{\sigma}') \prec_{\mathrm{q}} (\hat{\rho}, \hat{\sigma})$ is *not* equivalent to $(\hat{\rho}', \hat{\sigma}') \prec (\hat{\rho}, \hat{\sigma})$ of Definition 6.4 [130, 131]. Instead, it is worth noting the following theorem.

Theorem 6.4 (Theorem 2 of [125]) *Let $\hat{\rho}$, $\hat{\rho}'$, $\hat{\sigma}$, $\hat{\sigma}'$ be quantum states. The following are equivalent.*

(i) $(\hat{\rho}', \hat{\sigma}') \prec_{\mathrm{q}} (\hat{\rho}, \hat{\sigma})$.
(ii) *For all $t \in \mathbb{R}$,* $\|\hat{\rho}' - t\hat{\sigma}'\|_1 \leq \|\hat{\rho} - t\hat{\sigma}\|_1$.
(iii) *For all $0 \leq \eta \leq 1$,* $Q_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma}) \leq Q_{\mathrm{H}}^\eta(\hat{\rho}'\|\hat{\sigma}')$ *holds, where* $Q_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma}) := \min_{0 \leq \hat{Q} \leq \hat{I}, \mathrm{tr}[\hat{\rho}\hat{Q}] \geq \eta} \mathrm{tr}[\hat{\sigma}\hat{Q}]$.

Here, the classical counterpart of (i) $\Leftrightarrow$ (ii) is that of Theorem 3.3. In (iii) above, $Q_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma}) = \eta e^{-S_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma})}$ holds for $0 < \eta < 1$, where $S_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma})$ is the hypothesis testing divergence defined by Eq. (B.1) of Appendix B.

6.4 Remark on Resource Theory of Entanglement

Let us very briefly mention the relationship between majorization and entanglement. (See, e.g., Ref. [4] for details. We also note Refs. [127, 128] as seminal works that made an analogy between entanglement and thermodynamics.)

Suppose that a bipartite system AB is shared by Alice and Bob. They are allowed to perform any quantum operations on their own systems and to communicate by classical channels, which is called local operations and classical communications

(LOCC). For simplicity, we suppose that the state of AB is a pure state $|\Psi\rangle$. In this case, product states are free states and LOCC are free operations.

By the Schmidt decomposition, the pure state can be always written as $|\Psi\rangle = \sum_i \sqrt{p_i} |\varphi_{A,i}\rangle |\varphi_{B,i}\rangle$ with a classical probability distribution p_i and orthonormal bases $\{|\varphi_{A,i}\rangle\}$, $\{|\varphi_{B,i}\rangle\}$ of A, B. Then, the reduced density operator of A and B are respectively given by $\hat{\rho}_A = \sum_i p_i |\varphi_{A,i}\rangle\langle\varphi_{A,i}|$ and $\hat{\rho}_B = \sum_i p_i |\varphi_{B,i}\rangle\langle\varphi_{B,i}|$. Since these reduced density operators are the same up to a unitary, we will only focus on $\hat{\rho}_A$, and write it as $\hat{\rho}_\Psi$ to explicitly show its dependency on the global state $|\Psi\rangle$.

Entanglement of pure states can be quantified by the von Neumann entropy $S_1(\hat{\rho}_\Psi)$, which is called entanglement entropy. While the von Neumann entropy is a monotone under LOCC, it is not a complete monotone. On the other hand, majorization can completely characterize state convertibility under LOCC, which is shown in the following theorem.

Theorem 6.5 (Theorem 1 of [52]) $|\Psi\rangle$ *can be converted into* $|\Psi'\rangle$ *by LOCC, if and only if* $\hat{\rho}_\Psi \prec \hat{\rho}_{\Psi'}$.

This is reminiscent of Theorem 6.1 related to infinite-temperature thermodynamics. However, there is a crucial difference between entanglement and thermodynamics. In fact, the direction of state conversion is opposite to that of majorization in the case of entanglement, in contrast to the case of thermodynamics at infinite temperature. This is because the completely mixed state (the uniform distribution) corresponds to the maximally entangled state that has the maximum resource of entanglement, while in thermodynamics at infinite temperature, the completely mixed state is just a free state.

Chapter 7
Approximate and Asymptotic Majorization and Divergence

In this chapter, we consider an approximate (or "smoothed") version of the Rényi 0- and ∞-divergences, called the smooth divergences. We then consider the asymptotic limit of these smooth divergences, leading to the theory of *information spectrum* with the upper and lower spectral divergence rates corresponding to the 0- and ∞-divergences. This formalism naturally leads to characterization of asymptotic state convertibility of many-body systems even beyond independent and identically distributed (i.i.d.) situations.

In Sect. 7.1, we introduce the smooth 0- and ∞-divergences and the corresponding approximate majorization. In Sect. 7.2, we consider the asymptotic limit and introduce the spectral divergence rates, and show their relation to asymptotic d-majorization. In Sect. 7.3, we investigate the condition with which the upper and lower spectral divergence rates collapse to a single value. This is the quantum version of the asymptotic equipartition property (AEP). In such a case, asymptotic state convertibility by d-majorization can be characterized by a single complete monotone. In Sect. 7.4, we make some technical remarks on the smoothing of the divergences.

Throughout this chapter, we consider the general setup: quantum d-majorization and the corresponding quantum divergences. The classical counterpart is immediately obtained as a special case of the quantum formulation.

The central concepts introduced in this chapter are summarized as follows. First, for $\varepsilon > 0$, we introduce smooth divergences $S_0^\varepsilon(\hat{\rho}\|\hat{\sigma})$ and $S_\infty^\varepsilon(\hat{\rho}\|\hat{\sigma})$, which are ε-approximations of the 0- and ∞-divergences, respectively. We next consider the asymptotic limit, where $\hat{\rho}_n$ and $\hat{\sigma}_n$ are regarded as states of a large system of size n. A simplest example is the i.i.d. case, where $\hat{\rho}_n := \hat{\rho}^{\otimes n}$ and $\hat{\sigma}_n := \hat{\sigma}^{\otimes n}$. We emphasize, however, that our formulation here is very general and not restricted to the i.i.d. case. Then, we take the asymptotic limit of $S_0^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n)$ and $S_\infty^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n)$ by taking $n \to \infty$ first and then $\varepsilon \to +0$, and we obtain the lower and the upper spectral divergence rates, respectively. We show that the spectral divergence rates give a characterization of state convertibility in the asymptotic regime. Under certain conditions including ergodicity, we show that $S_0^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n) \simeq S_\infty^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n) \simeq S_1(\hat{\rho}_n\|\hat{\sigma}_n)$ holds with $n \to \infty$

and $\varepsilon \to +0$, which is a consequence of the quantum AEP and implies that the KL divergence rate serves as a complete monotone.

We note that quantum hypothesis testing is a useful tool in this chapter, which gives a divergence-like quantity called the hypothesis testing divergence $S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma})$. While we postpone the details of hypothesis testing to Appendix B, here we only remark that, roughly speaking, $S_{\mathrm{H}}^{\eta\simeq 1}(\hat{\rho}\|\hat{\sigma}) \simeq S_0^{\varepsilon\simeq 0}(\hat{\rho}\|\hat{\sigma})$ and $S_{\mathrm{H}}^{\eta\simeq 0}(\hat{\rho}\|\hat{\sigma}) \simeq S_{\infty}^{\varepsilon\simeq 0}(\hat{\rho}\|\hat{\sigma})$ hold (see Proposition B.1 for a rigorous formulation). Thus, the hypothesis testing divergence has essentially the same information as the smooth 0- and ∞-divergences; this is particularly the case when we consider the asymptotic limit, because the correction terms are independent of n.

7.1　Smooth Divergence and Approximate Majorization

We consider approximate state conversion and smooth divergences in line with the framework developed in Refs. [66, 67]. Specifically, we consider approximate quantum d-majorization, and correspondingly, the smooth quantum Rényi 0- and ∞-divergences introduced in Ref. [65].

Let $\varepsilon \geq 0$. We define the ε-neighborhood of a normalized state $\hat{\rho}$ by

$$B^{\varepsilon}(\hat{\rho}) := \{\hat{\tau} : \ D(\hat{\tau}, \hat{\rho}) \leq \varepsilon, \mathrm{tr}[\hat{\tau}] = 1, \hat{\tau} \geq 0\}. \tag{7.1}$$

Here we adopted the trace distance as a norm for smoothing, while this choice is not essential and other norms can be adopted equivalently. Also, only normalized states are allowed in the above definition, while this is again not essential. See also Sect. 7.4 for alternative ways of smoothing.

We now define approximate d-majorization and smooth divergences.

Definition 7.1 (*Approximate d-majorization*) Let $\varepsilon \geq 0$. We say that a pair $(\hat{\rho}, \hat{\sigma})$ ε-approximately d-majorizes another pair $(\hat{\rho}', \hat{\sigma}')$, written as $(\hat{\rho}', \hat{\sigma}') \prec^{\varepsilon} (\hat{\rho}, \hat{\sigma})$, if there exists $\hat{\tau}' \in B^{\varepsilon}(\hat{\rho}')$ such that $(\hat{\tau}', \hat{\sigma}') \prec (\hat{\rho}, \hat{\sigma})$.

Definition 7.2 (*Smooth Rényi 0/∞-divergence*)

$$S_{\infty}^{\varepsilon}(\hat{\rho}\|\hat{\sigma}) := \min_{\hat{\tau}\in B^{\varepsilon}(\hat{\rho})} S_{\infty}(\hat{\tau}\|\hat{\sigma}), \tag{7.2}$$

$$S_0^{\varepsilon}(\hat{\rho}\|\hat{\sigma}) := \max_{\hat{\tau}\in B^{\varepsilon}(\hat{\rho})} S_0(\hat{\tau}\|\hat{\sigma}). \tag{7.3}$$

The smooth divergences satisfy the monotonicity, which is the smoothed version of Theorem 5.4 or Theorem 6.3 (a).

Theorem 7.1 (Monotonicity of the smooth divergences) $(\hat{\rho}', \hat{\sigma}') \prec (\hat{\rho}, \hat{\sigma})$ *implies that for any $\varepsilon > 0$,*

(a) $S_{\infty}^{\varepsilon}(\hat{\rho}\|\hat{\sigma}) \geq S_{\infty}^{\varepsilon}(\hat{\rho}'\|\hat{\sigma}')$.

(b) $S_0^\varepsilon(\hat\rho\|\hat\sigma) \geq S_0^{\varepsilon^2/6}(\hat\rho'\|\hat\sigma') + \ln(\varepsilon^2/6).$

Proof Let $\mathscr{E}$ be a CPTP map such that $\hat\rho' = \mathscr{E}(\hat\rho)$ and $\hat\sigma' = \mathscr{E}(\hat\sigma)$.

(a) The proof is parallel to that of Lemma 59 of Ref. [46] for ordinary majorization:

$$S_\infty^\varepsilon(\hat\rho'\|\hat\sigma') := \min_{\hat\tau' \in B^\varepsilon(\hat\rho')} S_\infty(\hat\tau'\|\hat\sigma') \tag{7.4}$$

$$\leq \min_{\hat\tau \in B^\varepsilon(\hat\rho)} S_\infty(\mathscr{E}(\hat\tau)\|\hat\sigma') \tag{7.5}$$

$$\leq \min_{\hat\tau \in B^\varepsilon(\hat\rho)} S_\infty(\hat\tau\|\hat\sigma) \tag{7.6}$$

$$=: S_\infty^\varepsilon(\hat\rho\|\hat\sigma). \tag{7.7}$$

Here, to obtain the second line, we used that $\{\mathscr{E}(\hat\tau) : \hat\tau \in B^\varepsilon(\hat\rho)\} \subset B^\varepsilon(\hat\rho')$ holds from the monotonicity of the trace distance. We also used the monotonicity of $S_\infty(\hat\rho\|\hat\sigma)$ to obtain the third line.

(b) From the monotonicity of the hypothesis testing divergence $S_H^\eta(\hat\rho\|\hat\sigma)$ (Proposition B.2) and Proposition B.1 in Appendix B, we have

$$S_0^\varepsilon(\hat\rho\|\hat\sigma) \geq S_H^{1-\varepsilon'}(\hat\rho\|\hat\sigma) - \ln\frac{1-\varepsilon'}{\varepsilon'} \tag{7.8}$$

$$\geq S_H^{1-\varepsilon'}(\hat\rho'\|\hat\sigma') - \ln\frac{1-\varepsilon'}{\varepsilon'} \tag{7.9}$$

$$\geq S_0^{\varepsilon'}(\hat\rho'\|\hat\sigma') + \ln\varepsilon', \tag{7.10}$$

where $\varepsilon' := \varepsilon^2/6$. $\qquad\square$

The above theorem can be generalized to the case where state conversion is not exact.

Theorem 7.2 (Necessary conditions for approximate state conversion) *For any $\varepsilon \geq 0$, $(\hat\rho', \hat\sigma') \prec^\varepsilon (\hat\rho, \hat\sigma)$ implies that:*

(a) For any $\delta \geq 0$,

$$S_\infty^\delta(\hat\rho\|\hat\sigma) \geq S_\infty^{\varepsilon+\delta}(\hat\rho'\|\hat\sigma'). \tag{7.11}$$

(b) For any δ satisfying $\delta^2/6 > \varepsilon$,

$$S_0^\delta(\hat\rho\|\hat\sigma) \geq S_0^{\delta^2/6-\varepsilon}(\hat\rho'\|\hat\sigma') + \ln(\delta^2/6). \tag{7.12}$$

Proof If $(\hat\rho', \hat\sigma') \prec^\varepsilon (\hat\rho, \hat\sigma)$, there exists a CPTP map $\mathscr{E}$ such that $\hat\sigma' = \mathscr{E}(\hat\sigma)$ and $D(\hat\rho', \hat\tau') \leq \varepsilon$ with $\hat\tau' := \mathscr{E}(\hat\rho)$.

(a) The proof of this case is parallel to that of Lemma 64 of Ref. [46]. Theorem 7.1 (a) implies $S_\infty^\delta(\hat\rho\|\hat\sigma) \geq S_\infty^\delta(\hat\tau'\|\hat\sigma')$. We can take $\hat\tau''$ such that $S_\infty(\hat\tau''\|\hat\sigma') =$

$S_\infty^\delta(\hat{\tau}'\|\hat{\sigma}')$, and then $D(\hat{\tau}'', \hat{\rho}') \leq D(\hat{\tau}'', \hat{\tau}') + D(\hat{\tau}', \hat{\rho}') \leq \delta + \varepsilon$. Therefore, $S_\infty^\delta(\hat{\tau}'\|\hat{\sigma}') = S_\infty(\hat{\tau}''\|\hat{\sigma}') \geq S_\infty^{\varepsilon+\delta}(\hat{\rho}'\|\hat{\sigma}')$.

(b) The following proof is based on Ref. [119]. Let $\delta' := \delta^2/6$. From Theorem 7.1 (b), we have $S_0^\delta(\hat{\rho}\|\hat{\sigma}) \geq S_0^{\delta'}(\hat{\tau}'\|\hat{\sigma}') + \ln(\delta')$. Meanwhile, let $\hat{\tau}''$ be the optimal choice such that $S_0^{\delta'-\varepsilon}(\hat{\rho}'\|\hat{\sigma}') = S_0(\hat{\tau}''\|\hat{\sigma}')$ with $D(\hat{\tau}'', \hat{\rho}') \leq \delta' - \varepsilon$. From $D(\hat{\tau}', \hat{\tau}'') \leq D(\hat{\tau}', \hat{\rho}') + D(\hat{\rho}', \hat{\tau}'') \leq \delta'$, $\hat{\tau}''$ is a candidate for maximization in $S_0^{\delta'}(\hat{\tau}'\|\hat{\sigma}')$. Thus, $S_0^{\delta'}(\hat{\tau}'\|\hat{\sigma}') \geq S_0^{\delta'-\varepsilon}(\hat{\rho}'\|\hat{\sigma}')$. $\qquad\square$

We next consider the approximate version of Theorem 6.3 (b).

Theorem 7.3 (Sufficient condition for approximate state conversion) *For $\varepsilon \geq 0$, $(\hat{\rho}', \hat{\sigma}') \prec^\varepsilon (\hat{\rho}, \hat{\sigma})$ holds if (but not only if)*

$$S_\infty^{\varepsilon/2}(\hat{\rho}'\|\hat{\sigma}') \leq S_0^{\varepsilon/2}(\hat{\rho}\|\hat{\sigma}). \tag{7.13}$$

Proof The proof is parallel to that of Lemma 63 of Ref. [46]. We can take $\hat{\tau}'$, $\hat{\tau}$ such that $S_\infty^{\varepsilon/2}(\hat{\rho}'\|\hat{\sigma}') = S_\infty(\hat{\tau}'\|\hat{\sigma}')$, $S_0^{\varepsilon/2}(\hat{\rho}\|\hat{\sigma}) = S_0(\hat{\tau}\|\hat{\sigma})$, and $D(\hat{\tau}', \hat{\rho}') \leq \varepsilon/2$, $D(\hat{\tau}, \hat{\rho}) \leq \varepsilon/2$. From Theorem 6.3 (b), we have $(\hat{\tau}', \hat{\sigma}') \prec (\hat{\tau}, \hat{\sigma})$, that is, there exists $\mathscr{E}$ such that $\hat{\tau}' = \mathscr{E}(\hat{\tau})$ C $\hat{\sigma}' = \mathscr{E}(\hat{\sigma})$. By defining $\hat{\tau}'' := \mathscr{E}(\hat{\rho})$, we have

$$D(\hat{\tau}'', \hat{\rho}') \leq D(\hat{\tau}'', \hat{\tau}') + D(\hat{\tau}', \hat{\rho}') \leq D(\hat{\rho}, \hat{\tau}) + D(\hat{\tau}', \hat{\rho}') \leq \varepsilon/2 + \varepsilon/2 = \varepsilon, \tag{7.14}$$

which implies $(\hat{\rho}', \hat{\sigma}') \prec^\varepsilon (\hat{\rho}, \hat{\sigma})$. $\qquad\square$

7.2 Information Spectrum and Asymptotic Majorization

We next consider information spectrum and the asymptotic limit of majorization. The concept of information spectrum has been introduced by Han and Verdú [59, 60] and generalized to the quantum regime by Nagaoka and Hayashi [61]. The relevant quantities have been rewritten in terms of the smooth entropies and divergences by Datta and Renner [64, 65], on which the argument of this section is based.

To take the asymptotic limit of d-majorization, we consider a sequence of quantum states, $\widehat{P} := \{\hat{\rho}_n\}_{n \in \mathbb{N}}$, where $\hat{\rho}_n \in \mathscr{S}(\mathscr{H}^{\otimes n})$. This can be an arbitrary sequence, and is not restricted to an i.i.d. sequence. We note that, in the case of i.i.d., we can write $\hat{\rho}_n := \hat{\rho}^{\otimes n}$ with $\hat{\rho} \in \mathscr{S}(\mathscr{H})$. First, we define the asymptotic limit of the von Neumann entropy rate and the KL divergence rate:

Definition 7.3 Let $\widehat{P} := \{\hat{\rho}_n\}$, $\widehat{\Sigma} := \{\hat{\sigma}_n\}$ be sequences of quantum states. The von Neumann entropy rate is defined as

$$S_1(\widehat{P}) := \lim_{n \to \infty} \frac{1}{n} S_1(\hat{\rho}_n), \tag{7.15}$$

and the KL divergence rate is defined as

$$S_1(\widehat{P}\|\widehat{\Sigma}) := \lim_{n\to\infty} \frac{1}{n} S_1(\hat{\rho}_n\|\hat{\sigma}_n). \tag{7.16}$$

We note that these limits do not necessarily exist.

We next define asymptotic state conversion for sequences of quantum states in terms of d-majorization.

Definition 7.4 (*Asymptotic quantum d-majorization*) Let $\widehat{P} := \{\hat{\rho}_n\}$, $\widehat{\Sigma} := \{\hat{\sigma}_n\}$, $\widehat{P}' := \{\hat{\rho}'_n\}$, $\widehat{\Sigma}' := \{\hat{\sigma}'_n\}$ be sequences of quantum states. Then, $(\widehat{P}, \widehat{\Sigma})$ asymptotically d-majorizes $(\widehat{P}', \widehat{\Sigma}')$, written as $(\widehat{P}', \widehat{\Sigma}') \prec^a (\widehat{P}, \widehat{\Sigma})$, if there exists a sequence of CPTP maps $\{\mathscr{E}_n\}$ such that

$$\lim_{n\to\infty} D(\mathscr{E}_n(\hat{\rho}_n), \hat{\rho}'_n) = 0, \quad \mathscr{E}_n(\hat{\sigma}_n) = \hat{\sigma}'_n. \tag{7.17}$$

Or equivalently, for any $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $(\hat{\rho}'_n, \hat{\sigma}'_n) \prec^\varepsilon (\hat{\rho}_n, \hat{\sigma}_n)$ holds for all $n \geq N_\varepsilon$.

We emphasize that in the above definition the state conversion of $\hat{\sigma}_n$ should be exact. In terms of thermodynamics, this means that Gibbs-preserving maps should preserve Gibbs states exactly (see also Sect. 8.7).

We now introduce the upper and lower spectral divergence rates for sequences of quantum states, which are respectively given by the asymptotic limit of ∞- and 0-divergences [65].

Definition 7.5 (*Quantum spectral divergence rates*) The upper and lower spectral divergence rates are respectively defined as

$$\overline{S}(\widehat{P}\|\widehat{\Sigma}) := \lim_{\varepsilon\to+0} \limsup_{n\to\infty} \frac{1}{n} S_\infty^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n), \tag{7.18}$$

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) := \lim_{\varepsilon\to+0} \liminf_{n\to\infty} \frac{1}{n} S_0^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n). \tag{7.19}$$

These quantities play a key role in the theory of information spectrum. We note that $\lim_{\varepsilon\to+0}$ always exists, because $S_0^\varepsilon(\hat{\rho}\|\hat{\sigma})$ and $S_\infty^\varepsilon(\hat{\rho}\|\hat{\sigma})$ monotonically change in ε. Also, the order of the limits of n and ε is crucial in the above definition. In fact, the convergence in n is not uniform on $\varepsilon > 0$ in many cases.

These quantities have been originally introduced in Ref. [61] in the form

$$\overline{S}(\widehat{P}\|\widehat{\Sigma}) = \inf\left\{a : \limsup_{n\to\infty} \text{tr}\left[\hat{P}\left\{\hat{\rho}_n - e^{na}\hat{\sigma}_n \geq 0\right\}\hat{\rho}_n\right] = 0\right\}, \tag{7.20}$$

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) = \sup\left\{a \ : \ \liminf_{n\to\infty} \mathrm{tr}\left[\hat{P}\left\{\hat{\rho}_n - e^{na}\hat{\sigma}_n \geq 0\right\}\hat{\rho}_n\right] = 1\right\}, \qquad (7.21)$$

where $\hat{P}\left\{\hat{X} \geq 0\right\}$ denotes the projector onto the eigenspace of $\hat{X}$ with non-negative eigenvalues. The equivalence between the above expression and Definition 7.5 has been proved in Theorem 2 and Theorem 3 of Ref. [65].

We note that (Proposition of [62])

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) \leq \overline{S}(\widehat{P}\|\widehat{\Sigma}). \qquad (7.22)$$

In fact, from Proposition B.1 and inequality (B.2), we have for $0 < \varepsilon < 1/3$

$$\frac{1}{n}S_0^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n) \leq \frac{1}{n}S_\infty^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n) + \frac{1}{n}\ln\frac{2}{1-\varepsilon}. \qquad (7.23)$$

Then, take $\liminf_{n\to\infty}$ and $\limsup_{n\to\infty}$, and next $\varepsilon \to +0$.

We next show that the spectral divergence rates satisfy the monotonicity (and thus are monotones) under asymptotic state conversion. This is regarded as a necessary condition for state conversion as the asymptotic limit of Theorem 7.2. We note that the following theorem is slightly general than the monotonicity proved in Proposition 4 of Ref. [62] where state conversion is assumed to be exact.

Theorem 7.4 (Necessary conditions for state conversion; Monotonicity of the spectral divergence rates) *If $(\widehat{P}', \widehat{\Sigma}') \prec^a (\widehat{P}, \widehat{\Sigma})$, then*

$$\overline{S}(\widehat{P}'\|\widehat{\Sigma}') \leq \overline{S}(\widehat{P}\|\widehat{\Sigma}), \quad \underline{S}(\widehat{P}'\|\widehat{\Sigma}') \leq \underline{S}(\widehat{P}\|\widehat{\Sigma}). \qquad (7.24)$$

Proof From the assumption, for any $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $(\hat{\rho}_n', \hat{\sigma}_n') \prec^\varepsilon (\hat{\rho}_n, \hat{\sigma}_n)$ for all $n \geq N_\varepsilon$.

We first prove the former inequality. Theorem 7.2 (a) implies that for any $\delta > 0$, $S_\infty^{\varepsilon+\delta}(\hat{\rho}_n'\|\hat{\sigma}_n') \leq S_\infty^\delta(\hat{\rho}_n\|\hat{\sigma}_n)$ holds for all $n \geq N_\varepsilon$. We first take $\limsup_{n\to\infty}$ by dividing the above inequality by n and by fixing ε and δ, and then take the limit $\varepsilon \to +0$ and $\delta \to +0$. Then we have $\overline{S}(\widehat{P}'\|\widehat{\Sigma}') \leq \overline{S}(\widehat{P}\|\widehat{\Sigma})$.

For the latter inequality, the proof goes in a similar way. By applying Theorem 7.2 (b), we have for any δ satisfying $\delta^2/6 > \varepsilon$, $S_0^\delta(\hat{\rho}_n\|\hat{\sigma}_n) \geq S_0^{\delta^2/6-\varepsilon}(\hat{\rho}_n'\|\hat{\sigma}_n') + \ln(\delta^2/6)$ for all $n \geq N_\varepsilon$. We first take $\liminf_{n\to\infty}$ by dividing the above inequality by n and by fixing ε and δ, and then take the limit $\varepsilon \to +0$ and $\delta \to +0$ by keeping $\delta^2/6 > \varepsilon$. Then we have $\underline{S}(\widehat{P}'\|\widehat{\Sigma}') \leq \underline{S}(\widehat{P}\|\widehat{\Sigma})$. $\qquad\square$

We next consider a sufficient condition for asymptotic state conversion, which is obtained as the limit of Theorem 7.3.

Theorem 7.5 (Sufficient condition for asymptotic state conversion) $(\widehat{P}', \widehat{\Sigma}') \prec^a (\widehat{P}, \widehat{\Sigma})$ *holds, if (but not only if)*

$$\overline{S}(\widehat{P}'\|\widehat{\Sigma}') < \underline{S}(\widehat{P}\|\widehat{\Sigma}). \tag{7.25}$$

Proof Suppose that $\overline{S}(\widehat{P}'\|\widehat{\Sigma}') < \underline{S}(\widehat{P}\|\widehat{\Sigma})$. For any sufficiently small $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $S_\infty^{\varepsilon/2}(\hat{\rho}'_n\|\hat{\sigma}'_n) \leq S_0^{\varepsilon/2}(\hat{\rho}_n\|\hat{\sigma}_n)$ holds for all $n \geq N_\varepsilon$. Thus, from Theorem 7.3, we have $(\hat{\rho}'_n, \hat{\sigma}'_n) \prec^\varepsilon (\hat{\rho}_n, \hat{\sigma}_n)$ for all $n \geq N_\varepsilon$. This implies $(\widehat{P}', \widehat{\Sigma}') \prec^\mathrm{a} (\widehat{P}, \widehat{\Sigma})$. $\qquad\square$

We note that the equality is excluded from inequality (7.25). In fact, the above proof does not work in the equality case. Moreover, it is known that there indeed exists an example that the equality $\overline{S}(\widehat{P}'\|\widehat{\Sigma}') = \underline{S}(\widehat{P}\|\widehat{\Sigma})$ is satisfied but asymptotic state conversion from $(\widehat{P}, \widehat{\Sigma})$ to $(\widehat{P}', \widehat{\Sigma}')$ is impossible, which has been shown in Refs. [132, 133] (but we need to slightly change the setup there). This is a topic of the second-order asymptotics [134–137].

We now consider a special case that the upper and lower spectral divergence rates collapse to a single value $S(\widehat{P}\|\widehat{\Sigma})$. In such a case, we obtain a (almost) necessary and sufficient characterization for state conversion (where "almost" means that the equality case mentioned above is excluded). The characterization is given by a single scalar entropy-like function $S(\widehat{P}\|\widehat{\Sigma})$ (i.e., a (almost) complete monotone).

Theorem 7.6 *Suppose that the upper and lower spectral divergence rates coincide:* $\overline{S}(\widehat{P}\|\widehat{\Sigma}) = \underline{S}(\widehat{P}\|\widehat{\Sigma}) =: S(\widehat{P}\|\widehat{\Sigma})$ *and* $\overline{S}(\widehat{P}'\|\widehat{\Sigma}') = \underline{S}(\widehat{P}'\|\widehat{\Sigma}') =: S(\widehat{P}'\|\widehat{\Sigma}')$. *Then,*

(a) $(\widehat{P}', \widehat{\Sigma}') \prec^\mathrm{a} (\widehat{P}, \widehat{\Sigma})$ *implies* $S(\widehat{P}'\|\widehat{\Sigma}') \leq S(\widehat{P}\|\widehat{\Sigma})$.
(b) $S(\widehat{P}'\|\widehat{\Sigma}') < S(\widehat{P}\|\widehat{\Sigma})$ *implies* $(\widehat{P}', \widehat{\Sigma}') \prec^\mathrm{a} (\widehat{P}, \widehat{\Sigma})$.

The condition that the upper and lower spectral divergence rates collapse will be examined in Sect. 7.3. We note that the above theorem has been shown in Ref. [129], while some special cases have been discussed in Ref. [138] for i.i.d. states and in Ref. [139] for unital maps.

We next consider asymptotic (ordinary) majorization as a special case of d-majorization. We first define the following.

Definition 7.6 (*Spectral entropy rate*) Let $\widehat{P} := \{\hat{\rho}_n\}_{n\in\mathbb{N}}$ with $\hat{\rho}_n \in \mathscr{S}(\mathscr{H}^{\otimes n})$ be a sequence of quantum states, and let $\widehat{ID} := \{\hat{I}^{\otimes n}/d^n\}_{n\in\mathbb{N}}$ be the sequence of the maximally mixed states. We then define

$$\underline{S}(\widehat{P}) := \ln d - \overline{S}(\widehat{P}\|\widehat{ID}), \tag{7.26}$$

$$\overline{S}(\widehat{P}) := \ln d - \underline{S}(\widehat{P}\|\widehat{ID}). \tag{7.27}$$

Here, we note that "upper" and "lower" are the opposite to those of the divergence case. We can define $S_\infty^\varepsilon(\hat{\rho})$ and $S_0^\varepsilon(\hat{\rho})$ correspondingly. Under these definitions, the foregoing argument can reduce to the ordinary majorization case. In addition, we note the following theorem without a proof:

Proposition 7.1 (Lemma 3 of Ref. [62])

$$\underline{S}(\widehat{P}) \leq \liminf_{n \to \infty} \frac{1}{n} S_1(\hat{\rho}_n) \leq \limsup_{n \to \infty} \frac{1}{n} S_1(\hat{\rho}_n) \leq \overline{S}(\widehat{P}). \tag{7.28}$$

Thus, if $\underline{S}(\widehat{P}) = \overline{S}(\widehat{P})$ is satisfied, the von Neumann entropy rate $S_1(\widehat{P})$ exists and

$$\underline{S}(\widehat{P}) = \overline{S}(\widehat{P}) = S_1(\widehat{P}) \tag{7.29}$$

holds.

On the other hand, the above proposition with the spectral entropy rates replaced by the spectral divergence rates is not valid. In fact, there exists a counterexample even in the classical case (see Sect. IV.4.3. of Ref. [129]), for which $\overline{S}(\widehat{P}\|\widehat{\Sigma}) = \underline{S}(\widehat{P}\|\widehat{\Sigma}) =: S(\widehat{P}\|\widehat{\Sigma})$ holds and $S_1(\widehat{P}\|\widehat{\Sigma})$ exists, but $S(\widehat{P}\|\widehat{\Sigma}) \neq S_1(\widehat{P}\|\widehat{\Sigma})$.

We finally note the following general property of the spectral divergence rates. This plays an important role in ergodicity-broken systems as discussed at the end of the next section.

Proposition 7.2 (Proposition 11 of [129]) *Suppose that $\widehat{P} := \{\hat{\rho}_n\}$ is given by the mixture of $\widehat{P}^{(k)} := \{\hat{\rho}_n^{(k)}\}$ with probability $r_k > 0$ ($k = 1, 2, \ldots, K < \infty$), i.e., $\hat{\rho}_n = \sum_k r_k \hat{\rho}_n^{(k)}$ with r_k being independent of n. Then,*

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) = \min_k \{\underline{S}(\widehat{P}^{(k)}\|\widehat{\Sigma})\}, \ \overline{S}(\widehat{P}\|\widehat{\Sigma}) = \max_k \{\overline{S}(\widehat{P}^{(k)}\|\widehat{\Sigma})\}. \tag{7.30}$$

7.3　Quantum Asymptotic Equipartition Property

We next examine under what conditions the assumption of Theorem 7.6 is satisfied and a (almost) complete monotone emerges. That is, a goal of this section is to establish a condition that the upper and lower spectral divergence rates collapse to a single value.

The asymptotic equipartition property (AEP) plays a central role. In the classical case [3], the AEP characterizes typical asymptotic (large deviation) behavior of probability distributions, and is closely related to various concepts in classical probability theory: the law of large numbers, the ergodic theorem, and the Shannon-McMillan theorem. The AEP states that almost all events have almost the same probabilities, as represented by "equipartition" (see also Proposition C.1 in Appendix C). Therefore, if the AEP is satisfied, the Lorenz curve asymptotically consists of an almost straight line (and a horizontal line), implying the collapse of the upper and lower spectral entropy rates. Furthermore, we can consider the relative version of the AEP (see also Proposition C.2 in Appendix C), which is equivalent to the collapse of the upper and lower spectral divergence rates to a single value given by the KL divergence rate.

The relative AEP is also equivalent to the Stein's lemma for hypothesis testing. See Appendix C for details of the classical case.

In the following, we consider the quantum version of the (relative) AEP, which is equivalent to the quantum Stein's lemma for quantum hypothesis testing (see also Appendix B). Then, the goal of this section is rephrased as follows: Clarify under what conditions the quantum relative AEP holds.

We first note the i.i.d. case, where it is well-known that the relative AEP holds and the upper and lower spectral divergence rates collapse.

Theorem 7.7 (Theorem 2 of Ref. [61]) *If $\widehat{P} := \{\hat{\rho}^{\otimes n}\}$ and $\widehat{\Sigma} := \{\hat{\sigma}^{\otimes n}\}$ are both i.i.d.,*

$$\overline{S}(\widehat{P}\|\widehat{\Sigma}) = \underline{S}(\widehat{P}\|\widehat{\Sigma}) = S_1(\widehat{P}\|\widehat{\Sigma}). \tag{7.31}$$

To go beyond the i.i.d. situation, *ergodicity* plays a significant role, with which our answer to the above-mentioned question is as follows: If $\widehat{P}$ is translation invariant and ergodic and $\widehat{\Sigma}$ is the Gibbs state of a local and translation-invariant Hamiltonian, then $\overline{S}(\widehat{P}\|\widehat{\Sigma})$ and $\underline{S}(\widehat{P}\|\widehat{\Sigma})$ coincide. In the following, we clarify the definitions of these concepts.

We consider a quantum many-body system with many local spins on a lattice. Concretely, consider a quantum spin system on an infinite lattice $\mathbb{Z}^d$ in any spatial dimension ($d = 1, 2, 3, \ldots$), where a finite-dimensional Hilbert space $\mathscr{H}_i$ is attached to each lattice site $i \in \mathbb{Z}^d$. Let $\hat{\rho}$ represent a state on this infinite lattice system, and $\hat{\rho}_\Lambda$ be its reduced density operator on a bounded region $\Lambda \subset \mathbb{Z}^d$.

In this book, we do not go into the mathematical details of infinite spin systems with C^*-algebras. In this paragraph, however, let us just briefly remark on the rigorous formulation, which the readers can skip. We consider the C^*-algebra written as $\mathscr{A} := \overline{\otimes_{i \in \mathbb{Z}^d} \mathscr{A}_i}$ with $\mathscr{A}_i := \mathscr{L}(\mathscr{H}_i)$. Here, $\mathscr{A}$ is obtained as the C^*-inductive limit of the algebra of local operators $\mathscr{A}_{\text{loc}} := \cup_\Lambda \otimes_{i \in \Lambda} \mathscr{A}_i$ with $\Lambda \subset \mathbb{Z}^d$ being bounded. Then, $\hat{\rho}$ is the density operator (in the infinite-dimensional trace-class operator space) corresponding to a normal state $\rho : \mathscr{A} \to \mathbb{C}$. The expectation value of an operator $\hat{X}$ is given by $\rho(\hat{X}) = \text{tr}[\hat{\rho}\hat{X}]$. See Ref. [129] for details of the rigorous formulation, and Refs. [140–142] for the general mathematical theory of infinite spin systems.

We next introduce the shift superoperator $\mathscr{T}_i$ with $i \in \mathbb{Z}^d$, which maps any operator on site $j \in \mathbb{Z}^d$ to the same operator on site $j + i$. Then, we define translation invariance and ergodicity.

Definition 7.7 (*Translation-invariant states*) A state $\hat{\rho}$ is translation invariant, if it is invariant under the action of the shift operator: For any $i \in \mathbb{Z}^d$ and any $\hat{X} \in \mathscr{A}$,

$$\text{tr}[\hat{\rho}\hat{X}] = \text{tr}[\hat{\rho}\mathscr{T}_i(\hat{X})]. \tag{7.32}$$

Definition 7.8 (*Ergodic states*) Let $\hat{X} \in \mathscr{A}$ be a self-adjoint operator and define

$$\hat{X}_\Lambda := \frac{1}{|\Lambda|} \sum_{i \in \Lambda} \mathcal{T}_i(\hat{X}), \tag{7.33}$$

where $\Lambda \subset \mathbb{Z}^d$ is a bounded region and $|\Lambda|$ is the number of its elements. Then, a translation-invariant state $\hat{\rho}$ is ergodic, if the variance of any operator $\hat{X}_\Lambda$ of the form (7.33) vanishes in the limit of $\Lambda \to \mathbb{Z}^d$. Here, we take Λ as a hypercube to specify the meaning of this limit.

This definition implies that any macroscopic (i.e., extensive) observable of the form (7.33) (e.g., the total magnetization) does not exhibit macroscopic fluctuations, that is, any macroscopic observable has a definite value in the thermodynamic limit. We note that this is "spatial" ergodicity in the sense that the spatial average equals the ensemble average, while in standard statistical mechanics temporal ergodicity often plays significant roles. The above definition also implies that the state is in a "pure thermodynamic phase" without any phase mixture, which is formalized as follows.

Proposition 7.3 *A state $\hat{\rho}$ is translation invariant and ergodic, if and only if it is an extreme point of the set of translation-invariant states.*

In literature, the above expression of ergodicity is often adopted as a definition of it. See, for example, Theorem 6.3.3, Proposition 6.3.5, and Lemma 6.5.1 of Ref. [142] for other equivalent expressions.

We next formulate the Hamiltonian of the system.

Definition 7.9 (*Local and translation-invariant Hamiltonians*) A Hamiltonian of the system is local, if its truncation on any bounded region $\Lambda \subset \mathbb{Z}^d$, written as $\hat{H}_\Lambda$, is given of the form

$$\hat{H}_\Lambda = \sum_{i \in \Lambda} \hat{h}_i, \tag{7.34}$$

where $\hat{h}_i$ is a Hermitian operator acting on a bounded region around $i \in \mathbb{Z}^d$. Moreover, a local Hamiltonian is translation invariant, if $\hat{h}_i = \mathcal{T}_i(\hat{h}_0)$ holds.

The corresponding truncated Gibbs state is given by $\hat{\rho}_\Lambda^G := \exp(\beta(F_\Lambda - \hat{H}_\Lambda))$ with $F_\Lambda := -\beta^{-1}\mathrm{tr}[e^{-\beta\hat{H}_\Lambda}]$. Note that if the global Gibbs state (strictly speaking, the Kubo-Martin-Schwinger (KMS) state) is unique, it is obtained by the limit of $\hat{\rho}_\Lambda^G$ in the weak-$*$ (or ultraweak) topology (cf. Proposition 6.2.15 of Ref. [141]).

It is known that the KMS state of a local and translation-invariant Hamiltonian is ergodic, for any $\beta > 0$ in one dimension [143], and for sufficiently small β (i.e., for sufficiently high temperature) in higher dimensions (see, e.g., Ref. [144] and references therein).

We note that i.i.d. is ergodic. In addition, an i.i.d. state is regarded as the Gibbs state of a non-interacting Hamiltonian. On the other hand, the foregoing general formulation includes non-i.i.d. cases, in the sense that general ergodic states can be non-i.i.d. and Hamiltonians can be interacting.

We consider our setup of the asymptotic majorization in the previous section along with the foregoing infinite-lattice formulation. Let $\Lambda_n \subset \mathbb{Z}^d$ be a bounded region with $|\Lambda_n| = n$. We consider sequences $\widehat{P} = \{\hat{\rho}_n\}_{n=1}^{\infty}$ and $\widehat{\Sigma} = \{\hat{\sigma}_n\}_{n=1}^{\infty}$, where $\hat{\rho}_n$ is the reduced state on Λ_n of an ergodic state $\hat{\rho}$, and $\hat{\sigma}_n := \rho_{\Lambda_n}^{\mathrm{G}}$ is the truncated Gibbs state of a local and translation-invariant Hamiltonian.

Before going to the relative quantum AEP with $\widehat{P}$ and $\widehat{\Sigma}$, we consider the quantum version of the ordinary AEP of $\widehat{P}$, which is referred to as the quantum Shannon-McMillan theorem [145–147].

Theorem 7.8 (Quantum Shannon-McMillan Theorem, Theorem 1.4 of Ref. [147]) *Suppose that $\widehat{P}$ is ergodic. Then, for any $\varepsilon > 0$, there exists a sequence of projectors $\hat{\Pi}_{\widehat{P},n}^{\varepsilon}$ (called typical projectors) that satisfy, for sufficiently large n,*

$$e^{-n(s+\varepsilon)}\hat{\Pi}_{\widehat{P},n}^{\varepsilon} \leq \hat{\Pi}_{\widehat{P},n}^{\varepsilon}\hat{\rho}_n\hat{\Pi}_{\widehat{P},n}^{\varepsilon} \leq e^{-n(s-\varepsilon)}\hat{\Pi}_{\widehat{P},n}^{\varepsilon}, \tag{7.35}$$

$$e^{n(s-\varepsilon)} < \mathrm{tr}[\hat{\Pi}_{\widehat{P},n}^{\varepsilon}] < e^{n(s+\varepsilon)}, \tag{7.36}$$

$$\lim_{n\to\infty} \mathrm{tr}[\hat{\Pi}_{\widehat{P},n}^{\varepsilon}\hat{\rho}_n] = 1, \tag{7.37}$$

where $s := S_1(\widehat{P})$.

See Proposition C.1 in Appendix C for the classical counterpart. The following is a consequence of the quantum Shannon-McMillan theorem and is itself regarded as a representation of the quantum AEP.

Theorem 7.9 (Quantum AEP, Theorem 7 of [148]) *If $\widehat{P}$ is translation invariant and ergodic,*

$$\underline{S}(\widehat{P}) = \overline{S}(\widehat{P}) = S_1(\widehat{P}). \tag{7.38}$$

We now consider the relative version of the above theorem, which is a goal of this section. It is a quantum version of the relative Shannon-McMillan theorem or the relative AEP. We note that it is further equivalent to the quantum Stein's lemma for non-i.i.d. situations (see also Appendix B).

Theorem 7.10 (Relative quantum AEP, Theorem 3 of [129]) *If $\widehat{P}$ is translation invariant and ergodic, and $\widehat{\Sigma}$ is the Gibbs state of a local and translation invariant Hamiltonian, then*

$$\overline{S}(\widehat{P}\|\widehat{\Sigma}) = \underline{S}(\widehat{P}\|\widehat{\Sigma}) = S_1(\widehat{P}\|\widehat{\Sigma}). \tag{7.39}$$

Combining Theorem 7.10 with Theorem 7.6, we find that the KL divergence rate provides a (almost) necessary and sufficient condition for asymptotic state conversion:

Corollary 7.1 *Suppose that $\widehat{P}$ is translation invariant and ergodic, and $\widehat{\Sigma}$ is the Gibbs state of a local and translation-invariant Hamiltonian. Then:*

(a) $(\widehat{P}', \widehat{\Sigma}') \prec^{\mathrm{a}} (\widehat{P}, \widehat{\Sigma})$ *implies* $S_1(\widehat{P}'\|\widehat{\Sigma}') \leq S_1(\widehat{P}\|\widehat{\Sigma})$.
(b) $S_1(\widehat{P}'\|\widehat{\Sigma}') < S_1(\widehat{P}\|\widehat{\Sigma})$ *implies* $(\widehat{P}', \widehat{\Sigma}') \prec^{\mathrm{a}} (\widehat{P}, \widehat{\Sigma})$.

This corollary implies that under the above conditions, the KL divergence rate serves as a (almost) complete monotone in the asymptotic limit. This enables us to introduce a single thermodynamic potential that fully characterizes macroscopic state conversion, as will be discussed in Sect. 8.7.

We here remark on the history of the relative quantum AEP (or equivalently, the quantum Stein's lemma). Hiai and Petz [149] provided a half of the proof for a (completely) ergodic quantum state with respect to an i.i.d. state. Ogawa and Nagaoka [150] completed the proof for the case that both states are i.i.d.. Bjelaković and Siegmund-Schultze [151] proved a more general form for an ergodic (not necessarily completely ergodic) state with respect to an i.i.d. state (corresponding to the Gibbs state of a non-interacting Hamiltonian). Faist, Sagawa, et al. [129, 152] further generalized it for an ergodic state with respect to the Gibbs state of an interacting, local and translation-invariant Hamiltonian (Theorem 7.10 above).

If ergodicity is broken, the upper and lower divergence rates do not coincide in general. In particular, we consider a mixture of different ergodic states, where the entire state is not ergodic. With the notation of Proposition 7.2 in the previous section, let $\widehat{P}^{(k)}$ be a translation-invariant ergodic state and $\widehat{\Sigma}$ be the Gibbs state of a local and translation-invariant Hamiltonian. Then, Eq. (7.30) reduces to

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) = \min_k \{S_1(\widehat{P}^{(k)}\|\widehat{\Sigma})\}, \ \ \overline{S}(\widehat{P}\|\widehat{\Sigma}) = \max_k \{S_1(\widehat{P}^{(k)}\|\widehat{\Sigma})\}. \tag{7.40}$$

Thus, except for the special case that all of $S_1(\widehat{P}^{(k)}\|\widehat{\Sigma})$'s happen to take the same value, state convertibility can no longer be characterized by a single complete monotone (in particular by the KL divergence rate).

We note that, in the above setup,

$$S_1(\widehat{P}\|\widehat{\Sigma}) = \sum_k r_k S_1(\widehat{P}^{(k)}\|\widehat{\Sigma}). \tag{7.41}$$

In fact, $-\mathrm{tr}[\hat{\rho}_n^{(k)} \ln \hat{\sigma}_n]$ is additive with respect to k. Also because $\sum_k r_k S_1(\hat{\rho}_n^{(k)}) \leq S_1(\hat{\rho}_n) \leq \sum_k r_k S_1(\hat{\rho}_n^{(k)}) + S_1(r)$ [Eqs. (5.27) and (5.29)], we obtain $S_1(\widehat{P}) = \sum_k r_k S_1(\widehat{P}^{(k)})$. From Eq. (7.40) and Eq. (7.41), we have

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) \leq S_1(\widehat{P}\|\widehat{\Sigma}) \leq \overline{S}(\widehat{P}\|\widehat{\Sigma}). \tag{7.42}$$

However, as mentioned in Sect. 7.2, inequality (7.42) is not necessarily true in general, even in the classical case [129].

7.4 More on the Smoothing

As a side remark, we discuss some other ways of smoothing. First, the trace distance can be replaced by another distance for the definition of neighborhood $B^\varepsilon(\hat{\rho})$. A commonly used one is the purified distance defined as [45, 153]

$$P(\hat{\rho}, \hat{\sigma}) := \sqrt{1 - F(\hat{\rho}, \hat{\sigma})^2}, \tag{7.43}$$

where the fidelity is given by

$$F(\hat{\rho}, \hat{\sigma}) := \mathrm{tr}[(\hat{\rho}^{1/2}\hat{\sigma}\hat{\rho}^{1/2})^{1/2}] = \|\hat{\rho}^{1/2}\hat{\sigma}^{1/2}\|_1. \tag{7.44}$$

Here, $F(\hat{\rho}, \hat{\sigma}) = F(\hat{\sigma}, \hat{\rho})$ and thus $P(\hat{\rho}, \hat{\sigma}) = P(\hat{\sigma}, \hat{\rho})$. We note that $F(\hat{\rho}, \hat{\sigma})$ is sometimes called the square root fidelity and $F(\hat{\rho}, \hat{\sigma})^2$ is called the fidelity. The fidelity satisfies the monotonicity: it does not decrease under CPTP map $\mathscr{E}$:

$$F(\hat{\rho}, \hat{\sigma}) \le F(\mathscr{E}(\hat{\rho}), \mathscr{E}(\hat{\sigma})) \tag{7.45}$$

(see Theorem 9.6 of Ref. [4] for the proof). We note that the monotonicity is also satisfied under positive and TP maps (Corollary A.5 of Ref. [154]). Correspondingly, the purified distance satisfies the monotonicity $P(\hat{\rho}, \hat{\sigma}) \ge P(\mathscr{E}(\hat{\rho}), \mathscr{E}(\hat{\sigma}))$. The purified distance and the trace distance are related with each other as

$$D(\hat{\rho}, \hat{\sigma}) \le P(\hat{\rho}, \hat{\sigma}) \le \sqrt{2D(\hat{\rho}, \hat{\sigma})}, \tag{7.46}$$

and thus they give asymptotically the same smoothing of the 0- and ∞-divergences.

Meanwhile, it is sometimes convenient to define the lower spectral divergence rate in an alternative manner based on a Rényi divergence other than $S_0(\hat{\rho}\|\hat{\sigma})$ [153]. It is the sandwiched Rényi 1/2-divergence defined by the Fidelity as

$$S_{1/2}(\hat{\rho}\|\hat{\sigma}) := -2\ln F(\hat{\rho}, \hat{\sigma}). \tag{7.47}$$

See Appendix A.3 for details of the sandwiched Rényi divergences. We then define the smooth 1/2-divergence as

$$S_{1/2}^\varepsilon(\hat{\rho}\|\hat{\sigma}) := \max_{\hat{\tau} \in B^\varepsilon(\hat{\rho})} S_{1/2}(\hat{\tau}\|\hat{\sigma}). \tag{7.48}$$

The following proposition relates $S_{1/2}^\varepsilon(\hat{\rho}\|\hat{\sigma})$ to $S_0^\varepsilon(\hat{\rho}\|\hat{\sigma})$.

Proposition 7.4 (Proposition 2 of [129])

$$S_{1/2}^\varepsilon(\hat{\rho}\|\hat{\sigma}) - 6\ln\frac{3}{\varepsilon} \le S_0^{2\varepsilon}(\hat{\rho}\|\hat{\sigma}) \le S_{1/2}^{2\varepsilon}(\hat{\rho}\|\hat{\sigma}) - \ln(1 - 2\varepsilon). \tag{7.49}$$

We note that the above inequality is slightly different from that of Ref. [129], because the way of smoothing is slightly different (see Lemma 7.1 later). From this proposition, we see that $S_{1/2}(\hat{\rho}\|\hat{\sigma})$ has the same information as $S_0(\hat{\rho}\|\hat{\sigma})$ in the asymptotic limit. In this sense, $S_{1/2}(\hat{\rho}\|\hat{\sigma})$ is sometimes called the min divergence.

In the foregoing formulation of smooth divergences, we only adopted normalized states for the definition of neighborhood $B^\varepsilon(\hat{\rho})$. On the other hand, we can also adopt subnormalized states for smoothing as in Refs. [64, 65]. The corresponding neighborhood is given by

$$\bar{B}^\varepsilon(\hat{\rho}) := \{\hat{\tau} : \ \bar{D}(\hat{\tau}, \hat{\rho}) \le \varepsilon, \operatorname{tr}[\hat{\tau}] \le 1, \hat{\tau} \ge 0\}, \tag{7.50}$$

where the trace distance is replaced by a generalized one:

$$\bar{D}(\hat{\tau}, \hat{\rho}) := \frac{1}{2}\|\hat{\tau} - \hat{\rho}\|_1 + \frac{1}{2}|\operatorname{tr}[\hat{\tau}] - \operatorname{tr}[\hat{\rho}]|. \tag{7.51}$$

Suppose that $\hat{\rho}$ is normalized and $\hat{\tau}$ is subnormalized. Then, $\|\hat{\tau}/\operatorname{tr}[\hat{\tau}] - \hat{\rho}\|_1 \le \|\hat{\tau}/\operatorname{tr}[\hat{\tau}] - \hat{\tau}\|_1 + \|\hat{\tau} - \hat{\rho}\|_1 = |1 - \operatorname{tr}[\hat{\tau}]| + \|\hat{\tau} - \hat{\rho}\|_1$, and thus $D(\hat{\tau}/\operatorname{tr}[\hat{\tau}], \hat{\rho}) \le \bar{D}(\hat{\tau}, \hat{\rho})$. Also, by noting that $\|\hat{\rho}\|_1 - \|\hat{\tau}\|_1 \le \|\hat{\tau} - \hat{\rho}\|_1$, $\bar{D}(\hat{\tau}, \hat{\rho}) \le \varepsilon$ implies $\operatorname{tr}[\hat{\tau}] \ge 1 - \varepsilon$.

We define $\bar{S}^\varepsilon_\alpha(\hat{\rho}\|\hat{\sigma})$ with $\alpha = 0, 1/2, \infty$ as the smooth α-divergences with subnormalized neighborhood (7.50). (We note that $S_\alpha(\hat{\rho}\|\hat{\sigma})$ can be defined for subnormalized $\hat{\rho}$.) Then, the two ways of smoothing with normalized or subnormalized states are essentially equivalent [45], that is, $\bar{S}^\varepsilon_\alpha(\hat{\rho}\|\hat{\sigma})$ and $S^\varepsilon_\alpha(\hat{\rho}\|\hat{\sigma})$ give the same asymptotic limit.

Lemma 7.1 *For normalized $\hat{\rho}$, $\hat{\sigma}$,*

(a) $\bar{S}^\varepsilon_0(\hat{\rho}\|\hat{\sigma}) = S^\varepsilon_0(\hat{\rho}\|\hat{\sigma})$.
(b) $S^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma}) \le \bar{S}^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma}) \le S^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma}) - \ln(1 - \varepsilon)$.
(c) $\bar{S}^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma}) \le S^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma}) \le \bar{S}^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma}) - \ln(1 - \varepsilon)$.

Proof (a) $\bar{S}^\varepsilon_0(\hat{\rho}\|\hat{\sigma}) \ge S^\varepsilon_0(\hat{\rho}\|\hat{\sigma})$ is obvious from the definition. On the other hand, there exists $\hat{\tau} \in \bar{B}^\varepsilon(\hat{\rho})$ such that $\bar{S}^\varepsilon_0(\hat{\rho}\|\hat{\sigma}) = S_0(\hat{\tau}\|\hat{\sigma})$. Here, $\hat{\tau}/\operatorname{tr}[\hat{\tau}]$ is normalized and has the same support as $\hat{\tau}$, and is a candidate for maximization of $S^\varepsilon_0(\hat{\rho}\|\hat{\sigma})$. Thus, $\bar{S}^\varepsilon_0(\hat{\rho}\|\hat{\sigma}) \le S^\varepsilon_0(\hat{\rho}\|\hat{\sigma})$.

(b) $\bar{S}^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma}) \ge S^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma})$ is obvious from the definition. On the other hand, there exists $\hat{\tau} \in \bar{B}^\varepsilon(\hat{\rho})$ such that $\bar{S}^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma}) = S_{1/2}(\hat{\tau}\|\hat{\sigma})$. Since $\hat{\tau}/\operatorname{tr}[\hat{\tau}]$ is a candidate for maximization of $S^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma})$, we have $S^\varepsilon_{1/2}(\hat{\rho}\|\hat{\sigma}) \ge S_{1/2}(\hat{\tau}/\operatorname{tr}[\hat{\tau}]\|\hat{\sigma}) \ge S_{1/2}(\hat{\tau}\|\hat{\sigma}) + \ln(1 - \varepsilon)$. Thus we have the right inequality.

(c) $\bar{S}^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma}) \le S^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma})$ is obvious from the definition. On the other hand, there exists $\hat{\tau} \in \bar{B}^\varepsilon(\hat{\rho})$ such that $\bar{S}^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma}) = S_\infty(\hat{\tau}\|\hat{\sigma})$. Since $\hat{\tau}/\operatorname{tr}[\hat{\tau}]$ is a candidate for minimization of $S^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma})$, we have $S^\varepsilon_\infty(\hat{\rho}\|\hat{\sigma}) \le S_\infty(\hat{\tau}/\operatorname{tr}[\hat{\tau}]\|\hat{\sigma}) \le S_\infty(\hat{\tau}\|\hat{\sigma}) - \ln(1 - \varepsilon)$. Thus we have the right inequality. $\square$

Finally, we note that the fidelity for subnormalized states is given by $\bar{F}(\hat{\rho}, \hat{\sigma}) :=$ $F(\hat{\rho}, \hat{\sigma}) + \sqrt{(1 - \text{tr}[\hat{\rho}])(1 - \text{tr}[\hat{\sigma}])}$ [45, 153]. The corresponding purified distance is defined in the same manner as Eq. (7.43): $\bar{P}(\hat{\rho}, \hat{\sigma}) := \sqrt{1 - \bar{F}(\hat{\rho}, \hat{\sigma})^2}$, and satisfies the same inequality as (7.46): $\bar{D}(\hat{\rho}, \hat{\sigma}) \leq \bar{P}(\hat{\rho}, \hat{\sigma}) \leq \sqrt{2\bar{D}(\hat{\rho}, \hat{\sigma})}$.

Chapter 8
Quantum Thermodynamics

We consider a general setup of quantum thermodynamics and discuss the thermodynamic implications of quantum information theory developed in Chaps. 5, 6, and 7.

As a preliminary, we discuss the relation between the free energy and the divergences in Sect. 8.1. In Sect. 8.2, we formulate Gibbs-preserving maps and thermal operations, which are free operations in resource theory of thermodynamics. In Sect. 8.3, we introduce the work storage (weight) and the clock.

In Sect. 8.4, we consider the second law of thermodynamics at the level of ensemble average, where the work can fluctuate. This is the setup of stochastic thermodynamics and is a quantum counterpart of Sect. 4.1. A characteristic of the present formulation is that we explicitly take into account the work storage and the clock. We will also discuss the optimal protocol that saturates the average work bound.

From Sects. 8.5 to 8.8, we focus on the single-shot (or one-shot) scenario without allowing work fluctuations. This is a resource-theoretic perspective and is a quantum counterpart of Sect. 4.2. In Sect. 8.5, we derive the fundamental work bound of the single-shot scenario. We then consider the approximate and asymptotic versions of the single-shot work bound in Sects. 8.6 and 8.7, respectively. In particular, as a consequence of the quantum AEP discussed in Sect. 7.3, we see that a single complete thermodynamic potential (i.e., a complete monotone) emerges in the asymptotic limit and reduces to the KL divergence rate under certain conditions. In addition, we discuss an alternative formulation of the single-shot scenario with trace-nonincreasing maps in Sect. 8.8.

8.1 Nonequilibrium Free Energy

First of all, we clarify the definitions of the free energies of nonequilibrium quantum states. We consider a finite-dimensional quantum system with the Hamiltonian $\hat{H}$. Let $\beta \geq 0$. The Gibbs state is written as $\hat{\rho}^{\mathrm{G}} := e^{-\beta \hat{H}}/Z$, where $Z := \mathrm{tr}[e^{-\beta \hat{H}}]$ is the partition function and $F := -\beta^{-1} \ln Z$ is the equilibrium free energy.

For $\alpha = 0, 1, \infty$, we introduce the nonequilibrium α-free energy of state $\hat{\rho}$ by

$$F_\alpha(\hat{\rho}; \hat{H}) := \beta^{-1} S_\alpha(\hat{\rho} \| \hat{\rho}^G) + F. \tag{8.1}$$

Here, β equals the inverse temperature of the reference Gibbs state $\hat{\rho}^G$. Equation (8.1) can be rewritten as

$$F_\alpha(\hat{\rho}; \hat{H}) = \beta^{-1} S_\alpha(\hat{\rho} \| e^{-\beta \hat{H}}), \tag{8.2}$$

where we used the scaling property (5.47). This definition is a generalization of the classical 1-free energy defined in Eq. (4.9). Clearly,

$$F_\alpha(\hat{\rho}; \hat{H}) \geq F. \tag{8.3}$$

The 1-free energy can also be rewritten as

$$F_1(\hat{\rho}; \hat{H}) = E - \beta^{-1} S_1(\hat{\rho}), \tag{8.4}$$

where $E := \mathrm{tr}[\hat{H} \hat{\rho}]$ is the average energy. As mentioned above, $F_1(\hat{\rho}; \hat{H}) \geq F$, or equivalently

$$S_1(\hat{\rho}) \leq \beta(E - F) \tag{8.5}$$

holds, where the equality holds if and only if $\hat{\rho} = \hat{\rho}^G$. From this, we see that, under a fixed average energy E, the von Neumann entropy $S_1(\hat{\rho})$ takes the maximum value if and only if $\hat{\rho} = \hat{\rho}^G$. This guarantees that the Gibbs state is the maximum entropy state under the average energy constraint.

The Gibbs states are *free states* in resource theory of thermodynamics, which can be freely added to other states. This is based on the property called *complete passivity* [155, 156], which means that one cannot extract a positive amount of work from any number of copies of the state, $\hat{\rho}^{\otimes n}$, by any unitary operation with the initial and final Hamiltonians being the same (i.e., the operation is cyclic). It is known that a state is completely passive if and only if it is Gibbs.

8.2   Thermal Operations and Gibbs-Preserving Maps

We now generally and precisely formulate the classes of dynamics relevant to quantum thermodynamics. They are Gibbs-preserving maps and thermal operations, which are regarded as *free operations* in resource theory of thermodynamics. Correspondingly, the nonequilibrium free energies introduced in the previous section are monotones under these free operations, because of their monotonicity properties.

We consider the system S in contact with a heat bath B. Hereafter, we use subscripts S or B for states and Hamiltonians of them. Both of S and B are described by finite-dimensional Hilbert spaces. Mathematically, we do not necessarily suppose that B

is a "large" system, while physically it is natural to suppose so. In the following argument, $\beta \geq 0$ is interpreted as the inverse temperature of B.

We first reproduce the definition of the Gibbs-preserving map, which was already introduced in Eq. (4.1) of Sect. 4.1 for the classical case and Eq. (5.30) of Sect. 5.2 for the quantum case.

Definition 8.1 (*Gibbs-preserving maps*) A CPTP map $\mathscr{E}_S$ on S is a Gibbs-preserving map at $\beta \geq 0$ with Hamiltonian $\hat{H}_S$, if

$$\mathscr{E}_S(e^{-\beta \hat{H}_S}) = e^{-\beta \hat{H}_S}, \tag{8.6}$$

or equivalently, $\mathscr{E}_S(\hat{\rho}_S^G) = \hat{\rho}_S^G$.

We note that a CPTP unital map is a Gibbs-preserving map (with any $\hat{H}_S$) at $\beta = 0$ and also a Gibbs-preserving map of the trivial Hamiltonian $\hat{H}_S \propto \hat{I}_S$ (at any β) with $\hat{I}_S$ being the identity.

We next introduce another important class of thermodynamic operations, called thermal operations.

Definition 8.2 (*Thermal operations*) A CPTP map $\mathscr{E}_S$ on S is an exact thermal operation at $\beta \geq 0$ with Hamiltonian $\hat{H}_S$, if there exists a heat bath B with Hamiltonian $\hat{H}_B$ and the corresponding Gibbs state $\hat{\rho}_B^G := e^{-\beta \hat{H}_B}/Z_B$, and exists a unitary operator $\hat{U}$ acting on SB, such that

$$\mathscr{E}_S(\hat{\rho}_S) = \text{tr}_B \left[\hat{U} \hat{\rho}_S \otimes \hat{\rho}_B^G \hat{U}^\dagger \right] \tag{8.7}$$

and

$$[\hat{U}, \hat{H}_S + \hat{H}_B] = 0, \tag{8.8}$$

where $\hat{H}_S + \hat{H}_B$ is an abbreviation of $\hat{H}_S \otimes \hat{I}_B + \hat{I}_S \otimes \hat{H}_B$. Furthermore, a CPTP map $\mathscr{E}_S$ on S is a thermal operation at β with Hamiltonian $\hat{H}_S$, if there exists a sequence of exact thermal operations $\{\mathscr{E}_{S,n}\}_{n=1}^\infty$ with Hamiltonian $\hat{H}_S$ such that $\mathscr{E}_{S,n}$ converges to $\mathscr{E}_S$ in $n \to \infty$.

In the definition of thermal operation, the condition (8.8) implies that the sum of the free Hamiltonians of S and B, represented by $\hat{H}_S + \hat{H}_B$, is strictly conserved under the unitary operation $\hat{U}$. This is an idealized representation of the resource-theoretic viewpoint that the energy is a resource and is not free. If $\hat{U}$ is written as $\hat{U} = \exp(-i\hat{H}_{\text{tot}})$ with the total Hamiltonian

$$\hat{H}_{\text{tot}} = \hat{H}_S + \hat{H}_{\text{int}} + \hat{H}_B, \tag{8.9}$$

where $\hat{H}_{\text{int}}$ is the interaction Hamiltonian, then Eq. (8.8) can be expressed as

$$[\hat{H}_{\text{tot}}, \hat{H}_S + \hat{H}_B] = 0 \quad \Leftrightarrow \quad [\hat{H}_{\text{int}}, \hat{H}_S + \hat{H}_B] = 0. \tag{8.10}$$

The condition (8.8) or (8.10) is exactly satisfied for special systems such as the Jaynes-Cummings model at the resonant condition, whose Hamiltonian is given by

$$\hat{H}_{\mathrm{JC}} = \frac{1}{2}\omega\hat{\sigma}_z + g(\hat{\sigma}_+\hat{a} + \hat{\sigma}_-\hat{a}^\dagger) + \omega\hat{a}^\dagger\hat{a}. \tag{8.11}$$

Here, $\hat{\sigma}_z$, $\hat{\sigma}_\pm$ are the Pauli z and ladder operators of a two-level atom and $\hat{a}$ is the annihilation operator of a photon, and the second term on the right-hand side represents the interaction. More generally, the rotating wave approximation in the derivation of the Lindblad equation [157] approximately leads to the condition (8.8) or (8.10) (see also Ref. [35]). This is particularly relevant to the long time regime, in which one can expect that thermal operations can be realized at least approximately in a broad class of dynamics.

We note that a noisy operation (Definition 6.2) is a special thermal operation where both of the Hamiltonians, $\hat{H}_{\mathrm{S}}$ and $\hat{H}_{\mathrm{B}}$, are trivial (i.e., proportional to the identities). As a generalization of the fact that any noisy operation is unital (Proposition 6.1), we have the following.

Lemma 8.1 *If $\mathscr{E}_{\mathrm{S}}$ is a thermal operation, then $\mathscr{E}_{\mathrm{S}}$ is also a Gibbs-preserving map.*

Proof If $\mathscr{E}_{\mathrm{S}}$ is an exact thermal operation, $\mathscr{E}_{\mathrm{S}}(e^{-\beta\hat{H}_{\mathrm{S}}}) = Z_{\mathrm{B}}^{-1}\mathrm{tr}_{\mathrm{B}}[\hat{U}e^{-\beta(\hat{H}_{\mathrm{S}}+\hat{H}_{\mathrm{B}})}\hat{U}^\dagger] = Z_{\mathrm{B}}^{-1}\mathrm{tr}_{\mathrm{B}}[e^{-\beta\hat{U}(\hat{H}_{\mathrm{S}}+\hat{H}_{\mathrm{B}})\hat{U}^\dagger}] = Z_{\mathrm{B}}^{-1}\mathrm{tr}_{\mathrm{B}}[e^{-\beta(\hat{H}_{\mathrm{S}}+\hat{H}_{\mathrm{B}})}] = e^{-\beta\hat{H}_{\mathrm{S}}}$. For the non-exact case, take the limit. $\square$

The converse of the above lemma is not true: There are Gibbs-preserving maps that cannot be written as thermal operations, as is the case for noisy operations (Proposition 6.2). More strongly, we have the following proposition, which is now contrastive to the case of noisy operations (Corollary 6.1). The crucial observation is that any thermal operation cannot create coherence between energy eigenstates of $\hat{H}_{\mathrm{S}}$.

Proposition 8.1 (Main claim of [158]) *For any $\beta > 0$ and any Hamiltonian $\hat{H}_{\mathrm{S}}$ that is not proportional to the identity, there exists a pair of states $(\hat{\rho}_{\mathrm{S}}, \hat{\rho}_{\mathrm{S}}')$ such that there is a Gibbs-preserving map $\mathscr{E}_{\mathrm{S}}$ satisfying $\hat{\rho}_{\mathrm{S}}' = \mathscr{E}_{\mathrm{S}}(\hat{\rho}_{\mathrm{S}})$ but there is no such a thermal operation.*

Proof We will show that such a pair is given by $\hat{\rho}_{\mathrm{S}}$ without coherence and $\hat{\rho}_{\mathrm{S}}'$ with coherence in the eigenbasis of $\hat{H}_{\mathrm{S}}$. We first consider the case of exact thermal operations. By noting that the initial state of B is Gibbs in this case, we consider an eigenstate $|E_i^{\mathrm{S}}\rangle|E_j^{\mathrm{B}}\rangle$ of $\hat{H}_{\mathrm{S}} + \hat{H}_{\mathrm{B}}$. Then, from the condition (8.8), $\hat{U}|E_i^{\mathrm{S}}\rangle|E_j^{\mathrm{B}}\rangle$ should be given by a superposition of $|E_k^{\mathrm{S}}\rangle|E_l^{\mathrm{B}}\rangle$'s satisfying $E_i^{\mathrm{S}} + E_j^{\mathrm{B}} = E_k^{\mathrm{S}} + E_l^{\mathrm{B}}$. By tracing out the state of B, we find that the state of S cannot have any superposition of eigenstates with different energies. This property is kept unchanged even when one takes the limit for non-exact thermal operations.

On the other hand, such superposition can be created by a Gibbs-preserving map as shown below. Without loss of generality, let us focus on the situation that S

is a qubit system and the Hamiltonian is given by $\hat{H}_S = E_0|0\rangle\langle 0| + E_1|1\rangle\langle 1|$ with $E_0 < E_1$. (We omitted the superscript S in E_i.) As shown above, it is impossible to implement the transformation $|1\rangle \mapsto |+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$ by any thermal operation. (Here, $\hat{\rho}_S = |1\rangle\langle 1|$ and $\hat{\rho}'_S = |+\rangle\langle +|$.) We then construct a Gibbs-preserving map that enables this transformation. Let $p_i := e^{-\beta E_i}/Z_S$ with $Z_S := \sum_{i=0,1} e^{-\beta E_i}$. Then it is easy to show that $\hat{\sigma} := p_0^{-1}(\hat{\rho}_S^G - p_1|+\rangle\langle +|)$ is a normalized positive operator (i.e., a density operator). Then define

$$\mathscr{E}_S(\hat{\rho}_S) := \langle 0|\hat{\rho}_S|0\rangle\hat{\sigma} + \langle 1|\hat{\rho}_S|1\rangle|+\rangle\langle +|, \qquad (8.12)$$

which satisfies $\mathscr{E}_S(\hat{\rho}_S^G) = \hat{\rho}_S^G$ and $\mathscr{E}_S(|1\rangle\langle 1|) = |+\rangle\langle +|$. $\square$

We next remark on classical thermal operations.

Definition 8.3 (*Classical thermal operations*) A stochastic matrix T is an exact classical thermal operation with respect to a Hamiltonian of S, if there exists a heat bath B in finite dimension with some Hamiltonian and exists a permutation matrix P acting on the composite system such that

$$(Tp)_i = \sum_{jkl} P_{ij;kl} p_k p_{B,l}^G, \qquad (8.13)$$

where $p_{B,l}^G$ is the Gibbs state of B and P preserves the sum of the energies of S and B (that is, $P_{ij;kl}$ is nonzero only when $E_k^S + E_l^B = E_i^S + E_j^B$). Furthermore, a stochastic matrix T is a classical thermal operation with respect to a Hamiltonian of S, if there exists a sequence of exact thermal operations $\{T_n\}_{n=1}^\infty$ with the same Hamiltonian such that T_n converges to T in $n \to \infty$.

Based on the above definition, we can show that classical thermal operations are equivalent to classical Gibbs-preserving maps. A rigorous proof of this is given in Ref. [74]; see also Ref. [41].

Proposition 8.2 *For any classical stochastic map T and a given Hamiltonian, the following are equivalent.*
(i) T is a Gibbs-preserving map.
(ii) T is a thermal operation.

We also remark on the *semiclassical* case, where the initial state $\hat{\rho}_S$ and the final state $\hat{\rho}'_S$ are both (block-)diagonal in the eigenbasis of the Hamiltonian $\hat{H}_S$ (i.e., $[\hat{\rho}_S, \hat{H}_S] = [\hat{\rho}'_S, \hat{H}_S] = 0$). In the semiclassical case, thermal operations are basically equivalent to Gibbs-preserving maps, and therefore, most of the results in the subsequent sections are also valid even if the Gibbs-preserving maps are replaced with thermal operations. We notice, however, that this equivalence is not at the level of maps but only at the level of state convertibility, as in Theorem 6.1 and Corollary 6.1 (see also Proposition 6.2).

Going back to the quantum case, we consider the Lindblad equation as an illustrative example [157], which describes quantum Markovian dynamics. Let

$\hat{H}_S = \sum_i E_i |E_i\rangle\langle E_i|$ be the Hamiltonian, which is assumed to be non-degenerate for simplicity. The Lindblad equation is given by

$$\frac{d\hat{\rho}_S(t)}{dt} = -\mathrm{i}[\hat{H}_S, \hat{\rho}_S(t)] + \sum_{ij}\left(\hat{L}_{ij}\hat{\rho}_S(t)\hat{L}_{ij}^\dagger - \frac{1}{2}\{\hat{L}_{ij}^\dagger\hat{L}_{ij}, \hat{\rho}_S(t)\}\right). \qquad (8.14)$$

The solution of this gives a CPTP map: the map from $\hat{\rho}_S(0)$ to $\hat{\rho}_S(t)$ under Eq. (8.14) is a CPTP map.

We here assume that the Lindblad operators $\hat{L}_{ij}$ are labeled by (i, j) with both i and j representing the labels of the eigenvalues of $\hat{H}_S$, and that $[\hat{L}_{ij}, \hat{H}_S] = (E_j - E_i)\hat{L}_{ij}$, which represents that $\hat{L}_{ij}$ describes the quantum jump from $|E_j\rangle$ to $|E_i\rangle$. The detailed balance condition is given by

$$\hat{L}_{ji}^\dagger e^{-\beta E_i/2} = \hat{L}_{ij} e^{-\beta E_j/2}, \qquad (8.15)$$

which is a quantum counterpart of the classical detailed balance condition (4.2). It is straightforward to show that the detailed balance condition (8.15) implies that the Gibbs state $\hat{\rho}_S^G$ of $\hat{H}_S$ is the stationary solution of Eq. (8.14). In other words, the detailed balance condition (8.15) implies that the map from $\hat{\rho}_S(0)$ to $\hat{\rho}_S(t)$ is a Gibbs-preserving map for any t. See also Refs. [35, 94, 95] for quantum stochastic thermodynamics based on the Lindblad equations.

Moreover, by looking at the standard derivation of the Lindblad equation from unitary dynamics of the system and the bath (see, e.g., Ref. [157]), we can see that the Lindblad dynamics with the foregoing assumptions may be regarded as an approximate thermal operation for each infinitesimal time step [35]. In fact, as already mentioned before, the condition (8.8) is guaranteed by the rotating wave approximation. Also, we usually suppose that the bath is initially in the Gibbs state, which leads to the detailed balance condition (8.15).

8.3   Clock and Work Storage

We next introduce the work storage (or the weight) W and the clock C, which are already discussed in Sect. 4.2 for the classical case. Hereafter, we call the original part of the system just as the "system" and denote it by S. That is, the entire system consists of S, C, W, along with the heat bath B.

The work storage W has its own Hamiltonian, in which the work is stored. Let $\hat{H}_W$ be the Hamiltonian of W. At this stage, we do not specify its explicit form.

The change of the Hamiltonian of S by external driving, from $\hat{H}_S$ to $\hat{H}_S'$, can be implemented in an autonomous way by introducing the clock C such that the Hamiltonian of SC becomes time-independent [41, 159–165]. For example, suppose that the clock evolves from a pure state $|0\rangle$ to another pure state $|1\rangle$ with these states being orthogonal, and that the Hamiltonian of SC is given by the form $\hat{H}_{SC} =$

$\hat{H}_{S} \otimes |0\rangle\langle 0| + \hat{H}'_{S} \otimes |1\rangle\langle 1|$ as adopted in Ref. [41]. In this case, the Hamiltonian of S effectively changes from $\hat{H}_{S}$ to $\hat{H}'_{S}$.

The free Hamiltonian of SCW is now given by

$$\hat{H}_{SCW} = \hat{H}_{S} \otimes |0\rangle\langle 0| + \hat{H}'_{S} \otimes |1\rangle\langle 1| + \hat{H}_{W}. \tag{8.16}$$

In the subsequent sections, we will consider Gibbs-preserving maps and thermal operations with respect to the Hamiltonian (8.16). The condition of thermal operation on SCW (Eq. (8.8)) in the absence of CW is written as

$$[\hat{U}, \hat{H}_{SCW} + \hat{H}_{B}] = 0, \tag{8.17}$$

where $\hat{H}_{B}$ is the Hamiltonian of B and $\hat{U}$ is the unitary operator acting on SCWB. The condition (8.17) represents the strict energy conservation of the free Hamiltonian of SCWB.

In most of the remaining part of this chapter, we use the following notations. Let $\hat{\rho}^{G}_{SCW}, \hat{\rho}^{G}_{S}, \hat{\rho}^{G'}_{S}, \hat{\rho}^{G}_{W}$ be the Gibbs states of the Hamiltonians $\hat{H}_{SCW}, \hat{H}_{S}, \hat{H}'_{S}, \hat{H}_{W}$, and $Z_{SCW}, Z_{S}, Z'_{S}, Z_{W}$ and $F_{SCW}, F_{S}, F'_{S}, F_{W}$ be the corresponding partition functions and the free energies, respectively. Let $\Delta F_{S} := F'_{S} - F_{S}$. From Eq. (8.16), we have

$$\hat{\rho}^{G}_{SCW} = \hat{\rho}^{G}_{SC} \otimes \hat{\rho}^{G}_{W}, \tag{8.18}$$

where

$$\hat{\rho}^{G}_{SC} = \frac{Z_{S}}{Z_{S} + Z'_{S}} \hat{\rho}^{G}_{S} \otimes |0\rangle\langle 0| + \frac{Z'_{S}}{Z_{S} + Z'_{S}} \hat{\rho}^{G'}_{S} \otimes |1\rangle\langle 1|. \tag{8.19}$$

Concrete constructions of the work storage and the clock are discussed in Refs. [159–163, 165]. Here we illustrate the function of the clock by a (almost trivial) toy example. For simplicity, we only consider the composite SC. Let $\hat{H}_{SC} = \hat{\sigma}_{z} \otimes |0\rangle\langle 0| + \hat{\sigma}_{x} \otimes |1\rangle\langle 1|$ be the Hamiltonian and $\hat{U} = \hat{H} \otimes \hat{\sigma}_{x}$ be the unitary operation, where $\hat{\sigma}_{i}$ $(i = x, y, z)$ is the Pauli operator with $\{|0\rangle, |1\rangle\}$ being the eigenbasis of $\hat{\sigma}_{z}$, and $\hat{H}$ is the Hadamard operator satisfying $\hat{H}^{\dagger} = \hat{H}, \hat{H}\hat{\sigma}_{x}\hat{H} = \hat{\sigma}_{z}$, $\hat{H}\hat{\sigma}_{z}\hat{H} = \hat{\sigma}_{x}$. Then, $\hat{U}^{\dagger}\hat{H}_{SC}\hat{U} = \hat{H}_{SC}$, or equivalently $[\hat{U}, \hat{H}_{SC}] = 0$, holds. For any input state of the form $|\psi\rangle|0\rangle$, we have $\hat{U}|\psi\rangle|0\rangle = (\hat{H}|\psi\rangle)|1\rangle$. Thus in this case, the clock works perfectly for all initial states of S, while satisfying condition (8.17).

In general, however, not all the initial states of S are consistent with the evolution of C from $|0\rangle$ to $|1\rangle$, under the constraint of the strict energy conservation (8.17). This is because $\hat{H}_{S}$ and $\hat{H}'_{S}$ (or $\hat{H}_{S} + \hat{H}_{W} + \hat{H}_{B}$ and $\hat{H}'_{S} + \hat{H}_{W} + \hat{H}_{B}$) can have different energy spectra. Thus, in the formulation of the subsequent sections, we only require that the clock evolves from $|0\rangle$ to $|1\rangle$ for a *given* initial state of S.

Suppose that the unitary operator is generated by a time-independent Hamiltonian as $\hat{U} = \exp(-i\hat{H}_{tot}t)$, where $\hat{H}_{tot} = \hat{H}_{S} \otimes |0\rangle\langle 0| + \hat{H}'_{S} \otimes |1\rangle\langle 1| + \hat{H}_{W} + \hat{H}_{B} + \hat{H}_{int}$ is the total Hamiltonian and $\hat{H}_{int}$ is the interaction Hamiltonian. Here, $\hat{H}_{int}$ represents not only the interaction between S and B, but also the interaction between S and W,

and moreover, driving of C. The condition of thermal operation on SCW (8.17) is rewritten as $[\hat{H}_{\text{tot}}, \hat{H}_{\text{SCW}} + \hat{H}_{\text{B}}] = 0$, or equivalently $[\hat{H}_{\text{int}}, \hat{H}_{\text{SCW}} + \hat{H}_{\text{B}}] = 0$.

It is not trivial whether a desired unitary operation $\hat{U}$ can be implemented by a time-independent Hamiltonian $\hat{H}_{\text{tot}}$ in a physically feasible form without any additional thermodynamic cost. A way to implement such a unitary is that we introduce a "hyper-clock", named C', in addition to SCWB, as discussed in Ref. [161] (C' is called just a "clock" there). In that case, the total Hamiltonian is given by $\hat{H}_{\text{tot}} = \hat{H}_{\text{SCW}} + \hat{H}_{\text{B}} + \hat{H}'_{\text{int}} + \hat{H}_{\text{C'}}$, where $\hat{H}'_{\text{int}}$ describes the interaction between SCWB and C'. By imposing $[\hat{H}'_{\text{int}}, \hat{H}_{\text{SCW}} + \hat{H}_{\text{B}}] = 0$, the strict energy conservation (8.17) is guaranteed for $\hat{U} = \exp(-i\hat{H}_{\text{tot}}t)$. We note that a general bound of the quantum coherence cost to implement such (hyper-)clocks has been obtained in Ref. [166].

8.4 Average Work Bound

In this section, we consider the work bound at the level of ensemble average for the situation where work is a stochastic quantity [159, 160, 167], and derive the average work bound in two slightly different ways. This setup is contrastive to the single-shot case discussed in the subsequent sections and is relevant to stochastic thermodynamics in the quantum regime [33–35]. This also complements the arguments in Sects. 4.1 and 5.2. See also Ref. [34] for a more conventional approach to the average work bound.

First, we derive the second law by considering the Gibbs-preserving map of SCW. We adopt the setup and the notations of Sect. 8.3; Specifically, the Hamiltonian $\hat{H}_{\text{SCW}}$ is given by Eq. (8.16). We suppose that the initial state of SCW is given by $\hat{\rho} := \hat{\rho}_{\text{S}} \otimes |0\rangle\langle 0| \otimes \hat{\rho}_{\text{W}}$. (We omit the subscript SCW from $\hat{\rho}$.) At this stage, we do not make any assumption about the form of $\hat{H}_{\text{W}}$ and $\hat{\rho}_{\text{W}}$.

The final state after the Gibbs-preserving map of SCW, written as $\mathscr{E}_{\text{SCW}}$, is given by $\hat{\rho}' := \mathscr{E}_{\text{SCW}}(\hat{\rho})$. For a given initial state $\hat{\rho}_{\text{S}}$ of S, we assume that $\hat{\rho}'$ is given by the product of the final state of SW, written as $\hat{\rho}'_{\text{SW}}$, and the final pure state of C given by $|1\rangle\langle 1|$. This implies that the clock works perfectly for the given initial state. Let $\hat{\rho}'_{\text{S}} := \text{tr}_{\text{W}}[\hat{\rho}'_{\text{SW}}]$ and $\hat{\rho}'_{\text{W}} := \text{tr}_{\text{S}}[\hat{\rho}'_{\text{SW}}]$. Then, the average work is defined as

$$W := \text{tr}[(\hat{\rho}_{\text{W}} - \hat{\rho}'_{\text{W}})\hat{H}_{\text{W}}]. \tag{8.20}$$

We here make an assumption that $S_1(\hat{\rho}_{\text{W}}) = S_1(\hat{\rho}'_{\text{W}})$, which excludes the possibility of "cheating" that W works as an entropic source. This assumption can be satisfied at least approximately; For example, it has been shown in Ref. [160] that the entropy change can be arbitrarily small by appropriately designing the interaction and the initial state of W.

We now derive the work bound. By noting that the Gibbs states of SCW and SC are given by Eqs. (8.18) and (8.19), we have

$$S_1(\hat{\rho}\|\hat{\rho}_{SCW}^G) = S_1(\hat{\rho}_S\|\hat{\rho}_S^G) + \beta F_S + \ln(Z_S + Z_S') + S_1(\hat{\rho}_W\|\hat{\rho}_W^G), \qquad (8.21)$$

and

$$S_1(\hat{\rho}'\|\hat{\rho}_{SCW}^G) = S_1(\hat{\rho}_{SW}'\|\hat{\rho}_{SW}^G{}') + \beta F_S' + \ln(Z_S + Z_S'), \qquad (8.22)$$

where we used the scaling property (5.47) and defined $\hat{\rho}_{SW}^G{}' := \hat{\rho}_S^{G'} \otimes \hat{\rho}_W^G$. From the subadditivity (5.19) of the von Neumann entropy, we have

$$S_1(\hat{\rho}_{SW}'\|\hat{\rho}_{SW}^G{}') = -S_1(\hat{\rho}_{SW}') - \text{tr}[\hat{\rho}_S' \ln \hat{\rho}_S^{G'}] - \text{tr}[\hat{\rho}_W' \ln \hat{\rho}_W^G] \qquad (8.23)$$

$$\geq -S_1(\hat{\rho}_S') - S_1(\hat{\rho}_W') - \text{tr}[\hat{\rho}_S' \ln \hat{\rho}_S^{G'}] - \text{tr}[\hat{\rho}_W' \ln \hat{\rho}_W^G] \quad (8.24)$$

$$= S_1(\hat{\rho}_S'\|\hat{\rho}_S^{G'}) + S_1(\hat{\rho}_W'\|\hat{\rho}_W^G). \qquad (8.25)$$

Also, $S_1(\hat{\rho}_W\|\hat{\rho}_W^G) - S_1(\hat{\rho}_W'\|\hat{\rho}_W^G) = \beta W$ holds from the assumption $S_1(\hat{\rho}_W) = S_1(\hat{\rho}_W')$. By combining these relations, the monotonicity of the quantum KL divergence, i.e., $S_1(\hat{\rho}\|\hat{\rho}_{SCW}^G) \geq S_1(\hat{\rho}'\|\hat{\rho}_{SCW}^G)$, leads to

$$\beta(W - \Delta F_S) \geq S_1(\hat{\rho}_S'\|\hat{\rho}_S^{G'}) - S_1(\hat{\rho}_S\|\hat{\rho}_S^G). \qquad (8.26)$$

This is the bound of the average work in the present setup. By using the nonequilibrium 1-free energy introduced in Eq. (8.1), we rewrite the obtained work bound as

$$W \geq F_1(\hat{\rho}_S'; \hat{H}_S') - F_1(\hat{\rho}_S; \hat{H}_S). \qquad (8.27)$$

This is a quantum extension of inequality (4.10), while we did not explicitly take into account a clock and a work storage in the classical case.

It is reasonable to define the average heat absorbed by the system through the first law of thermodynamics:

$$Q := \text{tr}[\hat{H}_S'\hat{\rho}_S'] - \text{tr}[\hat{H}_S\hat{\rho}_S] - W. \qquad (8.28)$$

Then, we can rewrite the second law (8.27) in the form of the Clausius inequality:

$$S_1(\hat{\rho}_S') - S_1(\hat{\rho}_S) \geq \beta Q, \qquad (8.29)$$

which is a quantum counterpart of inequality (4.4).

To strictly justify the definition of heat (8.28), we need an additional assumption about the energy balance between SCW and the bath B. A strong assumption is the strict energy conservation (8.17) that is satisfied if $\mathscr{E}_{SCW}$ is a thermal operation, with which the heat (8.28) exactly coincides with the energy change in B. On the other hand, a weaker assumption is often enough to justify Eq. (8.28): the average energy conservation, which means that the expectation value of the free Hamiltonian

of SCWB, $\hat{H}_{\mathrm{SCW}} + \hat{H}_{\mathrm{B}}$, is conserved for a *given* initial state, which is the approach adopted in Ref. [160].

We next derive the work bound (8.26) in an alternative way without assuming that $S_1(\hat{\rho}_{\mathrm{W}}) = S_1(\hat{\rho}'_{\mathrm{W}})$ [161]. We explicitly take the heat bath B into account, and consider the Hamiltonian $\hat{H}_{\mathrm{SCW}} + \hat{H}_{\mathrm{B}}$ and a unitary operation $\hat{V}$ acting on SCWB. At this stage, we do not necessarily require the strict energy conservation (i.e., $[\hat{V}, H_{\mathrm{SCW}} + \hat{H}_{\mathrm{B}}] = 0$), but only require the average energy conservation at least for a given initial state $\hat{\rho}_{\mathrm{SCB}}$. Now we trace out W and define a CPTP map

$$\mathscr{E}_{\mathrm{SCB}}(\hat{\rho}_{\mathrm{SCB}}) := \mathrm{tr}_{\mathrm{W}}[\hat{V}\hat{\rho}_{\mathrm{SCB}} \otimes \hat{\rho}_{\mathrm{W}}\hat{V}^{\dagger}]. \tag{8.30}$$

A central assumption of the present approach is that $\mathscr{E}_{\mathrm{SCB}}$ is unital, which is indeed satisfied in some constructions of $\hat{V}$ discussed later. From this assumption,

$$S_1(\hat{\rho}_{\mathrm{SCB}}) \leq S_1(\hat{\rho}'_{\mathrm{SCB}}). \tag{8.31}$$

Let $\hat{\rho}_{\mathrm{SCB}} := \hat{\rho}_{\mathrm{S}} \otimes |0\rangle\langle 0| \otimes \hat{\rho}^{\mathrm{G}}_{\mathrm{B}}$, $\hat{\rho}'_{\mathrm{SCB}} := \mathscr{E}_{\mathrm{SCB}}(\hat{\rho}_{\mathrm{SCB}})$. For a given $\hat{\rho}_{\mathrm{S}}$, we assume that $\hat{\rho}'_{\mathrm{SCB}}$ is given by the product of the final state of SB, written as $\hat{\rho}'_{\mathrm{SB}}$, and the final pure state of C given by $|1\rangle\langle 1|$. We then have $S_1(\hat{\rho}_{\mathrm{SCB}}) = -S_1(\hat{\rho}_{\mathrm{S}}\|\hat{\rho}^{\mathrm{G}}_{\mathrm{S}}) - \beta(F_{\mathrm{S}} - \mathrm{tr}[\hat{H}_{\mathrm{S}}\hat{\rho}_{\mathrm{S}}]) - \beta(F_{\mathrm{B}} - \mathrm{tr}[\hat{H}_{\mathrm{B}}\hat{\rho}^{\mathrm{G}}_{\mathrm{B}}])$ and $S_1(\hat{\rho}'_{\mathrm{SCB}}) \leq S_1(\hat{\rho}'_{\mathrm{S}}) + S_1(\hat{\rho}'_{\mathrm{B}}) = -S_1(\hat{\rho}'_{\mathrm{S}}\|\hat{\rho}^{\mathrm{G}'}_{\mathrm{S}}) - \beta(F'_{\mathrm{S}} - \mathrm{tr}[\hat{H}'_{\mathrm{S}}\hat{\rho}'_{\mathrm{S}}]) - S_1(\hat{\rho}'_{\mathrm{B}}\|\hat{\rho}^{\mathrm{G}}_{\mathrm{B}}) - \beta(F_{\mathrm{B}} - \mathrm{tr}[\hat{H}_{\mathrm{B}}\hat{\rho}'_{\mathrm{B}}])$. Then, inequality (8.31) leads to

$$\beta\left(\mathrm{tr}[\hat{H}'_{\mathrm{S}}\hat{\rho}'_{\mathrm{S}}] + \mathrm{tr}[\hat{H}_{\mathrm{B}}\hat{\rho}'_{\mathrm{B}}] - \mathrm{tr}[\hat{H}_{\mathrm{S}}\hat{\rho}_{\mathrm{S}}] - \mathrm{tr}[\hat{H}_{\mathrm{B}}\hat{\rho}^{\mathrm{G}}_{\mathrm{B}}] - \Delta F_{\mathrm{S}}\right) \geq S_1(\hat{\rho}'_{\mathrm{S}}\|\hat{\rho}^{\mathrm{G}'}_{\mathrm{S}}) - S_1(\hat{\rho}_{\mathrm{S}}\|\hat{\rho}^{\mathrm{G}}_{\mathrm{S}}). \tag{8.32}$$

From the (at least average) energy conservation, we have

$$\mathrm{tr}[\hat{H}'_{\mathrm{S}}\hat{\rho}'_{\mathrm{S}}] + \mathrm{tr}[\hat{H}_{\mathrm{B}}\hat{\rho}'_{\mathrm{B}}] - \mathrm{tr}[\hat{H}_{\mathrm{S}}\hat{\rho}_{\mathrm{S}}] - \mathrm{tr}[\hat{H}_{\mathrm{B}}\hat{\rho}^{\mathrm{G}}_{\mathrm{B}}] = \mathrm{tr}[\hat{H}_{\mathrm{W}}\hat{\rho}_{\mathrm{W}}] - \mathrm{tr}[\hat{H}_{\mathrm{W}}\hat{\rho}'_{\mathrm{W}}] =: W. \tag{8.33}$$

Therefore, we obtain the same inequality as (8.26):

$$\beta(W - \Delta F_{\mathrm{S}}) \geq S_1(\hat{\rho}'_{\mathrm{S}}\|\hat{\rho}^{\mathrm{G}'}_{\mathrm{S}}) - S_1(\hat{\rho}_{\mathrm{S}}\|\hat{\rho}^{\mathrm{G}}_{\mathrm{S}}). \tag{8.34}$$

The optimal work (i.e., the equality case of the second law (8.34)) can be achieved by special protocols with certain limit [159–162]. In Refs. [159, 161], it was shown that there is an optimal protocol under the strict energy conservation, where a large amount of coherence is required in W. In Ref. [160], an optimal protocol was constructed under the average energy conservation, where the work value can be independent of the initial state of W. We will discuss these protocols in detail later.

The fact that the equality in (8.34) can be saturated implies that a necessary and sufficient condition for state conversion is given by the KL divergence (rigorously speaking, except for some technical issues), if the work fluctuation is allowed. This

is contrastive to the single-shot case discussed in Sect. 8.5, where the KL divergence does not provide such a condition and instead the Rényi 0- and ∞-divergences appear. The role of the work storage here would be reminiscent of "modestly non-exact" catalyst discussed in Sect. 3.3 where the KL divergence also appears. In fact, the work storage can be regarded as a kind of catalyst as discussed later in this section.

We now explicitly show that the equality in (8.34) can be achieved based on the protocol considered in Refs. [159–161]. We start with a general strategy and ignore the work storage W for a while. Also for simplicity, we assume that the clock C is absent and $\hat{H}_S = \hat{H}'_S$.

We will construct a unitary operator $\hat{U}$ acting on SB, which is not necessarily energy conserving. We suppose that the average energy change of SB equals the average work, and show that the average work saturates the equality in (8.34) by an optimal protocol. We will construct an energy-conserving unitary acting on SBW from $\hat{U}$ later.

Let $\hat{\rho}_S = \sum_i p_i |\varphi_i\rangle\langle\varphi_i|$, $\hat{\rho}'_S = \sum_i p'_i |\varphi'_i\rangle\langle\varphi'_i|$, and $\hat{H}_S = \sum_i E_i |E_i\rangle\langle E_i|$. Suppose that $\hat{\rho}'_S$ is positive definite. We define $\hat{\tau}_S := \sum_i p_i^{\downarrow} |E_i\rangle\langle E_i|$ and $\hat{\tau}'_S := \sum_i p_i'^{\downarrow} |E_i\rangle\langle E_i|$. We also define $p^{(n)}$ ($n = 0, 1, 2, \ldots, N$) satisfying $p^{(n)} = p^{(n)\downarrow}$ such that $p^{(0)} = p^{\downarrow}$, $D(p^{(n)}, p^{(n+1)}) = O(1/N)$ ($n = 0, 1, \ldots, N-1$), and $p^{(N)} = p'^{\downarrow}$. Let $\hat{\tau}_S^{(n)} := \sum_i p_i^{(n)} |E_i\rangle\langle E_i|$. We divide B into N subsystems, and introduce the Hamiltonian $\hat{H}_B^{(n)}$ of the nth subsystem such that its Gibbs state $\hat{\rho}_B^{G(n)}$ equals $\hat{\tau}_S^{(n)}$. The total Hamiltonian of B is given by $\hat{H}_B := \sum_{n=1}^{N} \hat{H}_B^{(n)}$. Then, we construct $\hat{U}$ acting on SB as follows.

 (i) Rotate the basis of S such that $\hat{\rho}_S$ is mapped to $\hat{\tau}_S = \hat{\tau}_S^{(0)}$.
(ii) Swap $\hat{\tau}_S^{(n-1)}$ with the nth subsystem of B with the state $\hat{\rho}_B^{G(n)} = \hat{\tau}_S^{(n)}$ ($n = 1, \ldots, N$).
(iii) Rotate the basis of S such that $\hat{\tau}'_S$ is mapped to $\hat{\rho}'_S$.

In each step of the above protocol, the change in the energy of SB equals the change in the 1-free energy of S at least approximately. In (i) and (iii), the change in the von Neumann entropy is zero, and thus the energy change of S equals the free energy change of S. In (ii), let $\Delta S_S^{(n)} := S_1(\hat{\tau}_S^{(n)}) - S_1(\hat{\tau}_S^{(n-1)}) =: -\Delta S_B^{(n)}$, $\Delta E_S^{(n)} := \mathrm{tr}[\hat{H}_S \hat{\tau}_S^{(n)}] - \mathrm{tr}[\hat{H}_S \hat{\tau}_S^{(n-1)}]$, and $\Delta E_B^{(n)} := \mathrm{tr}[\hat{H}_B^{(n)} \hat{\tau}_S^{(n-1)}] - \mathrm{tr}[\hat{H}_B^{(n)} \hat{\tau}_S^{(n)}]$ ($n = 1, 2, \ldots, N$). By noting that $\beta \Delta E_B^{(n)} - \Delta S_S^{(n)} = S_1(\tau_S^{(n-1)} \| \hat{\tau}_S^{(n)}) = O(1/N^2)$, we have $\Delta E_S^{(n)} + \Delta E_B^{(n)} = \Delta E_S^{(n)} - \beta^{-1} \Delta S_S^{(n)} + O(1/N^2)$, which is the 1-free energy change of S up to the error term $O(1/N^2)$. By summing these terms over $n = 1, \ldots, N$, the total error is given by $O(1/N)$. In the entire process (i)–(iii), therefore, the energy change of SB equals the 1-free energy change of S in the limit of $N \to \infty$.

We note that the above construction of the optimal process is slightly different from the quantum version of Fig. 4.2. In the present construction, the Hamiltonian of S is not quenched but the state of S rotates in (i) and (iii). Also, the gradual change of the state of S in (ii) is induced by a sequence of swapping processes, instead of ordinary relaxation processes.

If $\hat{H}_\text{S} \neq \hat{H}'_\text{S}$ in general, we can just insert the unitary change of the state of C from $|0\rangle$ to $|1\rangle$ into any stage of (i)-(iii) above. During this step, the energy change of S equals the 1-free energy change of S, because the state of S does not change at all.

So far, we have constructed unitary $\hat{U}$ acting on SB, which is not energy-conserving. Our next purpose is to construct an energy-conserving unitary acting on SBW that reproduces the same work value as the average energy change of SB by $\hat{U}$. This can be done by both the strict and average energy conservation protocols, as discussed below. In the following, we relabel SB just as "S".

First, we consider the case of the strict energy conservation. In line with Ref. [159], we show that if W is appropriately designed, any unitary $\hat{U}$ acting on S, which is not necessarily energy-conserving, can be implemented as a strictly energy conserving unitary $\hat{V}_{\hat{U}}$ acting on SW. Such a construction is inevitably accompanied by work fluctuations, and thus is not compatible with the single-shot scenario.

For the sake of simplicity, we first suppose that S is a two-level system spanned by $\{|0\rangle, |1\rangle\}$ with Hamiltonian $\hat{H}_\text{S} = |1\rangle\langle 1|$. We prepare W as an infinite-dimensional ladder system spanned by $\{|n\rangle\}_{n\in\mathbb{Z}}$ with Hamiltonian $\hat{H}_\text{W} = \sum_{n\in\mathbb{Z}} n|n\rangle\langle n|$. Here, we do not go into details of the technical aspect of unbounded operators of infinite-dimensional spaces. Also, a physical problem is that this Hamiltonian is not bounded from below, but this can be overcome by a slight modification of the setup [159]. We now define the unitary operator

$$\hat{V}_{\hat{U}} := \sum_{i,j=0,1} |i\rangle\langle i|\hat{U}|j\rangle\langle j| \otimes \hat{\Delta}^{j-i}, \tag{8.35}$$

where $\hat{\Delta} := \sum_{n\in\mathbb{Z}} |n+1\rangle\langle n|$ is the shift operator of the ladder. It is straightforward to check the strict energy conservation, i.e., $[\hat{V}_{\hat{U}}, \hat{H}_\text{S} + \hat{H}_\text{W}] = 0$. Also, $[\hat{V}_{\hat{U}}, \hat{\Delta}] = 0$ is obviously satisfied, implying that $\hat{V}_{\hat{U}}$ is translation invariant with respect to the ladder of W.

If S has multiple levels in general, W should consist of multiple ladders corresponding to all the energy gaps of S. The ideal limit is that W is a continuous system with $(\hat{x}, \hat{p})$ satisfying $[\hat{x}, \hat{p}] = \text{i}$ and the Hamiltonian of W is given by $\hat{H}_\text{W} = \hat{x}$. In the case of a general Hamiltonian $\hat{H}_\text{S} = \sum_i E_i|E_i\rangle\langle E_j|$, Eq. (8.35) is replaced by [161]

$$\hat{V}_{\hat{U}} := \sum_{i,j} |E_i\rangle\langle E_i|\hat{U}|E_j\rangle\langle E_j| \otimes e^{-\text{i}(E_j-E_i)\hat{p}}. \tag{8.36}$$

Going back to the two-level case, we consider the CPTP map

$$\mathscr{E}_{\hat{U}}(\hat{\rho}_\text{S}) := \text{tr}[\hat{V}_{\hat{U}}\hat{\rho}_\text{S} \otimes \hat{\rho}_\text{W}\hat{V}_{\hat{U}}^{\dagger}] \tag{8.37}$$

$$= \sum_{i,j,k,l=0,1} |i\rangle\langle i|\hat{U}|j\rangle\langle j|\hat{\rho}_\text{S}|l\rangle\langle l|\hat{U}^{\dagger}|k\rangle\langle k|\text{tr}[\hat{\Delta}^{k-l+j-i}\hat{\rho}_\text{W}]. \tag{8.38}$$

If $\mathrm{tr}[\hat{\Delta}^{k-l+j-i}\hat{\rho}_{\mathrm{W}}] = 1$ for all (i, j, k, l), we recover the unitary operation on S: $\mathscr{E}_{\hat{U}}(\hat{\rho}_{\mathrm{S}}) = \hat{U}\hat{\rho}_{\mathrm{S}}\hat{U}^{\dagger}$. This can be approximately satisfied if we choose the initial state of W with a large amount of coherence in the energy basis. In fact, if the initial state of W is given by a pure state $\sum_{n=0}^{N-1}|n\rangle/\sqrt{N}$, we have $\mathrm{tr}[\hat{\Delta}^m\hat{\rho}_{\mathrm{W}}] = (N - |m|)/N$, which goes to 1 in $N \to \infty$. In this limit, however, the variance of the energy distribution of W diverges and thus the work values can hardly be distinguished by measuring the energy of W [168].

In Ref. [159], it has been shown that the coherence of W can be used repeatedly in the following sense: Let $\hat{\rho}'_{\mathrm{W}}$ be the final state of W of the above operation. Then, we have $\mathrm{tr}[\hat{\Delta}^m\hat{\rho}_{\mathrm{W}}] = \mathrm{tr}[\hat{\Delta}^m\hat{\rho}'_{\mathrm{W}}]$. Therefore, W is regarded as a catalyst of coherence. We also note that, by applying the foregoing argument to SB instead of S, a thermal operation of SWB can create coherence of S, if W has coherence.

It is straightforward to check that $\mathscr{E}_{\hat{U}}$ constructed from $\hat{V}_{\hat{U}}$ is unital for any $\hat{U}$ and any initial state of W. Thus the work bound (8.34) is applicable. Then, by implementing the above-discussed optimal $\hat{U}$ that saturates the work bound (8.34), we obtain the optimal protocol including W under the strict energy conservation and with a large amount of initial coherence of W.

As a side remark, we consider the case that the initial state of W does not have any coherence in the foregoing construction. Suppose that $\hat{\rho}_{\mathrm{W}}$ in Eq. (8.37) is diagonal in the energy basis, which implies that $\mathrm{tr}[\hat{\Delta}^m\hat{\rho}_{\mathrm{W}}] = 0$ for $m \neq 0$. In this case, it is easy to show that

$$\mathrm{tr}[\mathscr{E}_{\hat{U}}(\mathscr{D}(\hat{\rho}_{\mathrm{S}}))\hat{H}_{\mathrm{S}}] = \mathrm{tr}[\mathscr{E}_{\hat{U}}(\hat{\rho}_{\mathrm{S}})\hat{H}_{\mathrm{S}}], \tag{8.39}$$

where $\mathscr{D}(\hat{\rho}_{\mathrm{S}}) := \sum_i |E_i\rangle\langle E_i|\hat{\rho}_{\mathrm{S}}|E_i\rangle\langle E_i|$ with $\hat{H}_{\mathrm{S}} = \sum_i E_i|E_i\rangle\langle E_i|$. Equation (8.39) implies that coherence of $\hat{\rho}_{\mathrm{S}}$ does not play any role in the energy balance during the process. When we explicitly consider B, notice that the initial state of B, $\hat{\rho}_{\mathrm{B}}^{\mathrm{G}}$, is already diagonal in the energy basis. Then, in the case that $\hat{H}_{\mathrm{S}} = \hat{H}'_{\mathrm{S}}$ and $\hat{\rho}'_{\mathrm{S}} = \hat{\rho}_{\mathrm{S}}^{\mathrm{G}}$, we obtain a tighter work bound with the incoherent work storage [169]:

$$-\beta W \leq S_1(\mathscr{D}(\hat{\rho}_{\mathrm{S}})\|\hat{\rho}_{\mathrm{S}}^{\mathrm{G}}), \tag{8.40}$$

where the right-hand side is equal to or smaller than $S_1(\hat{\rho}_{\mathrm{S}}\|\hat{\rho}_{\mathrm{S}}^{\mathrm{G}})$ in (8.34). This implies that, in the absence of coherence of W and under the strict energy conservation, one cannot harvest coherence of the initial state of S for the work extraction.

Next, we consider the case of the average energy conservation in line with Ref. [160]. Our goal is to implement a unitary $\hat{U}$ acting on S by a unitary $\hat{V}'_{\hat{U}}$ acting on SW that only conserves the average energy for a given initial state $\hat{\rho}_{\mathrm{S}}$.

Let $\hat{\rho}_{\mathrm{S}} = \sum_i p_i|\varphi_i\rangle\langle\varphi_i|$ be the initial state of S and $\hat{H}_{\mathrm{S}}$ be its Hamiltonian. Suppose that W is a continuous system with $(\hat{x}, \hat{p})$ satisfying $[\hat{x}, \hat{p}] = \mathrm{i}$ and its Hamiltonian is given by $\hat{H}_{\mathrm{W}} = \hat{x}$. Instead of Eqs. (8.35) or (8.36), we consider the following unitary operator that depends on $\hat{\rho}_{\mathrm{S}}$:

$$\hat{V}'_{\hat{U}} := \sum_i \hat{U}|\varphi_i\rangle\langle\varphi_i| \otimes e^{-\mathrm{i}\epsilon_i\hat{p}}, \tag{8.41}$$

where

$$\epsilon_i := \langle \varphi_i | \hat{H}_S | \varphi_i \rangle - \langle \varphi_i | \hat{U}^\dagger \hat{H}_S \hat{U} | \varphi_i \rangle. \tag{8.42}$$

The corresponding CPTP map on S is given by $\mathscr{E}'_{\hat{U}}(\bullet) := \text{tr}_W[\hat{V}'_{\hat{U}} \bullet \otimes \hat{\rho}_W \hat{V}'^{\dagger}_{\hat{U}}]$, where $\hat{\rho}_W$ is an initial state of W. Although $\mathscr{E}'_{\hat{U}}$ is not unitary in general, it reproduces the action of unitary $\hat{U}$ only for the given initial state $\hat{\rho}_S$; it is straightforward to show that

$$\text{tr}_W[\hat{V}'_{\hat{U}} \hat{\rho}_S \otimes \hat{\rho}_W \hat{V}'^{\dagger}_{\hat{U}}] = \hat{U} \hat{\rho}_S \hat{U}^\dagger. \tag{8.43}$$

Thus, the energy change of S by $\hat{V}'_{\hat{U}}$ equals that by $\hat{U}$, which is given by $\text{tr}_S[\hat{H}_S \hat{U} \hat{\rho}_S \hat{U}^\dagger] - \text{tr}_S[\hat{H}_S \hat{\rho}_S] = -\sum_i p_i \epsilon_i$. Let $\hat{\rho}'_W := \text{tr}_S[\hat{V}'_{\hat{U}} \hat{\rho}_S \otimes \hat{\rho}_W \hat{V}'^{\dagger}_{\hat{U}}]$ be the final state of W. The energy change of W is given by $\text{tr}_W[\hat{H}_W \hat{\rho}'_W] - \text{tr}_W[\hat{H}_W \hat{\rho}_W] = \sum_i p_i \epsilon_i$. Thus, $\hat{V}'_{\hat{U}}$ satisfies the average energy conservation.

We emphasize that $\hat{V}'_{\hat{U}}$ simulates the action of $\hat{U}$ and conserves the average energy only for the given initial state $\hat{\rho}_S$; In fact, $\hat{V}'_{\hat{U}}$ itself depends on $\hat{\rho}_S$. On the other hand, the initial state $\hat{\rho}_W$ of W is arbitrary; In particular, the initial coherence of W is not required for this protocol. We note that the translation invariance of W is satisfied: $[\hat{V}'_{\hat{U}}, e^{-i\varepsilon \hat{p}}] = 0$ for all $\varepsilon \in \mathbb{R}$.

It is straightforward to check that $\mathscr{E}'_{\hat{U}}$ constructed from $\hat{V}'_{\hat{U}}$ is unital for any $\hat{U}$ and any initial state of W. Thus the work bound (8.34) is again applicable. Then, by constructing $\hat{V}'_{\hat{U}}$ from the optimal $\hat{U}$ that saturates the work bound (8.34), we obtain the optimal protocol including W under the average energy conservation.

8.5   Single-Shot Work Bound: Exact Case

In the remaining part of this chapter, we consider the work bounds for the single-shot scenario, where work fluctuation is not allowed. This is the quantum extension of the resource-theoretic argument in Sect. 4.2.

The following argument is again based on the setup and the notations of Sect. 8.3, especially the Hamiltonian of SCW (8.16). In addition, in the case of the single-shot scenario, it is natural to take the work storage W as a two-level system, whose Hamiltonian is given by

$$\hat{H}_W = E_i |i\rangle\langle i| + E_f |f\rangle\langle f|, \quad E_i - E_f = w, \tag{8.44}$$

where w represents the single-shot work cost. In the following, we assume this form of $\hat{H}_W$. Then, we define thermodynamic processes of the single-shot setup with work cost w as follows.

Definition 8.4 (*Single-shot work-assisted state transformation*) Let $w \in \mathbb{R}$. A state $\hat{\rho}_S$ is w-assisted transformable to another state $\hat{\rho}'_S$ with the initial and final Hamilto-

nians $\hat{H}_S$ and $\hat{H}'_S$, if there exists a Gibbs-preserving map of SCW, written as $\mathscr{E}_{SCW}$, with respect to the Hamiltonian $\hat{H}_{SCW}$ of Eq. (8.16) with $\hat{H}_W$ satisfying Eq. (8.44), such that

$$\mathscr{E}_{SCW}\left(\hat{\rho}_S \otimes |0\rangle\langle 0| \otimes |E_i\rangle\langle E_i|\right) = \hat{\rho}'_S \otimes |1\rangle\langle 1| \otimes |E_f\rangle\langle E_f|. \qquad (8.45)$$

In the above definition, we required the existence of $\mathscr{E}_{SCW}$ satisfying Eq. (8.45) only for a given initial state $\hat{\rho}_S$, which represents that the function of the clock and the work transfer are supposed to be both perfect at least for $\hat{\rho}_S$.

The following theorem gives the work bounds of the single-shot scenario, which corresponds to Theorem 6.3. This has been shown in Refs. [41, 42, 45] (see also Proposition 16 of Ref. [129]).

Theorem 8.1 (Single-shot work bounds) *Let* $w \in \mathbb{R}$.

(a) *Necessary conditions for state conversion:*   If $\hat{\rho}_S$ is w-assisted transformable to $\hat{\rho}'_S$ with the Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$, then

$$\beta(w - \Delta F_S) \geq S_\alpha(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_\alpha(\hat{\rho}_S \| \hat{\rho}_S^G), \qquad (8.46)$$

for $\alpha = 0, 1, \infty$.

(b) *Sufficient condition for state conversion:* $\hat{\rho}_S$ is w-assisted transformable to $\hat{\rho}'_S$ with the Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$, if (but not only if)

$$\beta(w - \Delta F_S) \geq S_\infty(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_0(\hat{\rho}_S \| \hat{\rho}_S^G). \qquad (8.47)$$

Proof Let $\hat{\rho} := \hat{\rho}_S \otimes |0\rangle\langle 0| \otimes |E_i\rangle\langle E_i|$ and $\hat{\rho}' := \mathscr{E}_{SCW}(\hat{\rho}) = \hat{\rho}'_S \otimes |1\rangle\langle 1| \otimes |E_f\rangle\langle E_f|$ by noting Eq. (8.45). The Gibbs states of SCW and SC are respectively given by Eqs. (8.18) and (8.19), where in the present setup,

$$\hat{\rho}_W^G = \frac{e^{-\beta E_i}}{Z_W} |E_i\rangle\langle E_i| + \frac{e^{-\beta E_f}}{Z_W} |E_f\rangle\langle E_f| \qquad (8.48)$$

with $Z_W = e^{-\beta E_i} + e^{-\beta E_f}$. Thus, for $\alpha = 0, 1, \infty$,

$$S_\alpha(\hat{\rho} \| \hat{\rho}_{SCW}^G) = S_\alpha(\hat{\rho}_S \| \hat{\rho}_S^G) + \beta F_S + \ln(Z_S + Z'_S) + \beta E_i + \ln Z_W, \qquad (8.49)$$

and

$$S_\alpha(\hat{\rho}' \| \hat{\rho}_{SCW}^G) = S_\alpha(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) + \beta F'_S + \ln(Z_S + Z'_S) + \beta E_f + \ln Z_W, \qquad (8.50)$$

where we used the scaling property (5.47).

(a) From the monotonicity of the Rényi α-divergence of SCW, we have $S_\alpha(\hat{\rho} \| \hat{\rho}_{SCW}^G) \geq S_\alpha(\hat{\rho}' \| \hat{\rho}_{SCW}^G)$. By noting Eqs. (8.49) and (8.50), the monotonicity leads to inequality (8.46).

(b) From Theorem 6.3 (b), $S_\infty(\hat{\rho}\|\hat{\rho}^G_{SCW}) \geq S_0(\hat{\rho}'\|\hat{\rho}^G_{SCW})$ implies that the transformation from $\hat{\rho}$ to $\hat{\rho}'$ is possible. This is exactly the claim of Theorem 8.1 (b), again because of Eqs. (8.49) and (8.50). □

By using the α-free energy introduced in Eq. (8.1), we rewrite inequalities (8.46) and (8.47) in Theorem 8.1 as, respectively,

$$w \geq F_\alpha(\hat{\rho}'_S; \hat{H}'_S) - F_\alpha(\hat{\rho}_S; \hat{H}_S), \tag{8.51}$$

and

$$w \geq F_\infty(\hat{\rho}'_S; \hat{H}'_S) - F_0(\hat{\rho}_S; \hat{H}_S). \tag{8.52}$$

We next consider the special cases of the above theorem, which reproduce the work bounds of the classical case obtained by the Lorenz curve in Sect. 4.2.

Corollary 8.1 (Work extraction/state formation) *Let $w \in \mathbb{R}$.*
(a) $\hat{\rho}_S$ is w-assisted transformable to $\hat{\rho}^G_S$ with a fixed Hamiltonian $\hat{H}_S$, if and only if

$$-\beta w \leq S_0(\hat{\rho}_S\|\hat{\rho}^G_S). \tag{8.53}$$

(b) $\hat{\rho}^G$ is w-assisted transformable to $\hat{\rho}'_S$ with a fixed Hamiltonian $\hat{H}_S$, if and only if

$$\beta w \geq S_\infty(\hat{\rho}'_S\|\hat{\rho}^G_S). \tag{8.54}$$

(c) $\hat{\rho}^G$ is w-assisted transformable to $\hat{\rho}^{G'}_S$ with initial and final Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$, if and only if

$$w \geq \Delta F_S. \tag{8.55}$$

Proof To prove (a) and (b) above, apply Theorem 8.1 to the case that $\hat{\rho}'_S = \hat{\rho}^G_S$ and $\hat{\rho}_S = \hat{\rho}^G_S$, respectively. To prove (c), take $\hat{\rho}_S = \hat{\rho}^G_S$ and $\hat{\rho}'_S = \hat{\rho}^{G'}_S$. □

8.6   Single-Shot Work Bound: Approximate Case

We consider the single-shot work bound for approximate (or imperfect) thermodynamic processes. There can be several sources of "failure" in thermodynamic processes: the final state of the system deviates from the target state, the clock works imperfectly, or the work extraction fails with some probability, and so on. To incorporate such sources altogether, we adopt the following definition of approximate thermodynamic processes.

Definition 8.5 (*ε-approximate single-shot thermodynamic process*) Let $w \in \mathbb{R}$ and $\varepsilon \geq 0$. A state $\hat{\rho}_S$ is ε-approximate w-assisted transformable to another state $\hat{\rho}'_S$ with

Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$, if there exists a Gibbs-preserving map of SCW, written as $\mathscr{E}_{SCW}$, with respect to the Hamiltonian $\hat{H}_{SCW}$ of Eq. (8.16) with $\hat{H}_W$ satisfying Eq. (8.44), such that

$$D(\mathscr{E}_{SCW}(\hat{\rho}), \hat{\rho}') \leq \varepsilon, \tag{8.56}$$

where $\hat{\rho} := \hat{\rho}_S \otimes |0\rangle\langle 0| \otimes |E_i\rangle\langle E_i|$ and $\hat{\rho}' := \hat{\rho}'_S \otimes |1\rangle\langle 1| \otimes |E_f\rangle\langle E_f|$.

We remark that in Definition 8.5 the Hamiltonian of W may include additional levels; the Hamiltonian (8.44) can be replaced by $\hat{H}_W = E_i|E_i\rangle\langle E_i| + E_f|E_f\rangle\langle E_f| + \sum_k E_k|E_k\rangle\langle E_k|$ with $E_i - E_f = w$. For example, if the work extraction fails, the final state of W might end up with some state other than $|E_i\rangle, |E_f\rangle$.

From Definition 8.5, we see that the deviation of the final state of S and the failure probability of CW are both suppressed by ε. In fact, from the monotonicity of the trace distance under the partial trace, the condition (8.56) implies that $D(\hat{\rho}''_S, \hat{\rho}'_S) \leq \varepsilon$ with $\hat{\rho}''_S := \mathrm{tr}_{CW}[\mathscr{E}_{SCW}(\hat{\rho})]$. The condition (8.56) also implies that the success probability of the clock and the work storage (i.e., the probability that C ends up with $|1\rangle$ and W ends up with $|E_f\rangle$) equals or is greater than $1 - \varepsilon$, which is shown as follows. Let $\hat{\tau} := \mathscr{E}_{SCW}(\hat{\rho})$ and $|f\rangle := |1\rangle|E_f\rangle$. Then, $D(\hat{\tau}, \hat{\rho}') \leq \varepsilon$ implies that there exists $\hat{\Delta}$ such that $\hat{\tau} \geq \hat{\rho}' - \hat{\Delta}$, $\hat{\Delta} \geq 0$, and $\mathrm{tr}[\hat{\Delta}] \leq \varepsilon$. We then have $\langle f|\hat{\tau}|f\rangle \geq \hat{\rho}'_S - \langle f|\hat{\Delta}|f\rangle$, and therefore

$$\mathrm{tr}[\langle f|\hat{\tau}|f\rangle] \geq 1 - \varepsilon. \tag{8.57}$$

Another advantage of Definition 8.5 is that it is directly related to the asymptotic theory discussed in Sect. 8.7. Under this definition, however, the necessary condition for approximate state conversion is given by an apparently involved inequality with the hypothesis testing divergence S_H introduced in Appendix B, as shown in the following theorem. (We will discuss alternative definitions of approximate processes later in this section, where the smooth Rényi 0- and ∞-divergences gives simpler thermodynamic bounds, as obtained in Refs. [41, 42].)

Theorem 8.2 (Necessary condition for approximate state conversion; Proposition 4 of [129]) *Let $\varepsilon \geq 0$. If $\hat{\rho}_S$ is ε-approximate w-assisted transformable to $\hat{\rho}'_S$ with the Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$, then for $0 < \eta < 1 - \varepsilon$,*

$$\beta(w - \Delta F_S) \geq S_H^{\eta+\varepsilon}(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_H^{\eta}(\hat{\rho}_S \| \hat{\rho}_S^{G}) - \ln\left(\frac{\eta + \varepsilon}{\eta}\right). \tag{8.58}$$

Proof We use the notations $\hat{\rho}$ and $\hat{\rho}'$ in Definition 8.5. First, we have

$$S_H^{\eta+\varepsilon}(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) = S_H^{\eta+\varepsilon}(\hat{\rho}' \| \hat{\rho}_{SCW}^{G}) + \ln\frac{Z'_S}{Z_S + Z'_S} - \beta E_f \tag{8.59}$$

$$\leq S_H^{\eta}(\mathscr{E}_{SCW}(\hat{\rho}) \| \hat{\rho}_{SCW}^{G}) + \ln\left(\frac{\eta + \varepsilon}{\eta}\right) + \ln\frac{Z'_S}{Z_S + Z'_S} - \beta E_f, \tag{8.60}$$

where the equality in the first line is from Lemma B.2 and the inequality in the second line is from Lemma B.1. From the monotonicity under $\mathscr{E}_{\text{SCW}}$,

$$S_{\text{H}}^{\eta}(\mathscr{E}_{\text{SCW}}(\hat{\rho}) \| \hat{\rho}_{\text{SCW}}^{\text{G}}) \leq S_{\text{H}}^{\eta}(\hat{\rho} \| \hat{\rho}_{\text{SCW}}^{\text{G}}). \tag{8.61}$$

Again from Lemma B.2, we have

$$S_{\text{H}}^{\eta}(\hat{\rho} \| \hat{\rho}_{\text{SCW}}^{\text{G}}) = S_{\text{H}}^{\eta}(\hat{\rho}_{\text{S}} \| \hat{\rho}_{\text{S}}^{\text{G}}) - \ln \frac{Z_{\text{S}}}{Z_{\text{S}} + Z_{\text{S}}'} + \beta E_{\text{i}}. \tag{8.62}$$

By combining the foregoing relations, we obtain

$$S_{\text{H}}^{\eta+\varepsilon}(\hat{\rho}_{\text{S}}' \| \hat{\rho}_{\text{S}}^{\text{G}'}) \leq S_{\text{H}}^{\eta}(\hat{\rho}_{\text{S}} \| \hat{\rho}_{\text{S}}^{\text{G}}) + \ln \frac{Z_{\text{S}}'}{Z_{\text{S}}} + \beta(E_{\text{i}} - E_{\text{f}}) + \ln\left(\frac{\eta+\varepsilon}{\eta}\right), \tag{8.63}$$

which implies inequality (8.58). □

By using Proposition B.1, inequality (8.58) can be rewritten in terms of the smooth Rényi 0- and ∞-divergences, while we will not explicitly write it down here, as it is still involved.

We next consider a sufficient condition for state conversion, which is an approximate version of Theorem 8.1 (b) and now is given by a simple form.

Theorem 8.3 (Sufficient condition for approximate state conversion; Proposition 16 of [129]) $\hat{\rho}_{\text{S}}$ *is ε-approximate w-assisted transformable to $\hat{\rho}_{\text{S}}'$ with the Hamiltonians* $\hat{H}_{\text{S}}$ *and* $\hat{H}_{\text{S}}'$, *if (but not only if)*

$$\beta(w - \Delta F_{\text{S}}) \geq S_{\infty}^{\varepsilon/2}(\hat{\rho}_{\text{S}}' \| \hat{\rho}_{\text{S}}^{\text{G}'}) - S_{0}^{\varepsilon/2}(\hat{\rho}_{\text{S}} \| \hat{\rho}_{\text{S}}^{\text{G}}). \tag{8.64}$$

Proof Let $\hat{\tau}_{\text{S}}$ be optimal for $S_{0}^{\varepsilon/2}(\hat{\rho}_{\text{S}} \| \hat{\rho}_{\text{S}}^{\text{G}})$ and $\hat{\tau}_{\text{S}}'$ be optimal for $S_{\infty}^{\varepsilon/2}(\hat{\rho}_{\text{S}}' \| \hat{\rho}_{\text{S}}^{\text{G}'})$. We then have

$$\beta(w - \Delta F_{\text{S}}) \geq S_{\infty}(\hat{\tau}_{\text{S}}' \| \hat{\rho}_{\text{S}}^{\text{G}'}) - S_{0}(\hat{\tau}_{\text{S}} \| \hat{\rho}_{\text{S}}^{\text{G}}). \tag{8.65}$$

From this inequality and Theorem 8.1 (b), $\hat{\tau}_{\text{S}}$ is w-assisted transformable to $\hat{\tau}_{\text{S}}'$. Define $\hat{\tau} := \hat{\tau}_{\text{S}} \otimes |0\rangle\langle 0| \otimes |E_{\text{i}}\rangle\langle E_{\text{i}}|$ and $\hat{\tau}' := \hat{\tau}_{\text{S}}' \otimes |1\rangle\langle 1| \otimes |E_{\text{f}}\rangle\langle E_{\text{f}}|$, and note that $D(\hat{\rho}, \hat{\tau}) \leq \varepsilon/2$ and $D(\hat{\rho}', \hat{\tau}') \leq \varepsilon/2$. Then, there exists a Gibbs-preserving map $\mathscr{E}_{\text{SCW}}$ such that $\mathscr{E}_{\text{SCW}}(\hat{\tau}) = \hat{\tau}'$. Then, $D(\mathscr{E}_{\text{SCW}}(\hat{\rho}), \hat{\rho}') \leq D(\mathscr{E}_{\text{SCW}}(\hat{\rho}), \hat{\tau}') + D(\hat{\tau}', \hat{\rho}') \leq D(\hat{\rho}, \hat{\tau}) + D(\hat{\tau}', \hat{\rho}') \leq \varepsilon$. □

As mentioned before, we can also introduce alternative definitions of approximate processes, which are given by slightly stronger conditions than Definition 8.5, leading to simpler work bounds. We here define two kinds of approximate processes, which we label by $\alpha = 0, \infty$, having in mind that these two definitions correspond to the Rényi 0- and ∞-divergences.

Definition 8.6 (*Strongly ε-approximate single-shot thermodynamic process*) Let $w \in \mathbb{R}$ and $\varepsilon \geq 0$. A state $\hat{\rho}_S$ is α-strongly ε-approximate w-assisted transformable to another state $\hat{\rho}'_S$ with Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$, if:

$\alpha = 0$; there exists a state $\hat{\tau}_S$ satisfying $D(\hat{\tau}_S, \hat{\rho}_S) \leq \varepsilon$ such that $\hat{\tau}_S$ is w-assisted transformable to $\hat{\rho}'_S$.

$\alpha = \infty$; there exists a state $\hat{\tau}'_S$ satisfying $D(\hat{\tau}'_S, \hat{\rho}'_S) \leq \varepsilon$ such that $\hat{\rho}_S$ is w-assisted transformable to $\hat{\tau}'_S$.

Here, the roles of the initial and final states are exchanged for $\alpha = 0, \infty$; in this sense, these two definitions are dual to each other. More importantly, for both of $\alpha = 0, 1$, "α-strongly ε-approximate transformable" automatically implies "ε-approximate transformable" (Definition 8.5).

To see this, we set the notations $\hat{\rho} := \hat{\rho}_S \otimes |1\rangle\langle 1| \otimes |E_f\rangle\langle E_f|$, $\hat{\tau} := \hat{\tau}_S \otimes |0\rangle\langle 0| \otimes |E_i\rangle\langle E_i|$, $\hat{\rho}' := \hat{\rho}'_S \otimes |1\rangle\langle 1| \otimes |E_f\rangle\langle E_f|$, $\hat{\tau}' := \hat{\tau}'_S \otimes |1\rangle\langle 1| \otimes |E_f\rangle\langle E_f|$, and let $\mathscr{E}_{SCW}$ be the relevant Gibbs preserving map with respect to the Hamiltonian $\hat{H}_{SCW}$ of Eq. (8.16), satisfying $\mathscr{E}_{SCW}(\hat{\tau}) = \hat{\rho}'$ for $\alpha = 0$ or $\mathscr{E}_{SCW}(\hat{\rho}) = \hat{\tau}'$ for $\alpha = \infty$. For $\alpha = 0$, if the condition of Definition 8.6 is satisfied, then $D(\mathscr{E}_{SCW}(\hat{\rho}), \hat{\rho}') \leq D(\hat{\rho}, \hat{\tau}) = D(\hat{\rho}_S, \hat{\tau}_S) \leq \varepsilon$ holds from the monotonicity of the trace distance. For $\alpha = \infty$, if the condition of Definition 8.6 is satisfied, then obviously $D(\mathscr{E}_{SCW}(\hat{\rho}), \hat{\rho}') = D(\hat{\tau}'_S, \hat{\rho}'_S) \leq \varepsilon$.

A physical motivation behind Definition 8.6 can be illustrated as follows. For $\alpha = 0$, suppose that $\hat{\rho}_S$ can be written as $\hat{\rho}_S = (1 - \varepsilon)\hat{\tau}_S + \varepsilon\hat{\gamma}_S$, where $0 < \varepsilon < 1$ is some constant and $\hat{\tau}_S$ and $\hat{\gamma}_S$ are normalized states orthogonal to each other. Note that $D(\hat{\tau}_S, \hat{\rho}_S) \leq \varepsilon$. Also suppose that the condition of Definition 8.6 is satisfied for this $\hat{\tau}_S$ (i.e., $\mathscr{E}_{SCW}(\hat{\tau}) = \hat{\rho}'$ with the above notation). In such a case, $\hat{\tau}_S$ represents the "success event" in the initial state of the transformation with success probability $1 - \varepsilon$. On the other hand, for $\alpha = \infty$, Definition 8.6 simply means that state transformation of S is imperfect but CW goes to the desired final state with unit probability.

With the foregoing "strong" definitions of approximate transformation, we have the following simpler work bounds. They are approximate versions of Corollary 8.1 (a), (b), and can be regarded as essentially the same bounds as obtained in Refs. [41, 42]. (See also Refs. [170–173] for the related work bounds for information processing.)

Theorem 8.4 (Simple work bounds for strongly approximate processes) *Let $w \in \mathbb{R}$ and $\varepsilon \geq 0$. Consider states $\hat{\rho}_S$, $\hat{\rho}'_S$ and a fixed Hamiltonian $\hat{H}_S$ with the Gibbs state $\hat{\rho}^G_S$.*

(a) $\hat{\rho}_S$ *is 0-strongly ε-approximate w-assisted transformable to $\hat{\rho}^G_S$, if and only if*

$$- \beta w \leq S^\varepsilon_0(\hat{\rho}_S \| \hat{\rho}^G_S). \tag{8.66}$$

(b) $\hat{\rho}^G_S$ *is ∞-strongly ε-approximate w-assisted transformable to $\hat{\rho}'_S$, if and only if*

$$\beta w \geq S^\varepsilon_\infty(\hat{\rho}'_S \| \hat{\rho}^G_S). \tag{8.67}$$

Proof We first prove (a). For the sake of generality at this stage, suppose that $\hat{\rho}_S$ is 0-strongly ε-approximate w-assisted transformable to $\hat{\rho}'_S$, by allowing different initial and final Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$. From inequality (8.46) of Theorem 8.1 (a), we have $\beta(w - \Delta F_S) \geq S_0(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_0(\hat{\tau}_S \| \hat{\rho}_S^G)$. Meanwhile, because $\hat{\tau}_S$ is a candidate for optimization in $S_0^\varepsilon(\hat{\rho}_S \| \hat{\rho}_S^G)$, we have $S_0(\hat{\tau}_S \| \hat{\rho}_S^G) \leq S_0^\varepsilon(\hat{\rho}_S \| \hat{\rho}_S^G)$. We thus have

$$\beta(w - \Delta F_S) \geq S_0(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_0^\varepsilon(\hat{\rho}_S \| \hat{\rho}_S^G). \tag{8.68}$$

By letting $\hat{H}_S = \hat{H}'_S$ and $\hat{\rho}'_S = \hat{\rho}_S^G$, we obtain inequality (8.66).

To prove the converse, suppose that

$$\beta(w - \Delta F_S) \geq S_\infty(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_0^\varepsilon(\hat{\rho}_S \| \hat{\rho}_S^G) \tag{8.69}$$

for general $\hat{\rho}_S$, $\hat{\rho}'_S$, $\hat{H}_S$, $\hat{H}'_S$. Let $\hat{\tau}_S$ be optimal for $S_0^\varepsilon(\hat{\rho}_S \| \hat{\rho}_S^G)$, which satisfies $S_0^\varepsilon(\hat{\rho}_S \| \hat{\rho}_S^G) = S_0(\hat{\tau}_S \| \hat{\rho}_S^G)$ and $D(\hat{\tau}_S, \hat{\rho}_S) \leq \varepsilon$. We have $\beta(w - \Delta F_S) \geq S_\infty(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_0(\hat{\tau}_S \| \hat{\rho}_S^G)$. Then, from Theorem 8.1 (b), $\hat{\tau}_S$ is w-assisted transformable to $\hat{\rho}'_S$, which implies that $\hat{\rho}_S$ is 0-strongly ε-approximate w-assisted transformable to $\hat{\rho}'_S$. By letting $\hat{H}_S = \hat{H}'_S$ and $\hat{\rho}'_S = \hat{\rho}_S^G$, we obtain the claim of (a).

We next prove (b). Again for the sake of generality at this stage, suppose that $\hat{\rho}_S$ is ∞-strongly ε-approximate w-assisted transformable to $\hat{\rho}'_S$, by allowing different initial and final Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$. From inequality (8.46) of Theorem 8.1 (a), we have $\beta(w - \Delta F_S) \geq S_\infty(\hat{\tau}'_S \| \hat{\rho}_S^{G'}) - S_\infty(\hat{\rho}_S \| \hat{\rho}_S^G)$. Meanwhile, because $\hat{\tau}'_S$ is a candidate for optimization in $S_\infty^\varepsilon(\hat{\rho}'_S \| \hat{\rho}_S^{G'})$, we have $S_\infty(\hat{\tau}'_S \| \hat{\rho}_S^{G'}) \geq S_\infty^\varepsilon(\hat{\rho}'_S \| \hat{\rho}_S^{G'})$. We thus have

$$\beta(w - \Delta F_S) \geq S_\infty^\varepsilon(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_\infty(\hat{\rho}_S \| \hat{\rho}_S^G). \tag{8.70}$$

By letting $\hat{H}_S = \hat{H}'_S$ and $\hat{\rho}_S = \hat{\rho}_S^G$, we obtain inequality (8.67).

To prove the converse, suppose that

$$\beta(w - \Delta F_S) \geq S_\infty^\varepsilon(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) - S_0(\hat{\rho}_S \| \hat{\rho}_S^G) \tag{8.71}$$

for general $\hat{\rho}_S$, $\hat{\rho}'_S$, $\hat{H}_S$, $\hat{H}'_S$. Let $\hat{\tau}'_S$ be optimal for $S_\infty^\varepsilon(\hat{\rho}'_S \| \hat{\rho}_S^{G'})$, which satisfies $S_\infty^\varepsilon(\hat{\rho}'_S \| \hat{\rho}_S^{G'}) = S_\infty(\hat{\tau}'_S \| \hat{\rho}_S^{G'})$ and $D(\hat{\tau}'_S, \hat{\rho}'_S) \leq \varepsilon$. We have $\beta(w - \Delta F_S) \geq S_\infty(\hat{\tau}'_S \| \hat{\rho}_S^{G'}) - S_0(\hat{\rho}_S \| \hat{\rho}_S^G)$. Then, from Theorem 8.1 (b), $\hat{\rho}_S$ is w-assisted transformable to $\hat{\tau}'_S$, which implies that $\hat{\rho}_S$ is ∞-strongly ε-approximate w-assisted transformable to $\hat{\rho}'_S$. By letting $\hat{H}_S = \hat{H}'_S$ and $\hat{\rho}_S = \hat{\rho}_S^G$, we obtain the claim of (a). $\square$

A protocol that saturates the equality in inequality (8.66) is presented in Ref. [42], which is a slight modification of the protocol of Fig. 4.5 in Sect. 4.2, by allowing failure of work extraction with small probability ($\leq \varepsilon$). We also note that, by substituting $\hat{\rho}_S = \hat{\rho}_S^G$, inequality (8.66) suggests that a positive amount of work can be extracted even from the Gibbs state if a small probability of failure is allowed.

8.7 Single-Shot Work Bound: Asymptotic Case

We consider the single-shot work bound for macroscopic systems, by taking the asymptotic limit of approximate thermodynamic processes in the sense of Definition 8.5. This is an application of the general asymptotic theory developed in Sects. 7.2 and 7.3. In terms of thermodynamics, the asymptotic limit represents the thermodynamic limit of many-body systems, where we do not necessarily assume the i.i.d. setup. Physically, i.i.d. systems represent non-interacting systems, while we can treat interacting systems in the following argument. In particular, we show that if the state is ergodic and the Hamiltonian is local and translation invariant, then the KL divergence provides the work bound in a necessary and sufficient manner, which is a thermodynamic consequence of the quantum AEP (Theorem 7.10).

Let $\widehat{P}_S := \{\hat{\rho}_{S,n}\}_{n\in\mathbb{N}}$ and $\widehat{P}'_S := \{\hat{\rho}'_{S,n}\}_{n\in\mathbb{N}}$ be sequences of states of S and $\widehat{H}_S := \{\hat{H}_{S,n}\}_{n\in\mathbb{N}}$ and $\widehat{H}'_S := \{\hat{H}'_{S,n}\}_{n\in\mathbb{N}}$ be sequences of Hamiltonians of S. Let $\widehat{\Sigma}_S := \{\hat{\rho}^G_{S,n}\}_{n\in\mathbb{N}}$ and $\widehat{\Sigma}'_S := \{\hat{\rho}^G_{S,n}{}'\}_{n\in\mathbb{N}}$ be the corresponding sequences of the Gibbs states. Suppose that the equilibrium free-energy rates exist, defined as

$$F_S := \lim_{n\to\infty} \frac{1}{n} F_{S,n}, \quad F'_S := \lim_{n\to\infty} \frac{1}{n} F'_{S,n}, \tag{8.72}$$

where $F_{S,n}$ and $F'_{S,n}$ are the free energies corresponding to $\hat{H}_{S,n}$ and $\hat{H}'_{S,n}$, respectively. Let $\Delta F_{S,n} := F'_{S,n} - F_{S,n}$ and $\Delta F_S := F'_S - F_S$. We then define the asymptotic limit of single-shot thermodynamic processes.

Definition 8.7 (*Asymptotic thermodynamic process*) Let $w \in \mathbb{R}$. A sequence $\widehat{P}_S$ is asymptotically w-assisted transformable to another sequence $\widehat{P}'_S$ with respect to $\widehat{H}_S$ and $\widehat{H}'_S$, if there exist sequences $\{w_n\}_{n\in\mathbb{N}}$ and $\{\varepsilon_n\}_{n\in\mathbb{N}}$ with $w_n \in \mathbb{R}$ and $\varepsilon_n > 0$, such that $\hat{\rho}_{S,n}$ is ε_n-approximate w_n-assisted transformable to $\hat{\rho}'_{S,n}$, and $\lim_{n\to\infty} w_n/n = w$, $\lim_{n\to\infty} \varepsilon_n = 0$.

Based on this definition, we consider the work bounds. We first state the necessary conditions corresponding to Theorem 8.1 (a).

Theorem 8.5 (Necessary condition for asymptotic state conversion; Proposition 5 of [129]) *If $\widehat{P}_S$ is asymptotically w-assisted transformable to $\widehat{P}'_S$ with respect to $\widehat{H}_S$ and $\widehat{H}'_S$, then*

$$\beta(w - \Delta F_S) \geq \underline{S}(\widehat{P}'_S \| \widehat{\Sigma}'_S) - \underline{S}(\widehat{P}_S \| \widehat{\Sigma}_S), \tag{8.73}$$

$$\beta(w - \Delta F_S) \geq \overline{S}(\widehat{P}'_S \| \widehat{\Sigma}'_S) - \overline{S}(\widehat{P}_S \| \widehat{\Sigma}_S). \tag{8.74}$$

Proof Take the limit of Theorem 8.2, by applying Proposition B.1. $\qquad\square$

We next state the sufficient condition of state conversion corresponding to Theorem 8.1 (b).

Theorem 8.6 (Sufficient condition for asymptotic state conversion; Proposition 16 of [129]) *$\widehat{P}_S$ is asymptotically w-assisted transformable to $\widehat{P}_S'$ with respect to $\widehat{H}_S$ and $\widehat{H}_S'$, if (but not only if)*

$$\beta(w - \Delta F_S) \geq \overline{S}(\widehat{P}_S' \| \widehat{\Sigma}_S') - \underline{S}(\widehat{P}_S \| \widehat{\Sigma}_S). \tag{8.75}$$

Proof Suppose that inequality (8.75) holds. Then, for any $\varepsilon > 0$ and $n \in \mathbb{N}$, there exists $\Delta_{n,\varepsilon} \geq 0$ such that

$$\frac{\beta}{n}(nw - \Delta F_{S,n}) + \Delta_{n,\varepsilon} \geq \frac{1}{n} S_\infty^{\varepsilon/2}(\hat{\rho}_{S,n}' \| \hat{\rho}_{S,n}^{G\ '}) - \frac{1}{n} S_0^{\varepsilon/2}(\hat{\rho}_{S,n} \| \hat{\rho}_{S,n}^G) \tag{8.76}$$

and

$$\lim_{\varepsilon \to +0} \limsup_{n \to \infty} \Delta_{n,\varepsilon} = 0. \tag{8.77}$$

Let $w_{n,\varepsilon} := nw + n\Delta_{n,\varepsilon}$ that satisfies $\lim_{\varepsilon \to +0} \limsup_{n \to \infty} w_{n,\varepsilon}/n = w$. From Theorem 8.3, $\hat{\rho}_{S,n}$ is ε-approximate $w_{n,\varepsilon}$-assisted transformable to $\hat{\rho}_{S,n}'$. From Lemma 13 of Ref. [129], this implies that $\widehat{P}_S$ is asymptotically w-assisted transformable to $\widehat{P}_S'$.

For the sake of self-containedness, we here reproduce the proof of the above last step. Let $w_\varepsilon := \limsup_{n \to \infty} w_{n,\varepsilon}/n$, $N(\varepsilon) := \min\{N\ :\ \forall n \geq N, w_{n,\varepsilon}/n \leq w_\varepsilon + \varepsilon\}$, and $\varepsilon(n) := \inf\{\varepsilon : N(\varepsilon) \leq n\}$. Because of the existence of the limit superior, $N(\varepsilon)$ is finite for any $\varepsilon > 0$, and thus we see that $\lim_{n \to \infty} \varepsilon(n) = 0$. Then, by defining $w_n' := w_{n,\varepsilon(n)}$, we obtain $\limsup_{n \to \infty} w_n'/n \leq \limsup_{n \to \infty}(w_{\varepsilon(n)} + \varepsilon(n)) = w$. Because the state convertibility is not affected by adding any positive amount of work, we can construct w_n such that $\lim_{n \to \infty} w_n/n = w$, by adding some positive constant to w_n' if necessary. □

In contrast to Theorem 7.5, inequality (8.75) in the above theorem includes the equality case, thanks to the role of w in Definition 8.7. The following are the two special cases corresponding to Corollary 8.1 (a) (b).

Corollary 8.2 (Asymptotic work extraction/state formation) *Let $w \in \mathbb{R}$.*
(a) $\widehat{P}_S$ is asymptotically w-assisted transformable to $\widehat{\Sigma}_S$ with $\widehat{H}_S = \widehat{H}_S'$, if and only if

$$-\beta w \leq \underline{S}(\widehat{P}_S \| \widehat{\Sigma}_S). \tag{8.78}$$

(b) $\widehat{\Sigma}_S$ is asymptotically w-assisted transformable to $\widehat{P}_S'$ with $\widehat{H}_S = \widehat{H}_S'$, if and only if

$$\beta w \geq \overline{S}(\widehat{P}_S' \| \widehat{\Sigma}_S). \tag{8.79}$$

Finally, we state a necessary and sufficient characterization of state conversion in terms of the work bound for the case where the upper and lower spectral divergence rates collapse to a single value for the initial and final states.

Corollary 8.3 *Suppose that* $\overline{S}(\widehat{P}_S\|\widehat{\Sigma}_S) = \underline{S}(\widehat{P}_S\|\widehat{\Sigma}_S) =: S(\widehat{P}_S\|\widehat{\Sigma}_S)$ *and* $\overline{S}(\widehat{P}'_S\|\widehat{\Sigma}'_S)$ $= \underline{S}(\widehat{P}'_S\|\widehat{\Sigma}'_S) =: S(\widehat{P}'_S\|\widehat{\Sigma}'_S)$. *Then,* $\widehat{P}_S$ *is asymptotically w-assisted transformable to* $\widehat{P}'_S$ *with respect to* $\widehat{H}_S$ *and* $\widehat{H}'_S$, *if and only if*

$$\beta(w - \Delta F_S) \geq S(\widehat{P}'_S\|\widehat{\Sigma}'_S) - S(\widehat{P}_S\|\widehat{\Sigma}_S). \tag{8.80}$$

The above corollary implies that the upper and lower spectral divergence rates give a complete thermodynamic potential that can be defined for out-of-equilibrium situations. Correspondingly, we can introduce the nonequilibrium free-energy rate by

$$F(\widehat{P}_S; \widehat{H}_S) := \beta^{-1}S(\widehat{P}_S\|\widehat{\Sigma}_S) + F_S. \tag{8.81}$$

This illustrates the significance of information spectrum in thermodynamics.

We now consider a many-body quantum spin system on a lattice $\mathbb{Z}^d$ as in Sect. 7.3, where an explicit condition of the collapse of the upper and lower spectral divergence rates has been shown in Theorem 7.10. Combining it with Corollary 8.3 above, we have the following asymptotic work bound, which states that the complete thermodynamic potential emerges and is given by the KL divergence rate even for out-of-equilibrium and quantum situations, if the state is ergodic and the Hamiltonian is local and translation invariant.

Corollary 8.4 *Suppose that* $\widehat{P}_S$ *and* $\widehat{P}'_S$ *are translation invariant and ergodic, and that the Hamiltonians* $\widehat{H}_S$ *and* $\widehat{H}'_S$ *are local and translation invariant with the corresponding Gibbs states* $\widehat{\Sigma}_S$ *and* $\widehat{\Sigma}'_S$. *Then,* $\widehat{P}_S$ *is asymptotically w-assisted transformable to* $\widehat{P}'_S$ *with respect to* $\widehat{H}_S$ *and* $\widehat{H}'_S$, *if and only if*

$$\beta(w - \Delta F_S) \geq S_1(\widehat{P}'_S\|\widehat{\Sigma}'_S) - S_1(\widehat{P}_S\|\widehat{\Sigma}_S). \tag{8.82}$$

Corollary 8.4 may provide an information-theoretic and statistical-mechanical foundation of, and furthermore a nonequilibrium generalization of, phenomenological macroscopic thermodynamics of Lieb and Yngvason [2] where entropy is a complete monotone of equilibrium transitions. We also remark that inequality (8.82) is in the same form as the average work bound (8.26), which implies that the two formulations of thermodynamics with and without work fluctuations eventually give the same form of the second law in the macroscopic limit.

The emergence of the complete thermodynamic potential ensures the reversibility of thermodynamic transformations in the macroscopic limit. As an illustrative situation, let us consider a simple cycle in the single-shot scenario: one first transforms an equilibrium state $\hat{\rho}_S^G$ to a nonequilibrium state $\hat{\rho}_S$, and then restores $\hat{\rho}_S$ to $\hat{\rho}_S^G$. As shown in Corollary 8.1, this cycle requires work of $S_\infty(\hat{\rho}_S\|\hat{\rho}_S^G) - S_0(\hat{\rho}_S\|\hat{\rho}_S^G)$, which is positive if $\hat{\rho}_S$ is out of equilibrium. This implies that it is in general impossible to obtain a single thermodynamic potential that can completely characterize state

convertibility in the single-shot scenario. This is also contrastive to conventional thermodynamics for equilibrium transitions, in which a crucial postulate is that one can perform a reversible cyclic operation without remaining any effect on the outside world. On the other hand, Corollary 8.4 reveals that a thermodynamic potential emerges in the asymptotic limit, which is related to the concept of reversibility in resource theory [128, 174].

   We also note that one can see a characteristic of the single-shot scenario in light of the asymptotic theory; Thanks to the single-shot formulation, the work automatically becomes a deterministic quantity in the asymptotic limit, which is a desirable property in macroscopic thermodynamics.

   Furthermore, as shown in Refs. [129, 152], the operation in Corollary 8.4 can be replaced by an asymptotic thermal operation even in the fully quantum case, if the aid of a small amount of quantum coherence is available (Theorem 2 of Ref. [129]). This implies that thermal operation can work even in the fully quantum regime, if we take the asymptotic limit. The key of the proof is the fact that any ergodic state (more generally, any state with which the upper and lower spectral divergence rates are close) has a small coherence in the energy basis (Lemma 4 of Ref. [129]).

8.8   Trace-Nonincreasing Formulation

As a side remark, we consider a way to "trace out" the clock degrees of freedom from our formulation introduced in Sect. 8.3, where the notion of trace-nonincreasing naturally appears as a consequence of the assumption that the clock does not necessarily work perfectly.

   For simplicity, we ignore the work storage W and only focus on the role of the clock C. Suppose that the Hamiltonian is given by $\hat{H}_{\mathrm{SC}}$ of the form Eq. (8.16), where the corresponding Gibbs state $\hat{\rho}_{\mathrm{SC}}$ is given by Eq. (8.19).

Lemma 8.2 *Let $\mathscr{E}_{\mathrm{SC}}$ be a Gibbs-preserving map with the Hamiltonian $\hat{H}_{\mathrm{SC}}$ of Eq. (8.16). Then,*

$$\mathscr{E}_{\mathrm{S}}(\hat{\rho}_{\mathrm{S}}) := \langle 1|\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}} \otimes |0\rangle\langle 0|)|1\rangle \tag{8.83}$$

is CP and trace-nonincreasing, and satisfies

$$\mathscr{E}_{\mathrm{S}}(e^{-\beta \hat{H}_{\mathrm{S}}}) \leq e^{-\beta \hat{H}'_{\mathrm{S}}}. \tag{8.84}$$

Proof We prove inequality (8.84). Let $\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}}^{\mathrm{G}} \otimes |0\rangle\langle 0|) = \sum_{i,j=0,1} \hat{\sigma}_{ij} \otimes |i\rangle\langle j|$ and $\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}}^{\mathrm{G}'} \otimes |1\rangle\langle 1|) = \sum_{i,j=0,1} \hat{\sigma}'_{ij} \otimes |i\rangle\langle j|$.   Note   that $\mathscr{E}_{\mathrm{S}}(\hat{\rho}_{\mathrm{S}}^{\mathrm{G}}) = \hat{\sigma}_{11}$.   From Eq. (8.19), we have

$$\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{SC}}^{\mathrm{G}}) = \frac{Z_{\mathrm{S}}}{Z_{\mathrm{S}} + Z_{\mathrm{S}}'} \sum_{i,j=0,1} \hat{\sigma}_{ij} \otimes |i\rangle\langle j| + \frac{Z_{\mathrm{S}}'}{Z_{\mathrm{S}} + Z_{\mathrm{S}}'} \sum_{i,j=0,1} \hat{\sigma}_{ij}' \otimes |i\rangle\langle j|, \quad (8.85)$$

which equals $\hat{\rho}_{\mathrm{SC}}^{\mathrm{G}}$ because $\mathscr{E}_{\mathrm{SC}}$ is Gibbs-preserving. Thus, we find that $Z_{\mathrm{S}}\hat{\sigma}_{00} + Z_{\mathrm{S}}'\hat{\sigma}_{00}' = Z_{\mathrm{S}}\hat{\rho}_{\mathrm{S}}^{\mathrm{G}}, Z_{\mathrm{S}}\hat{\sigma}_{11} + Z_{\mathrm{S}}'\hat{\sigma}_{11}' = Z_{\mathrm{S}}'\hat{\rho}_{\mathrm{S}}^{\mathrm{G}\prime}, Z_{\mathrm{S}}\hat{\sigma}_{01} + Z_{\mathrm{S}}'\hat{\sigma}_{01}' = 0, Z_{\mathrm{S}}\hat{\sigma}_{10} + Z_{\mathrm{S}}'\hat{\sigma}_{10}' = 0$. Because $\mathscr{E}_{\mathrm{SC}}$ is CPTP, $\hat{\sigma}_{11}' \geq 0$. Thus, we obtain $Z_{\mathrm{S}}\hat{\sigma}_{11} \leq Z_{\mathrm{S}}'\hat{\rho}_{\mathrm{S}}^{\mathrm{G}\prime}$, which implies inequality (8.84). $\qquad\square$

If the clock works perfectly for all the initial states, i.e., if for any $\hat{\rho}_{\mathrm{S}}$ there exists $\hat{\rho}_{\mathrm{S}}'$ such that

$$\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}} \otimes |0\rangle\langle 0|) = \hat{\rho}_{\mathrm{S}}' \otimes |1\rangle\langle 1| \qquad (8.86)$$

holds, then $\mathscr{E}_{\mathrm{S}} : \hat{\rho}_{\mathrm{S}} \mapsto \hat{\rho}_{\mathrm{S}}'$ is TP. We have inequality (8.84) also in this case. However, the condition (8.86) for *all* $\hat{\rho}_{\mathrm{S}}$ is very strong, which is not necessarily satisfied. If fact, if $\mathscr{E}_{\mathrm{S}}$ is TP, we have $\mathrm{tr}[e^{-\beta\hat{H}_{\mathrm{S}}}] \leq \mathrm{tr}[e^{-\beta\hat{H}_{\mathrm{S}}'}]$ from (8.84), or equivalently $F_{\mathrm{S}} \geq F_{\mathrm{S}}'$, which is not necessarily satisfied in thermodynamic processes.

The converse of Lemma 8.2 is also true in the following sense.

Lemma 8.3 *For any CP and trace-nonincreasing map $\mathscr{E}_{\mathrm{S}}$ satisfying inequality (8.84), there exists a CPTP Gibbs-preserving map of SC with the Hamiltonian $\hat{H}_{\mathrm{SC}}$, written as $\mathscr{E}_{\mathrm{SC}}$, such that Eq. (8.83) holds.*

Proof We can construct $\mathscr{E}_{\mathrm{SC}}$ as follows. First, there exists a CP and trace-nonincreasing map $\mathscr{E}_{\mathrm{S}}'$ such that $\mathscr{E}_{\mathrm{S}}'(\hat{\rho}_{\mathrm{S}}) \leq \hat{\rho}_{\mathrm{S}}$ holds for any $\hat{\rho}_{\mathrm{S}}$ and $\mathscr{E}_{\mathrm{S}} + \mathscr{E}_{\mathrm{S}}'$ is CPTP. In fact, if the Kraus representation of $\mathscr{E}_{\mathrm{S}}$ is given by $\mathscr{E}_{\mathrm{S}}(\hat{\rho}_{\mathrm{S}}) = \sum_k \hat{M}_k \hat{\rho}_{\mathrm{S}} \hat{M}_k^\dagger$ with $\sum_k \hat{M}_k^\dagger \hat{M}_k \leq \hat{I}$, we can define $\mathscr{E}_{\mathrm{S}}'(\hat{\rho}_{\mathrm{S}}) := \sqrt{\hat{I} - \hat{E}}\,\hat{\rho}_{\mathrm{S}}\sqrt{\hat{I} - \hat{E}}$ with $\hat{E} := \sum_k \hat{M}_k^\dagger \hat{M}_k$. Then define

$$\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}} \otimes |0\rangle\langle 0|) := \mathscr{E}_{\mathrm{S}}'(\hat{\rho}_{\mathrm{S}}) \otimes |0\rangle\langle 0| + \mathscr{E}_{\mathrm{S}}(\hat{\rho}_{\mathrm{S}}) \otimes |1\rangle\langle 1|. \qquad (8.87)$$

We also define

$$\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}} \otimes |1\rangle\langle 1|) := \frac{e^{-\beta\hat{H}_{\mathrm{S}}} - \mathscr{E}_{\mathrm{S}}'(e^{-\beta\hat{H}_{\mathrm{S}}})}{Z_{\mathrm{S}}'} \otimes |0\rangle\langle 0| + \frac{e^{-\beta\hat{H}_{\mathrm{S}}'} - \mathscr{E}_{\mathrm{S}}(e^{-\beta\hat{H}_{\mathrm{S}}})}{Z_{\mathrm{S}}'} \otimes |1\rangle\langle 1|, \tag{8.88}$$

for any $\hat{\rho}_{\mathrm{S}}$. Finally, we let $\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}} \otimes |0\rangle\langle 1|) := 0$ and $\mathscr{E}_{\mathrm{SC}}(\hat{\rho}_{\mathrm{S}} \otimes |1\rangle\langle 0|) := 0$. By construction, along with the assumption (8.84), $\mathscr{E}_{\mathrm{SC}}$ is CPTP. It is also straightforward to check that $\mathscr{E}_{\mathrm{SC}}$ is Gibbs-preserving. $\qquad\square$

Now, we may adopt inequality (8.84) as an alternative generalized definition of Gibbs-preserving maps for the situation that the input and the output Hamiltonians are not the same and the clock does not necessarily work perfectly [45, 129], which we refer to as Gibbs-sub-preserving maps. We formally state the definition as follows:

Definition 8.8 (*Gibbs-sub-preserving maps*) A CP and trace-nonincreasing map $\mathscr{E}_S$ is Gibbs-sub-preserving with respect to the initial and final Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$, if

$$\mathscr{E}(e^{-\beta \hat{H}_S}) \le e^{-\beta H'_S}. \tag{8.89}$$

We can rephrase the second law in the form of Theorem 8.1 (a) as follows.

Corollary 8.5 (*Proposition 3 of* [129]) *Let $\hat{\rho}_S$, $\hat{\rho}'_S$ be (normalized) states satisfying $\hat{\rho}'_S = \mathscr{E}_S(\hat{\rho}_S)$, where $\mathscr{E}_S$ is a Gibbs-sub-preserving map with respect to the initial and final Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$. Then, for $\alpha = 0, 1, \infty$,*

$$S_\alpha(\hat{\rho}_S \| e^{-\hat{H}_S}) \ge S_\alpha(\hat{\rho}'_S \| e^{-\hat{H}'_S}). \tag{8.90}$$

Proof We embed $\mathscr{E}_S$ into a CPTP Gibbs-preserving map $\mathscr{E}_{SC}$ with Hamiltonian $\hat{H}_{SC}$ in the same manner as Lemma 8.3. As in the proof of Theorem 8.1, we have $S_\alpha(\hat{\rho} \| \hat{\rho}^G_{SC}) = S_\alpha(\hat{\rho}_S \| e^{-\hat{H}_S}) + C$ and $S_\alpha(\hat{\rho}' \| \hat{\rho}^G_{SC}) = S_\alpha(\hat{\rho}'_S \| e^{-\hat{H}'_S}) + C$, where $\hat{\rho} := \hat{\rho}_S \otimes |0\rangle\langle 0|$, $\hat{\rho}' := \hat{\rho}'_S \otimes |1\rangle\langle 1|$, and $C := \ln(Z_S + Z'_S)$. Because $\hat{\rho}_S$ and $\hat{\rho}'_S$ are both normalized, $\mathscr{E}'_S(\hat{\rho}_S) = 0$ must be satisfied and thus $\hat{\rho}' = \mathscr{E}_{SC}(\hat{\rho})$. Therefore, the monotonicity of $\mathscr{E}_{SC}$ implies inequality (8.90). $\square$

We now explicitly recover the work storage W and consider the single-shot situations. From the above argument, we have the following.

Corollary 8.6 *A (normalized) state $\hat{\rho}_S$ is w-assisted single-shot Gibbs-preserving transformable to another (normalized) state $\hat{\rho}'_S$ with the initial and final Hamiltonians $\hat{H}_S$ and $\hat{H}'_S$ in the sense of Definition 8.4, if and only if there exists a Gibbs-sub-preserving map $\mathscr{E}_{SW}$ on SW such that*

$$\mathscr{E}_{SW} \left(\hat{\rho}_S \otimes |E_i\rangle\langle E_i| \right) = \hat{\rho}'_S \otimes |E_f\rangle\langle E_f|. \tag{8.91}$$

We note that the initial and final Hamiltonians of W can be different by applying the clock to W as well. We can thus restrict the Hilbert spaces of W for the initial and final states to one-dimensional; the initial (resp. final) Hilbert space of W is spanned only by $|E_i\rangle$ (resp. $|E_f\rangle$) with the initial and final Hamiltonians $\hat{H}_{W,i} := E_i|E_i\rangle\langle E_i|$ (resp. $\hat{H}_{W,f} := E_f|E_f\rangle\langle E_f|$). In this setup, the condition (8.84) of Gibbs-sub-preserving maps on SW is written as [45]

$$\mathscr{E}_S(e^{-\beta H_S}) \le e^{\beta w} e^{-\beta \hat{H}'_S}. \tag{8.92}$$

Meanwhile, the corresponding, trace-nonincreasing, notion of thermal operations can be defined as follows [129]. We again ignore W.

Definition 8.9 (*Generalized thermal operations*) A CP and trace-nonincreasing map $\mathscr{E}_S$ is a (generalized) thermal operation with the initial and final Hamiltonians $\hat{H}_S$ and

$\hat{H}_S'$, if there exists a heat bath B with Hamiltonian $\hat{H}_B$ and the corresponding Gibbs state $\hat{\rho}_B^G$, and exists a partial isometry $\hat{V}$ such that

$$\mathscr{E}_S(\hat{\rho}_S) = \text{tr}_B\left[\hat{V}\hat{\rho}_S \otimes \hat{\rho}_B^G \hat{V}^\dagger\right] \tag{8.93}$$

and

$$\hat{V}(\hat{H}_S + \hat{H}_B) = (\hat{H}_S' + \hat{H}_B)\hat{V}. \tag{8.94}$$

The corresponding non-exact (generalized) thermal operation is defined in the same manner as Definition 8.2.

Here, an operator $\hat{V}$ is a partial isometry if $\hat{V}\hat{V}^\dagger$ and $\hat{V}^\dagger\hat{V}$ are projectors. The condition (8.94) above implies that the sum of the energies of the system and the bath is conserved even when the Hamiltonian of the system is changed. In parallel to Lemma 8.3, we can also construct a TP thermal operation (with a unitary operator) of an extended system, starting from any trace-nonincreasing thermal operation (Proposition 13 of Ref. [129]).

It is also known that any trace-nonincreasing thermal operation $\mathscr{E}_S$ in the above sense is a trace-nonincreasing Gibbs-sub-preserving map that satisfies $\mathscr{E}_S(e^{-\beta\hat{H}_S}) \leq e^{-\beta\hat{H}_S'}$ (Lemma 1 of [129]). We note that if $\hat{V}$ is unitary and the thermal operation is TP, we have $\mathscr{E}_S(e^{-\beta\hat{H}_S}) = e^{-\beta H_S'}$.

Appendix A
General Quantum Divergences and Their Monotonicity

We prove the monotonicity of quantum divergences discussed in Chap. 5 from a general point of view, by introducing a class of general quantum divergence-like quantities called the Petz's quasi-entropies [16, 17, 112, 113]. The classical counterpart is presented mainly in Sects. 2.4 and 2.5. In this Appendix, superoperator $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ allows different input and output spaces.

We start with proving some operator inequalities in Sect. A.1, and discuss operator convex and operator monotone in Sect. A.2. In Sect. A.3, we prove the monotonicity of general divergence-like quantities. In addition, we prove the monotonicity of the quantum Fisher information in Sect. A.4.

A.1 Some Operator Inequalities

As a preliminary, we here note some useful operator inequalities.

Lemma A.1 $\hat{A} \leq \hat{B}$ *implies* $\hat{A}^{-1} \geq \hat{B}^{-1}$ *for* $\hat{A} > 0$ *and* $\hat{B} > 0$.

Proof $\hat{B}^{-1/2}\hat{A}\hat{B}^{-1/2} \leq \hat{I}$ implies $\hat{B}^{1/2}\hat{A}^{-1}\hat{B}^{1/2} = (\hat{B}^{-1/2}\hat{A}\hat{B}^{-1/2})^{-1} \geq \hat{I}$. $\qquad\square$

Lemma A.2 *Suppose that* $\hat{Z}$ *is positive definite. An operator-valued matrix*

$$\begin{bmatrix} \hat{X} & \hat{Y} \\ \hat{Y}^\dagger & \hat{Z} \end{bmatrix} \tag{A.1}$$

is positive, if and only if

$$\hat{X} \geq \hat{Y}\hat{Z}^{-1}\hat{Y}^\dagger. \tag{A.2}$$

Proof This can be seen from

$$\begin{bmatrix} \hat{I} & -\hat{Y}\hat{Z}^{-1} \\ 0 & \hat{I} \end{bmatrix} \begin{bmatrix} \hat{X} & \hat{Y} \\ \hat{Y}^\dagger & \hat{Z} \end{bmatrix} \begin{bmatrix} \hat{I} & -\hat{Y}\hat{Z}^{-1} \\ 0 & \hat{I} \end{bmatrix}^\dagger = \begin{bmatrix} \hat{X} - \hat{Y}\hat{Z}^{-1}\hat{Y}^\dagger & 0 \\ 0 & \hat{Z} \end{bmatrix}. \qquad (A.3)$$

$\square$

Proposition A.1 (Kadison's inequality, Lemma 3.5 of [112]) *Let $\mathscr{E}$ be a positive superoperator. Let $\hat{X}$ be Hermitian and $\hat{Y}$, $\mathscr{E}(\hat{Y})$ be positive definite. Then,*

$$\mathscr{E}(\hat{X}\hat{Y}^{-1}\hat{X}) \geq \mathscr{E}(\hat{X})\mathscr{E}(\hat{Y})^{-1}\mathscr{E}(\hat{X}). \qquad (A.4)$$

In particular, if $\mathscr{E}$ is positive and unital,

$$\mathscr{E}(\hat{X}^2) \geq \mathscr{E}(\hat{X})^2. \qquad (A.5)$$

Proof Let $\hat{X} = \sum_k x_k |\varphi_k\rangle\langle\varphi_k|$ be the spectral decomposition of $\hat{X}$. We first consider the case that $\hat{Y} = \hat{I}$ and define

$$\hat{X}' := \begin{bmatrix} \mathscr{E}(\hat{X}^2) & \mathscr{E}(\hat{X}) \\ \mathscr{E}(\hat{X}) & \mathscr{E}(\hat{I}) \end{bmatrix} = \sum_k \begin{bmatrix} x_k^2 & x_k \\ x_k & 1 \end{bmatrix} \otimes \mathscr{E}(|\varphi_k\rangle\langle\varphi_k|). \qquad (A.6)$$

The right-hand side is positive, because the 2×2 matrix and $\mathscr{E}(|\varphi_k\rangle\langle\varphi_k|)$ are both positive. Thus, from Lemma A.2, we obtain

$$\mathscr{E}(\hat{X}^2) \geq \mathscr{E}(\hat{X})\mathscr{E}(\hat{I})^{-1}\mathscr{E}(\hat{X}). \qquad (A.7)$$

If $\hat{Y} \neq \hat{I}$, we replace $\hat{X}$ by $\hat{Y}^{-1/2}\hat{X}\hat{Y}^{-1/2}$ and $\mathscr{E}(*)$ by $\mathscr{E}(\hat{Y}^{1/2} * \hat{Y}^{1/2})$ (that is also positive), and then obtain inequality (A.4). $\square$

Corollary A.1 *Let $\mathscr{E}$ be positive and TP. Let $\hat{X}$ be Hermitian and $\hat{Y}$, $\mathscr{E}(\hat{Y})$ be positive definite. Then,*

$$\mathrm{tr}[\hat{X}^2\hat{Y}^{-1}] \geq \mathrm{tr}[\mathscr{E}(\hat{X})^2\mathscr{E}(\hat{Y})^{-1}]. \qquad (A.8)$$

Proof Take the trace of inequality (A.4). $\square$

The quantum Rényi 2-divergence introduced in Eq. (A.35) is given by $\tilde{S}_2(\hat{\rho}\|\hat{\sigma}) := \ln\left(\mathrm{tr}[\hat{\rho}^2\hat{\sigma}^{-1}]\right)$. Thus, inequality (A.8) implies the monotonicity of $\tilde{S}_2(\hat{\rho}\|\hat{\sigma})$.

The above proof of Proposition A.1 does not work in general if $\hat{X}$ is not Hermitian. If $\hat{X}$ is not necessarily Hermitian, we have the following proposition by additionally assuming that $\mathscr{E}$ is 2-positive.

Proposition A.2 (Schwarz's operator inequality [175]) *Let $\mathscr{E}$ be 2-positive and let $\hat{Y}$, $\mathscr{E}(\hat{Y})$ be positive definite. Then,*

$$\mathscr{E}(\hat{X}^\dagger \hat{Y}^{-1} \hat{X}) \geq \mathscr{E}(\hat{X})^\dagger \mathscr{E}(\hat{Y})^{-1} \mathscr{E}(\hat{X}). \tag{A.9}$$

In particular, if $\mathscr{E}$ is 2-positive and unital,

$$\mathscr{E}(\hat{X}^\dagger \ \hat{X}) \geq \mathscr{E}(\hat{X})^\dagger \mathscr{E}(\hat{X}). \tag{A.10}$$

Proof We define

$$\hat{X}' := \begin{bmatrix} \hat{Y} & \hat{X} \\ \hat{X}^\dagger & \hat{X}^\dagger \hat{Y}^{-1} \hat{X} \end{bmatrix}. \tag{A.11}$$

Because $\mathscr{E}$ is 2-positive, $\mathscr{E} \otimes \mathscr{I}_2$ is positive, and thus

$$(\mathscr{E} \otimes \mathscr{I}_2)(\hat{X}') = \begin{bmatrix} \mathscr{E}(\hat{Y}) & \mathscr{E}(X) \\ \mathscr{E}(\hat{X})^\dagger & \mathscr{E}(\hat{X}^\dagger \hat{Y}^{-1} \hat{X}) \end{bmatrix} \tag{A.12}$$

is positive. From Lemma A.2, we obtain inequality (A.9). □

A.2 Operator Convex and Operator Monotone

We consider operator convex/concave functions and operator monotone functions. The operator convexity and the operator monotonicity are much stronger properties than the ordinary convexity and the ordinary monotonicity of functions. In this section, we will omit proofs of several important theorems; see Refs. [7, 8, 112] for details.

Definition A.1 (*Operator convex/concave functions*) A function $f : (0, \infty) \to \mathbb{R}$ is operator convex, if for any positive-definite operators $\hat{X}$, $\hat{Y}$ and any $0 \leq \lambda \leq 1$,

$$f(\lambda \hat{X} + (1 - \lambda) \hat{Y}) \leq \lambda f(\hat{X}) + (1 - \lambda) f(\hat{Y}) \tag{A.13}$$

holds. If $\leq$ above is replaced by $\geq$, f is operator concave.

Obviously, f is operator convex if and only if $-f$ is operator concave. Since every convex/concave function on any open interval is continuous, every operator convex/concave function is automatically continuous on $(0, \infty)$.

Definition A.2 (*Operator monotone functions*) A function $f : (0, \infty) \to \mathbb{R}$ is operator monotone, if for any positive-definite operators $\hat{X}$, $\hat{Y}$ with $\hat{X} \leq \hat{Y}$, $f(\hat{X}) \leq f(\hat{Y})$ holds. If $f(\hat{X}) \geq f(\hat{Y})$ holds, f is operator decreasing-monotone.

Obviously, f is operator monotone if and only if $-f$ is operator decreasing-monotone. We note the following propositions.

Proposition A.3 (Theorem 2.4 of [176]) *Let $f : (0, \infty) \to \mathbb{R}$ be a continuous function and suppose that $f(0) := \lim_{x \to +0} f(x) \in (-\infty, 0]$ exists. Then, f is operator convex if and only if $g(x) := x^{-1} f(x)$ is operator monotone on $(0, \infty)$.*

Proposition A.4 (Theorem 2.5 of [176]) *Let $f : (0, \infty) \to (-\infty, 0]$ be a continuous function and suppose that $f(0) := \lim_{x \to +0} f(x) \in (-\infty, 0]$ exists. Then, f is operator convex if and only if it is operator decreasing-monotone.*

It is known that the full characterizations of operator convex functions and operator monotone functions are given by some integral representations, which is often referred to as the Löwner's theorem [7, 8, 112]. While there are several variants of such integral representations, we here state one of them for operator convex functions.

Theorem A.1 (Theorem 8.1 of [112]) *Suppose that $f : (0, \infty) \to \mathbb{R}$ is a continuous function and $f(0) := \lim_{x \to +0} f(x) \in \mathbb{R}$ exists. f is operator convex, if and only if there exist $a \in \mathbb{R}$, $b \geq 0$, and a non-negative measure μ on $(0, \infty)$ satisfying $\int_{(0,\infty)} (1 + t)^{-2} d\mu(t) < \infty$, such that*

$$f(x) = f(0) + ax + bx^2 + \int_{(0,\infty)} \left(\frac{x}{1+t} - \frac{x}{x+t} \right) d\mu(t). \qquad \text{(A.14)}$$

Moreover, a, b, and μ are uniquely determined by f, and

$$b = \lim_{x \to \infty} \frac{f(x)}{x^2}, \quad a = f(1) - f(0) - b. \qquad \text{(A.15)}$$

A simple example of the integral representation (A.14) is given by

$$x \ln x = \int_0^\infty \left(\frac{x}{1+t} - \frac{x}{x+t} \right) dt. \qquad \text{(A.16)}$$

We here list examples of operator convex/concave functions and operator monotone functions [7, 8, 112].

Proposition A.5 (Löwner-Heinz Theorem) *On $(0, \infty)$,*

(a) *For $0 \leq t < \infty$, $f(x) = (x + t)^{-1}$ is operator convex and operator decreasing-monotone.*
(b) *For $0 < t < \infty$, $f(x) = x/(x + t)$ is operator concave and operator monotone.*
(c) *$f(x) = -\ln x$ is operator convex and operator decreasing-monotone.*
(d) *$f(x) = x \ln x$ is operator convex (but not (decreasing-)monotone).*
(e) *For $0 \leq \alpha \leq 1$, $f(x) = x^\alpha$ is operator concave and operator monotone.*

(f) For $1 < \alpha \leq 2$, $f(x) = x^\alpha$ is operator convex (but not operator monotone).

Proof Here we only prove (a)–(d).

We first show (a). From the convexity of x^{-1}, we have $(\lambda \hat{Y}^{-1/2} \hat{X} \hat{Y}^{-1/2} + (1 - \lambda) \hat{I})^{-1} \leq \lambda \hat{Y}^{1/2} \hat{X}^{-1} \hat{Y}^{1/2} + (1 - \lambda) \hat{I}$, which implies $(\lambda \hat{X} + (1 - \lambda) \hat{Y})^{-1} \leq \lambda \hat{X}^{-1} + (1 - \lambda) \hat{Y}^{-1}$. By shifting $\hat{X}$ and $\hat{Y}$ by $t\hat{I}$, we prove the operator convexity. The operator decreasing-monotonicity is obvious, because $\hat{X} + t\hat{I} \leq \hat{Y} + t\hat{I}$ implies $(\hat{X} + t\hat{I})^{-1} \geq (\hat{Y} + t\hat{I})^{-1}$.

Then, (b) follows from (a) and $x/(x + t) = 1 - t/(x + t)$; (c) follows from (a) and $-\ln x = \int_0^\infty \left((x + t)^{-1} - (1 + t)^{-1} \right) dt$; (d) follows from (b) and Eq. (A.16).

(e) is Theorem V.1.9 of Ref. [7] or Example 8.3 of Ref. [112] or Example 2.5.9 of Ref. [8]; (f) is Example 8.3 of Ref. [112] or Example 2.5.9 of Ref. [8]. Here we only note that the operator convexity of $f(x) = x^2$ is obvious from $\lambda \hat{X}^2 + (1 - \lambda) \hat{Y}^2 - (\lambda \hat{X} + (1 - \lambda) \hat{Y})^2 = \lambda(1 - \lambda)(\hat{X} - \hat{Y})^2 \geq 0$ (Example V.1.3 of [7]). See also Ref. [7] for that $f(x) = x^2$ is not operator monotone (Example V.1.2).

We finally note that $f(x) = x^\alpha$ for $2 < \alpha < \infty$ is not operator convex (Example 2.5.9 of Ref. [8]; see also Example V.1.4 of Ref. [7]). □

We next show a useful property of operator convex functions (see also Theorem V.2.3 of [7]).

Proposition A.6 (Jensen's operator inequality; Theorem 2.1 of [176]) *Consider $f : (0, \infty) \to \mathbb{R}$ and suppose that $f(0) := \lim_{x \to +0} f(x) \in \mathbb{R}$ exists. Then, the following are equivalent.*
(i) f is operator convex on $(0, \infty)$ and $f(0) \leq 0$.
(ii) For any $\hat{X} \geq 0$ and any contraction $\hat{V}$,

$$f(\hat{V}^\dagger \hat{X} \hat{V}) \leq \hat{V}^\dagger f(\hat{X}) \hat{V}. \tag{A.17}$$

Here, a linear operator $\hat{V}$ is called a contraction, if it satisfies $\langle \varphi | \hat{V}^\dagger \hat{V} | \varphi \rangle \leq \langle \varphi | \varphi \rangle$ for all $|\varphi\rangle$.

Proof Here we only prove (i) $\Rightarrow$ (ii) (see Ref. [7] for the converse). Note that $\hat{V}^\dagger \hat{V} \leq \hat{I}$ and $\hat{V} \hat{V}^\dagger \leq \hat{I}$ because $\hat{V}$ is a contraction. We can define $\hat{U} := (I - \hat{V} \hat{V}^\dagger)^{1/2}$ and $\hat{U}' := (I - \hat{V}^\dagger \hat{V})^{1/2}$. Then we consider

$$\hat{X}' := \begin{bmatrix} \hat{X} & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{V}_1 := \begin{bmatrix} \hat{V} & \hat{U} \\ \hat{U}' & -\hat{V}^\dagger \end{bmatrix}, \quad \hat{V}_2 := \begin{bmatrix} \hat{V} & -\hat{U} \\ \hat{U}' & \hat{V}^\dagger \end{bmatrix}. \tag{A.18}$$

By using the singular-value decomposition of $\hat{V}$, we can see that $\hat{V}_1$ and $\hat{V}_2$ are unitary. We compute

$$\hat{V}_1^\dagger \hat{X}' \hat{V}_1 = \begin{bmatrix} \hat{V}^\dagger \hat{X} \hat{V} & \hat{V}^\dagger \hat{X} \hat{U} \\ \hat{U} \hat{X} \hat{V} & \hat{U} \hat{X} \hat{U} \end{bmatrix}, \quad \hat{V}_2^\dagger \hat{X}' \hat{V}_2 = \begin{bmatrix} \hat{V}^\dagger \hat{X} \hat{V} & -\hat{V}^\dagger \hat{X} \hat{U} \\ -\hat{U} \hat{X} \hat{V} & \hat{U} \hat{X} \hat{U} \end{bmatrix}, \tag{A.19}$$

and thus obtain

$$\frac{\hat{V}_1^\dagger \hat{X}' \hat{V}_1 + \hat{V}_2^\dagger \hat{X}' \hat{V}_2}{2} = \begin{bmatrix} \hat{V}^\dagger \hat{X} \hat{V} & 0 \\ 0 & \hat{U} \hat{X} \hat{U} \end{bmatrix}. \qquad (A.20)$$

From the operator convexity of $f(x)$, we have

$$f\left(\frac{\hat{V}_1^\dagger \hat{X}' \hat{V}_1 + \hat{V}_2^\dagger \hat{X}' \hat{V}_2}{2}\right) \le \frac{f(\hat{V}_1^\dagger \hat{X}' \hat{V}_1) + f(\hat{V}_2^\dagger \hat{X}' \hat{V}_2)}{2} = \frac{\hat{V}_1^\dagger f(\hat{X}') \hat{V}_1 + \hat{V}_2^\dagger f(\hat{X}') \hat{V}_2}{2},$$
$$(A.21)$$

which leads to

$$\begin{bmatrix} f(\hat{V}^\dagger \hat{X} \hat{V}) & 0 \\ 0 & f(\hat{U} \hat{X} \hat{U}) \end{bmatrix} \le \begin{bmatrix} \hat{V}^\dagger f(\hat{X}) \hat{V} + \hat{U}' f(0) \hat{U}' & 0 \\ 0 & \hat{U} f(\hat{X}) \hat{U} + \hat{V} f(0) \hat{V}^\dagger \end{bmatrix}.$$
$$(A.22)$$

From $f(0) \le 0$, we finally obtain

$$\begin{bmatrix} f(\hat{V}^\dagger \hat{X} \hat{V}) & 0 \\ 0 & f(\hat{U} \hat{X} \hat{U}) \end{bmatrix} \le \begin{bmatrix} \hat{V}^\dagger f(\hat{X}) \hat{V} & 0 \\ 0 & \hat{U} f(\hat{X}) \hat{U} \end{bmatrix}, \qquad (A.23)$$

which implies (ii). $\qquad \square$

A.3 General Monotonicity Properties

We now go to the main part of this Appendix: the proof of the monotonicity. In this section, we basically assume that $\hat{\rho}$, $\hat{\sigma}$, $\mathscr{E}(\hat{\rho})$, $\mathscr{E}(\hat{\sigma})$ are all positive definite, while $\hat{\rho}$ and $\mathscr{E}(\hat{\rho})$ can be just positive if $f(0) := \lim_{x \to +0} f(x) \in \mathbb{R}$ exists for function $f : (0, \infty) \to \mathbb{R}$ introduced below. We do not necessarily assume that these operators are normalized, unless stated otherwise.

To introduce general divergence-like quantities, we first define the left and right multiplications of $\hat{\rho}$:

$$\mathscr{L}_{\hat{\rho}}(\hat{X}) := \hat{\rho} \hat{X}, \quad \mathscr{R}_{\hat{\rho}}(\hat{X}) := \hat{X} \hat{\rho}. \qquad (A.24)$$

Here, $\mathscr{L}_{\hat{\rho}}$ and $\mathscr{R}_{\hat{\rho}}$ are commutable and are both Hermitian with respect to the Hilbert-Schmidt inner product. Then, we introduce the modular operator $\mathscr{D}_{\hat{\rho},\hat{\sigma}}$ for $\hat{\rho}$, $\hat{\sigma}$ by

$$\mathscr{D}_{\hat{\rho},\hat{\sigma}}(\hat{X}) := \mathscr{L}_{\hat{\rho}} \mathscr{R}_{\hat{\sigma}^{-1}}(\hat{X}) = \hat{\rho} \hat{X} \hat{\sigma}^{-1}, \qquad (A.25)$$

which is also Hermitian.

Let $\hat{\rho} = \sum_i p_i \hat{P}_i$ and $\hat{\sigma} = \sum_i q_i \hat{Q}_i$ be the spectrum decompositions, where $\hat{P}_i$, $\hat{Q}_i$ are rank-one projectors. The spectrum decomposition of $\mathscr{D}_{\hat{\rho},\hat{\sigma}}$ is then given by

$$\mathscr{D}_{\hat{\rho},\hat{\sigma}} = \sum_{ij} \frac{p_i}{q_j} \mathscr{P}_{ij}, \qquad (A.26)$$

where $\mathscr{P}_{ij}$ is a projection superoperator defined as $\mathscr{P}_{ij}(\hat{X}) := \hat{P}_i \hat{X} \hat{Q}_j$.

Let $f : (0, \infty) \to \mathbb{R}$ be a function. We can define $f(\mathscr{D}_{\hat{\rho}, \hat{\sigma}})$ by

$$f(\mathscr{D}_{\hat{\rho}, \hat{\sigma}}) = \sum_{ij} f\left(\frac{p_i}{q_j}\right) \mathscr{P}_{ij}, \tag{A.27}$$

or equivalently

$$f(\mathscr{D}_{\hat{\rho}, \hat{\sigma}})(\hat{X}) = \sum_{ij} f\left(\frac{p_i}{q_j}\right) \hat{P}_i \hat{X} \hat{Q}_j. \tag{A.28}$$

Now, we consider a divergence-like quantity

$$D_f(\hat{\rho}\|\hat{\sigma}) := \langle \hat{\sigma}^{1/2}, f(\mathscr{D}_{\hat{\rho}, \hat{\sigma}})(\hat{\sigma}^{1/2})\rangle_{\mathrm{HS}} = \sum_{ij} q_j f\left(\frac{p_i}{q_j}\right) \mathrm{tr}[\hat{P}_i \hat{Q}_j], \tag{A.29}$$

which is a special case of the Petz's quasi-entropies [16, 17] (see also Refs. [5, 112]). We discuss two special examples of this quantity:

Quantum KL divergence. The first example is $f(x) = x \ln x$, which is operator convex. By noting that $f(\mathscr{D}_{\hat{\rho}, \hat{\sigma}}) = \mathscr{L}_{\hat{\rho}} \mathscr{R}_{\hat{\sigma}^{-1}}(\ln \mathscr{L}_{\hat{\rho}} + \ln \mathscr{R}_{\hat{\sigma}^{-1}})$ and $\mathscr{L}_{\hat{\rho}}$ and $\mathscr{R}_{\hat{\sigma}^{-1}}$ are commutable, we obtain

$$f(\mathscr{D}_{\hat{\rho}, \hat{\sigma}})(\hat{X}) = (\hat{\rho} \ln \hat{\rho})\hat{X}\hat{\sigma}^{-1} - \hat{\rho}\hat{X}(\hat{\sigma}^{-1} \ln \hat{\sigma}), \tag{A.30}$$

which leads to

$$D_f(\hat{\rho}\|\hat{\sigma}) = \mathrm{tr}[\hat{\sigma}^{1/2}(\hat{\rho} \ln \hat{\rho})\hat{\sigma}^{1/2}\hat{\sigma}^{-1}] - \mathrm{tr}[\hat{\sigma}^{1/2}\hat{\rho}\hat{\sigma}^{1/2}(\hat{\sigma}^{-1} \ln \hat{\sigma})] = S_1(\hat{\rho}\|\hat{\sigma}). \tag{A.31}$$

Note that Eq. (A.29) implies

$$D_f(\hat{\rho}\|\hat{\sigma}) = \sum_{ij} p_i \ln \frac{p_i}{q_j} \mathrm{tr}[\hat{P}_i \hat{Q}_j]. \tag{A.32}$$

Quantum Rényi divergence. The second example is $f_\alpha(x) := x^\alpha$ with $0 < \alpha < 1$ and $1 < \alpha < \infty$. By noting that $f_\alpha(\mathscr{D}_{\hat{\rho}, \hat{\sigma}})(\hat{X}) = \hat{\rho}^\alpha \hat{X}\hat{\sigma}^{-\alpha}$, we have

$$D_{f_\alpha}(\hat{\rho}\|\hat{\sigma}) = \mathrm{tr}[\hat{\sigma}^{1/2}\hat{\rho}^\alpha \hat{\sigma}^{1/2}\hat{\sigma}^{-\alpha}] = \mathrm{tr}[\hat{\rho}^\alpha \hat{\sigma}^{1-\alpha}]. \tag{A.33}$$

We note that, from Eq. (A.29),

$$D_{f_\alpha}(\hat{\rho}\|\hat{\sigma}) = \sum_{ij} \frac{p_i^\alpha}{q_j^{\alpha-1}} \mathrm{tr}[\hat{P}_i \hat{Q}_j]. \tag{A.34}$$

Then, we define a simple version of the quantum Rényi α-divergence for $0 < \alpha < 1$ and $1 < \alpha < \infty$ as [112, 114]

$$\tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) := \frac{1}{\alpha - 1} \ln D_{f_\alpha}(\hat{\rho}\|\hat{\sigma}) = \frac{1}{\alpha - 1} \ln \left(\mathrm{tr}[\hat{\rho}^\alpha \hat{\sigma}^{1-\alpha}] \right). \tag{A.35}$$

It is straightforward to see that

$$\tilde{S}_0(\hat{\rho}\|\hat{\sigma}) := \lim_{\alpha \to +0} \tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) = S_0(\hat{\rho}\|\hat{\sigma}), \tag{A.36}$$

and we can also show that

$$\tilde{S}_1(\hat{\rho}\|\hat{\sigma}) := \lim_{\alpha \to 1} \tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) = S_1(\hat{\rho}\|\hat{\sigma}). \tag{A.37}$$

Thus $\tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma})$ is well-defined for $0 \leq \alpha < \infty$. On the other hand, $\lim_{\alpha \to \infty} \tilde{S}_\alpha$ $(\hat{\rho}\|\hat{\sigma})$ does not equal $S_\infty(\hat{\rho}\|\hat{\sigma})$. We finally note that, if $\hat{\rho}$ is normalized,

$$\tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) \leq \tilde{S}_{\alpha'}(\hat{\rho}\|\hat{\sigma}) \text{ for } \alpha \leq \alpha', \tag{A.38}$$

which can be proved in the same manner as the proof of Proposition 2.3, by noting that $\sum_{ij} p_i \mathrm{tr}[\hat{P}_i \hat{Q}_j] = 1$.

We now investigate the fundamental properties of $D_f(\hat{\rho}\|\hat{\sigma})$. First, we consider the following theorem, from which the non-negativity of the quantum KL divergence follows.

Theorem A.2 *Suppose that $\hat{\rho}(> 0)$, $\hat{\sigma}(> 0)$ are both normalized. If f is convex on $(0, \infty)$ and strictly-convex at $x = 1$,*

$$D_f(\hat{\rho}\|\hat{\sigma}) \geq f(1), \tag{A.39}$$

where the equality holds if and only if $\hat{\rho} = \hat{\sigma}$. If f is concave on $(0, \infty)$ and strictly-concave at $x = 1$, the opposite inequality holds.

Proof Consider the expression of $D_f(\hat{\rho}\|\hat{\sigma})$ of Eq. (A.29). Because $T_{ij} := \mathrm{tr}[\hat{P}_i \hat{Q}_j]$ is a doubly stochastic matrix and f is convex, we have $\sum_i f(p_i/q_j) T_{ij} \geq f(p'_j/q_j)$ with $p'_j := \sum_i p_i T_{ij}$ being a probability distribution. We thus have

$$D_f(\hat{\rho}\|\hat{\sigma}) \geq D_f(p'\|q), \tag{A.40}$$

where $D_f(p'\|q) := \sum_i q_i f(p'_i/q_i)$. Also, as is the case for inequality (2.35), we have

$$D_f(p'\|q) \geq f(1), \tag{A.41}$$

where the equality holds if and only if $p' = q$, because of the strict convexity of f at $x = 1$. In the case that $p' = q$ holds, the equality of (A.40) holds if and only if $T_{ij} = 0$ for (i, j) with $p_i \neq p_j$, again from the strict convexity of f at $x = 1$. Thus, $D_f(\hat{\rho}\|\hat{\sigma}) = f(1)$ holds if and only if $\hat{\rho} = \hat{\sigma}$. We can prove the concave case in the same manner. □

We emphasize that we only used the ordinary convexity, not the operator convexity, of f (and the strict convexity at $x = 1$) in the above proof. The following are two special cases.

Corollary A.2 *Suppose that $\hat{\rho}(\geq 0)$, $\hat{\sigma}(> 0)$ are both normalized. The quantum KL divergence satisfies $S_1(\hat{\rho}\|\hat{\sigma}) \geq 0$ [inequality (5.16)], where the equality holds if and only if $\hat{\rho} = \hat{\sigma}$.*

Proof Take $f(x) = x \ln x$ with $f(1) = 0$, which is convex on $(0, \infty)$ and strictly-convex at $x = 1$. □

Corollary A.3 *Suppose that $\hat{\rho}(\geq 0)$, $\hat{\sigma}(> 0)$ are both normalized. For $0 < \alpha < 1$ and $1 < \alpha < \infty$, the quantum Rényi divergence satisfies $\tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) \geq 0$, where the equality holds if and only if $\hat{\rho} = \hat{\sigma}$.*

Proof We take $f(x) = x^\alpha$ with $f(1) = 1$, which is concave (convex) on $(0, \infty)$ and strictly-concave (convex) at $x = 1$ for $0 < \alpha < 1$ ($1 < \alpha < \infty$). By noting the sign of $\alpha - 1$, we prove the claim. □

We now consider the monotonicity, where $\hat{\rho}$, $\hat{\sigma}$ can be unnormalized.

Theorem A.3 (Monotonicity, Theorem 4.3 of [112]) *Let $\hat{\rho}$, $\mathcal{E}(\hat{\rho})$ be positive and $\hat{\sigma}$, $\mathcal{E}(\hat{\sigma})$ be positive definite. Suppose that $f : (0, \infty) \to \mathbb{R}$ is operator convex and $f(0) := \lim_{x \to +0} f(x) \in \mathbb{R}$ exists. Let $\mathcal{E}$ be 2-positive and TP. Then,*

$$D_f(\hat{\rho}\|\hat{\sigma}) \geq D_f(\mathcal{E}(\hat{\rho})\|\mathcal{E}(\hat{\sigma})). \tag{A.42}$$

If f is operator concave, the opposite inequality holds.

To prove this theorem, we prepare the following lemma.

Lemma A.3 (Lemma 4.2 of [112]) *Let $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ be 2-positive and TP. Define $\mathcal{V} : \mathcal{L}(\mathcal{H}') \to \mathcal{L}(\mathcal{H})$ by*

$$\mathcal{V} := \mathcal{R}_{\hat{\sigma}}^{1/2} \mathcal{E}^\dagger \mathcal{R}_{\mathcal{E}(\hat{\sigma})}^{-1/2}, \tag{A.43}$$

or equivalently,

$$\mathcal{V}(\hat{X}) := \mathcal{E}^\dagger(\hat{X}\mathcal{E}(\hat{\sigma})^{-1/2})\hat{\sigma}^{1/2} \quad \Leftrightarrow \quad \mathcal{V}(\hat{X}\mathcal{E}(\hat{\sigma})^{1/2}) = \mathcal{E}^\dagger(\hat{X})\hat{\sigma}^{1/2}. \tag{A.44}$$

Then, $\mathcal{V}$ is a contraction with respect to the Hilbert-Schmidt inner product. Moreover,

$$\mathscr{V}^\dagger \mathscr{D}_{\hat\rho,\hat\sigma} \mathscr{V} \le \mathscr{D}_{\mathscr{E}(\hat\rho),\mathscr{E}(\hat\sigma)}. \tag{A.45}$$

Proof Because $\mathscr{E}^\dagger$ that is 2-positive and unital, we apply the Schwarz's operator inequality (A.10) of Proposition A.2 and obtain

$$\langle \hat{X}\mathscr{E}(\hat\sigma)^{1/2}, \hat{X}\mathscr{E}(\hat\sigma)^{1/2}\rangle_{\mathrm{HS}} = \mathrm{tr}[\hat\sigma\mathscr{E}^\dagger(\hat{X}^\dagger\hat{X})] \tag{A.46}$$

$$\ge \mathrm{tr}[\hat\sigma\mathscr{E}^\dagger(\hat{X})^\dagger\mathscr{E}^\dagger(\hat{X})] = \langle \mathscr{E}^\dagger(\hat{X})\hat\sigma^{1/2}, \mathscr{E}^\dagger(\hat{X})\hat\sigma^{1/2}\rangle_{\mathrm{HS}}, \tag{A.47}$$

which implies that $\mathscr{V}$ is a contraction.

Inequality (A.45) is shown as

$$\langle \hat{X}\mathscr{E}(\hat\sigma)^{1/2}, \mathscr{V}^\dagger\mathscr{D}_{\hat\rho,\hat\sigma}\mathscr{V}(\hat{X}\mathscr{E}(\hat\sigma)^{1/2})\rangle_{\mathrm{HS}} \tag{A.48}$$

$$= \langle \mathscr{V}(\hat{X}\mathscr{E}(\hat\sigma)^{1/2}), \mathscr{D}_{\hat\rho,\hat\sigma}\mathscr{V}(\hat{X}\mathscr{E}(\hat\sigma)^{1/2})\rangle_{\mathrm{HS}} \tag{A.49}$$

$$= \langle \mathscr{E}^\dagger(\hat{X})\hat\sigma^{1/2}, \mathscr{D}_{\hat\rho,\hat\sigma}(\mathscr{E}^\dagger(\hat{X})\hat\sigma^{1/2})\rangle_{\mathrm{HS}} \tag{A.50}$$

$$= \mathrm{tr}\left[\hat\sigma^{1/2}\mathscr{E}^\dagger(\hat{X}^\dagger)\mathscr{D}_{\hat\rho,\hat\sigma}(\mathscr{E}^\dagger(\hat{X})\hat\sigma^{1/2})\right] \tag{A.51}$$

$$= \mathrm{tr}\left[\hat\rho\mathscr{E}^\dagger(\hat{X})\mathscr{E}^\dagger(\hat{X}^\dagger)\right] \tag{A.52}$$

$$\le \mathrm{tr}\left[\hat\rho\mathscr{E}^\dagger(\hat{X}\hat{X}^\dagger)\right] \tag{A.53}$$

$$= \mathrm{tr}\left[\mathscr{E}(\hat\rho)\hat{X}\hat{X}^\dagger\right] \tag{A.54}$$

$$= \mathrm{tr}[\mathscr{E}(\hat\sigma)^{1/2}\hat{X}^\dagger\mathscr{E}(\hat\rho)\hat{X}\mathscr{E}(\hat\sigma)^{1/2}\mathscr{E}(\hat\sigma)^{-1}] \tag{A.55}$$

$$= \langle \hat{X}\mathscr{E}(\hat\sigma)^{1/2}, \mathscr{D}_{\mathscr{E}(\hat\rho),\mathscr{E}(\hat\sigma)}(\hat{X}\mathscr{E}(\hat\sigma)^{1/2})\rangle_{\mathrm{HS}}, \tag{A.56}$$

where we again used the Schwarz's operator inequality (A.10) for $\mathscr{E}^\dagger$. □

Proof of Theorem A.3. First, we assume that f is operator convex and operator decreasing-monotone, and satisfies $f(0) \le 0$. Consider $\mathscr{V}$ defined in Lemma A.3.

By applying the Jensen's operator inequality (A.17) in Proposition A.6 to f and contraction $\mathscr{V}$, we have

$$f(\mathscr{V}^\dagger\mathscr{D}_{\hat\rho,\hat\sigma}\mathscr{V}) \le \mathscr{V}^\dagger f(\mathscr{D}_{\hat\rho,\hat\sigma})\mathscr{V}. \tag{A.57}$$

Also, by substituting inequality (A.45) in Lemma A.3 to operator-decreasing f, we have

$$f(\mathscr{D}_{\mathscr{E}(\hat\rho),\mathscr{E}(\hat\sigma)}) \le f(\mathscr{V}^\dagger\mathscr{D}_{\hat\rho,\hat\sigma}\mathscr{V}). \tag{A.58}$$

By combining inequalities (A.57) and (A.58), we obtain

$$f(\mathscr{D}_{\mathscr{E}(\hat\rho),\mathscr{E}(\hat\sigma)}) \le \mathscr{V}^\dagger f(\mathscr{D}_{\hat\rho,\hat\sigma})\mathscr{V}. \tag{A.59}$$

Thus,

$$\langle \mathcal{E}(\hat{\sigma})^{1/2}, f(\mathcal{D}_{\mathcal{E}(\hat{\rho}), \mathcal{E}(\hat{\sigma})})(\mathcal{E}(\hat{\sigma})^{1/2}) \rangle_{\mathrm{HS}} \leq \langle \mathcal{E}(\hat{\sigma})^{1/2}, \mathcal{V}^{\dagger} f(\mathcal{D}_{\hat{\rho}, \hat{\sigma}}) \mathcal{V} (\mathcal{E}(\hat{\sigma})^{1/2}) \rangle_{\mathrm{HS}},$$
(A.60)

where the left-hand side equals $D_f(\mathcal{E}(\hat{\rho}) \| \mathcal{E}(\hat{\sigma}))$ and the right-hand side equals $D_f(\hat{\rho} \| \hat{\sigma})$ because of $\mathcal{V}(\mathcal{E}(\hat{\sigma})^{1/2}) = \hat{\sigma}^{1/2}$. Thus, we obtain

$$D_f(\mathcal{E}(\hat{\rho}) \| \mathcal{E}(\hat{\sigma})) \leq D_f(\hat{\rho} \| \hat{\sigma}).$$
(A.61)

For general f that is just operator convex, we invoke the integral representation (A.14), which leads to

$$D_f(\hat{\rho} \| \hat{\sigma}) = \mathrm{tr}[\hat{\sigma}] f(0) + \mathrm{tr}[\hat{\rho}] a + \mathrm{tr}[\hat{\rho}^2 \hat{\sigma}^{-1}] b + \int_{(0,\infty)} \left(\frac{\mathrm{tr}[\hat{\rho}]}{1+t} + D_{\varphi_t}(\hat{\rho} \| \hat{\sigma}) \right) d\mu(t),$$
(A.62)

where we defined $\varphi_t(x) := -x/(x+t)$ for $0 < t < \infty$. The first and the second terms on the right-hand side above are just constants, because $\mathcal{E}$ is TP. The monotonicity of the third term, $\mathrm{tr}[\hat{\rho}^2 \hat{\sigma}^{-1}] b$ with $b \geq 0$, was already proved in Corollary A.1. The first term inside the integral again gives a constant. Finally, $D_{\varphi_t}(\hat{\rho} \| \hat{\sigma})$ satisfies the monotonicity, because φ_t is operator convex and operator decreasing-monotone, and $\varphi_t(0) \leq 0$.   □

We discuss the following two special cases.

Corollary A.4 *Let $\hat{\rho}$, $\mathcal{E}(\hat{\rho})$ be positive and $\hat{\sigma}$, $\mathcal{E}(\hat{\sigma})$ be positive definite. Let $\mathcal{E}$ be 2-positive and TP. The quantum KL divergence $S_1(\hat{\rho} \| \hat{\sigma})$ satisfies the monotonicity*

$$S_1(\hat{\rho} \| \hat{\sigma}) \geq S_1(\mathcal{E}(\hat{\rho}) \| \mathcal{E}(\hat{\sigma})).$$
(A.63)

From this, Theorem 5.1 immediately follows.

Proof Take $f(x) = x \ln x$, which is operator convex.   □

Corollary A.5 ([112, 114]) *Let $\hat{\rho}$, $\mathcal{E}(\hat{\rho})$ be positive and $\hat{\sigma}$, $\mathcal{E}(\hat{\sigma})$ be positive definite. Let $\mathcal{E}$ be 2-positive and TP. For $0 \leq \alpha \leq 2$, the quantum Rényi α-divergence (A.35) satisfies the monotonicity:*

$$\tilde{S}_{\alpha}(\hat{\rho} \| \hat{\sigma}) \geq \tilde{S}_{\alpha}(\mathcal{E}(\hat{\rho}) \| \mathcal{E}(\hat{\sigma})).$$
(A.64)

Proof For $\alpha \neq 0, 1$, take $f(x) = x^{\alpha}$, which is operator concave for $0 < \alpha < 1$ and operator convex for $1 < \alpha \leq 2$. By noting the sign of $\alpha - 1$, we obtain inequality (A.64). For the case of $\alpha = 0, 1$, we can take the limit $\alpha \to +0$ and $\alpha \to 1$, respectively.   □

From the monotonicity, we have the joint convexity of $D_f(\hat{\rho} \| \hat{\sigma})$; we will omit the proof, as it is completely parallel to that of Theorem 5.2 for the quantum KL divergence.

Corollary A.6 (Joint convexity, Corollary 4.7 of [112]) *Let $\hat{\rho}$, $\mathscr{E}(\hat{\rho})$ be positive and $\hat{\sigma}$, $\mathscr{E}(\hat{\sigma})$ be positive definite. Suppose that $f : (0, \infty) \to \mathbb{R}$ is operator convex and $f(0) := \lim_{x \to +0} f(x) \in \mathbb{R}$ exists. Let $\hat{\rho} = \sum_k p_k \hat{\rho}_k$ and $\hat{\sigma} = \sum_k p_k \hat{\sigma}_k$, where $\hat{\rho}_k$ and $\hat{\sigma}_k$ are normalized states and p_k's represent a classical distribution with $p_k > 0$. Then,*

$$D_f(\hat{\rho}\|\hat{\sigma}) \le \sum_k p_k D_f(\hat{\rho}_k\|\hat{\sigma}_k). \tag{A.65}$$

If f is operator concave, the opposite inequality holds. The equality holds if $\mathscr{P}_k$'s are orthogonal with each other, where $\mathscr{P}_k$ is the subspace spanned by the supports of $\hat{\rho}_k$ and $\hat{\sigma}_k$.

Corollary A.7 (Joint convexity of the quantum Rényi divergence) *Let $\hat{\rho}$, $\mathscr{E}(\hat{\rho})$ be positive and $\hat{\sigma}$, $\mathscr{E}(\hat{\sigma})$ be positive definite. For $0 < \alpha < 1$, the quantum Rényi α-divergence (A.35) satisfies the joint convexity (with the same notations as Corollary A.6 and with $p_k > 0$):*

$$\tilde{S}_\alpha(\hat{\rho}\|\hat{\sigma}) \le \sum_k p_k \tilde{S}_\alpha(\hat{\rho}_k\|\hat{\sigma}_k). \tag{A.66}$$

The equality holds if the same equality condition as in Corollary A.6 is satisfied and $\tilde{S}_\alpha(\hat{\rho}_k\|\hat{\sigma}_k)$'s are the same for all k.

Proof From Corollary A.6 with f_α being concave and from the convexity of $\frac{1}{\alpha-1}\ln t$,

$$\frac{1}{\alpha-1}\ln D_{f_\alpha}(\hat{\rho}\|\hat{\sigma}) \le \frac{1}{\alpha-1}\ln \sum_k p_k D_{f_\alpha}(\hat{\rho}_k\|\hat{\sigma}_k) \le \frac{1}{\alpha-1}\sum_k p_k \ln D_{f_\alpha}(\hat{\rho}_k\|\hat{\sigma}_k). \tag{A.67}$$

We note that for $\alpha > 1$, the right inequality fails. The equality in (A.66) holds if the equality conditions for the above two inequalities are satisfied. $\qquad\square$

We can take the limit of inequality (A.66) for $\alpha = 0, 1$, while the joint convexity for these cases is already proved in Theorem 5.5 and Theorem 5.2, respectively.

Quantum f-divergence. We remark on the quantum f-divergence [112], which is a special case of the Petz's quasi-entropies [16, 17]. Let $\hat{\rho}$ and $\hat{\sigma}$ be normalized states. $D_f(\hat{\rho}\|\hat{\sigma})$ is called the quantum f-divergence, if f is operator convex on $(0, \infty)$, strictly convex at $x = 1$, and $f(1) = 0$. For example, $S_1(\hat{\rho}\|\hat{\sigma})$ is the quantum f-divergence in this sense. From Theorem A.2, the quantum f-divergence is non-negative: $D_f(\hat{\rho}\|\hat{\sigma}) \ge 0$, where the equality holds if and only if $\hat{\rho} = \hat{\sigma}$. From Theorem A.3, $D_f(\hat{\rho}\|\hat{\sigma})$ satisfies the monotonicity under CPTP maps.

The trace distance $D(\hat{\rho}, \hat{\sigma})$ is not an f-divergence in the quantum case. We note that $f(x) = |x - 1|$ is not operator convex in any interval that contains $x = 1$ [7]. The quantum fidelity (7.44) is not related to a quantum f-divergence too. In fact, $f(x) = 1 - \sqrt{x}$ is operator convex and gives $D_f(\hat{\rho}\|\hat{\sigma}) = 1 - \text{tr}[\hat{\rho}^{1/2}\hat{\sigma}^{1/2}]$, which is not equivalent to Eq. (7.44).

Sandwiched Rényi divergence. We can also introduce another version of the quantum Rényi α-divergence, called the sandwiched Rényi α-divergence [115–118]. Let $\hat{\rho}$, $\hat{\sigma}$ be normalized states. The sandwiched Rényi α-divergence with $0 < \alpha < 1$ and $1 < \alpha < \infty$ is defined as

$$S_\alpha(\hat{\rho}\|\hat{\sigma}) := \frac{1}{\alpha - 1} \ln \left(\mathrm{tr} \left[(\hat{\sigma}^{\frac{1-\alpha}{2\alpha}} \hat{\rho} \hat{\sigma}^{\frac{1-\alpha}{2\alpha}})^\alpha \right] \right). \tag{A.68}$$

It is known that $S_\alpha(\hat{\rho}\|\hat{\sigma}) \geq 0$ holds, where the equality is achieved if and only if $\hat{\rho} = \hat{\sigma}$ (e.g., Theorem 5 of [118]). The limit $\alpha \to \infty$ of $S_\alpha(\hat{\rho}\|\hat{\sigma})$ equals $S_\infty(\hat{\rho}\|\hat{\sigma})$ defined by Eq. (5.36) (Theorem 5 of [116]). Also, the limit $\alpha \to 1$ gives the quantum KL divergence. Thus, $S_\alpha(\hat{\rho}\|\hat{\sigma})$ is well-defined for $0 < \alpha \leq \infty$. It is known [116, 118] that $S_\alpha(\hat{\rho}\|\hat{\sigma}) \leq S_{\alpha'}(\hat{\rho}\|\hat{\sigma})$ for $\alpha \leq \alpha'$. We note that the fidelity $F(\hat{\rho}, \hat{\sigma})$ corresponds to $S_{1/2}(\hat{\rho}\|\hat{\sigma})$ as shown in Eq. (7.47). It is also known that the sandwiched Rényi divergence satisfies the monotonicity under CPTP map $\mathscr{E}$ (Theorem 1 of [117]): For $1/2 \leq \alpha \leq \infty$,

$$S_\alpha(\hat{\rho}\|\hat{\sigma}) \geq S_\alpha(\mathscr{E}(\hat{\rho})\|\mathscr{E}(\hat{\sigma})). \tag{A.69}$$

We remark that it has been further proved in Ref. [177] that $S_\alpha(\hat{\rho}\|\hat{\sigma})$ satisfies the monotonicity under *positive* and TP maps for $\alpha \geq 1$; in particular, the quantum KL divergence satisfies the monotonicity under positive and TP maps. $S_{1/2}(\hat{\rho}\|\hat{\sigma})$ also satisfies the monotonicity under positive and TP maps, as the fidelity satisfies it as mentioned in Sect. 7.4.

Characterization of the KL divergence. So far, we have shown the monotonicity of various divergence-like quantities. To put it in another way, the monotonicity is not sufficient to specify a single divergence such as the KL divergence. Then, an interesting question is: Under which additional conditions, the KL divergence can be uniquely characterized? This question has been answered in Refs. [83, 138] as follows.

Theorem A.4 (Theorem 1 of [83]) *Let $S(\hat{\rho}\|\hat{\sigma})$ be a real-valued function of two normalized quantum states with $\hat{\sigma}$ being positive definite. Suppose that the following four properties are satisfied:*

Continuity: $S(\hat{\rho}\|\hat{\sigma})$ *is a continuous function of $\hat{\rho}$.*
Monotonicity: *For any CPTP map $\mathscr{E}$, $S(\hat{\rho}\|\hat{\sigma}) \geq S(\mathscr{E}(\hat{\rho})\|\mathscr{E}(\hat{\sigma}))$.*
Additivity: $S(\hat{\rho}_A \otimes \hat{\rho}_B\|\hat{\sigma}_A \otimes \hat{\sigma}_B) = S(\hat{\rho}_A\|\hat{\sigma}_A) + S(\hat{\rho}_B\|\hat{\sigma}_B)$.
Super-additivity: *Let $\hat{\rho}_{AB}$ be a state of a composite system AB with partial states $\hat{\rho}_A$ and $\hat{\rho}_B$. Then, $S(\hat{\rho}_{AB}\|\hat{\sigma}_A \otimes \hat{\sigma}_B) \geq S(\hat{\rho}_A\|\hat{\sigma}_A) + S(\hat{\rho}_B\|\hat{\sigma}_B)$.*

Then, $S(\hat{\rho}\|\hat{\sigma})$ equals the KL divergence up to normalization, i.e., $S(\hat{\rho}\|\hat{\sigma}) = C S_1(\hat{\rho}\|\hat{\sigma})$ for some constant $C > 0$.

A.4 Quantum Fisher Information

We briefly discuss a quantum generalization of the Fisher information in line with
Ref. [19, 20], where again the concepts of operator monotone and operator concave
play crucial roles. The quantum Fisher information is related to quantum estimation
theory, which we will not go into details in this book. In contrast to the classical
case discussed in Sect. 2.5, the quantum Fisher information is not unique, but can be
characterized by operator monotone (and operator concave) functions.

We consider smooth parametrization of positive-definite states, written as $\hat{\rho}(\theta)$
with $\theta := (\theta^1, \theta^2, \ldots, \theta^m) \in \mathbb{R}^m$. We denote $\partial_k := \partial/\partial\theta^k$.

Definition A.3 (*Quantum Fisher information*) Let $f : (0, \infty) \to (0, \infty)$ be oper-
ator monotone and suppose that $f(0) := \lim_{x \to +0} f(x) \in [0, \infty)$ exists. (f is thus
operator concave from Proposition A.4.) Let $\hat{\rho}$ be a positive-definite (normalized)
state. We define

$$\mathcal{K}_{\hat{\rho}} := f(\mathcal{D}_{\hat{\rho},\hat{\rho}})\mathcal{R}_{\hat{\rho}}. \tag{A.70}$$

Then, the quantum Fisher information matrix $K_{\hat{\rho}(\theta)}$, whose matrix component is
written as $K_{\hat{\rho}(\theta),kl}$, is defined as

$$K_{\hat{\rho}(\theta),kl} := \langle \partial_k \hat{\rho}(\theta), \mathcal{K}_{\hat{\rho}(\theta)}^{-1}(\partial_l \hat{\rho}(\theta)) \rangle_{\mathrm{HS}} = \mathrm{tr}[\partial_k \hat{\rho}(\theta) \mathcal{K}_{\hat{\rho}(\theta)}^{-1}(\partial_l \hat{\rho}(\theta))]. \tag{A.71}$$

We discuss two important examples. A simplest case is $f(x) = 1$, for which the
quantum Fisher information is called the RLD (right logarithmic derivative) Fisher
information. In this case,

$$K_{\hat{\rho}(\theta),kl} = \mathrm{tr}[\partial_k \hat{\rho}(\theta) \partial_l \hat{\rho}(\theta) \hat{\rho}(\theta)^{-1}]. \tag{A.72}$$

We note that this can be rewritten as

$$K_{\hat{\rho}(\theta),kl} = \mathrm{tr}[\hat{\rho}(\theta) \hat{L}_l(\theta) \hat{L}_k(\theta)], \tag{A.73}$$

where $\hat{L}_k(\theta)$ is defined as the solution of $\partial_k \hat{\rho}(\theta) = \hat{\rho}(\theta) \hat{L}_k(\theta)$ and is given by
$\hat{L}_k(\theta) = \hat{\rho}(\theta)^{-1} \partial_k \hat{\rho}(\theta)$.

Another example is $f(x) = (1 + x)/2$, for which the quantum Fisher information
is called the SLD (symmetric logarithmic derivative) Fisher information. In this case,
we have $\mathcal{K}_{\hat{\rho}(\theta)} = (\mathcal{L}_{\hat{\rho}(\theta)} + \mathcal{R}_{\hat{\rho}(\theta)})/2$, and $\hat{L}_l(\theta) := \mathcal{K}_{\hat{\rho}(\theta)}^{-1}(\partial_l \hat{\rho}(\theta))$ is given by the
solution of $\partial_l \hat{\rho}(\theta) = (\hat{\rho}(\theta) \hat{L}_l(\theta) + \hat{L}_l(\theta) \hat{\rho}(\theta))/2$. Thus,

$$K_{\hat{\rho}(\theta),kl} = \mathrm{tr}[\partial_k \hat{\rho}(\theta) \hat{L}_l(\theta)] = \frac{1}{2} \mathrm{tr}[\hat{\rho}(\theta)(\hat{L}_k(\theta) \hat{L}_l(\theta) + \hat{L}_l(\theta) \hat{L}_k(\theta))]. \tag{A.74}$$

These two versions of the quantum Fisher information are more or less straightforward generalizations of the classical Fisher information, but here we need to take into account the fact that the logarithmic derivative $\partial_k \ln p(\theta)$ is not uniquely extended to the quantum case because of the non-commutability of operators. We note that there is also another useful quantity called the ALD (anti-symmetric logarithmic derivative) Fisher information [178], while it is not one of the quantum Fisher information in the sense of of Definition A.3.

The classical Fisher information is a special case of the quantum Fisher information:

Lemma A.4 *Suppose that $\hat{\rho}(\theta)$'s are diagonalizable in the same basis for all θ. Let $p(\theta)$ be the diagonal distribution of $\hat{\rho}(\theta)$. Then, the quantum Fisher information matrix of $\hat{\rho}(\theta)$ for any f reduces to the classical Fisher information matrix of $p(\theta)$ up to normalization.*

Proof Let $\hat{\rho}(\theta) := \sum_i p_i(\theta) \hat{P}_i$ be the spectral decomposition. From the expression like Eq. (A.28), if $\hat{X} = \sum_i x_i \hat{P}_i$ is also diagonalizable in the same basis, we have $\mathscr{K}_{\hat{\rho}(\theta)}(\hat{X}) = f(1) \sum_i p_i(\theta) x_i \hat{P}_i$, and thus $\mathscr{K}_{\hat{\rho}(\theta)}^{-1}(\hat{X}) = \frac{1}{f(1)} \sum_i \frac{x_i}{p_i(\theta)} \hat{P}_i$. By substituting $\hat{X} = \partial_l \hat{\rho}(\theta) = \sum_i \partial_l p_i(\theta) \hat{P}_i$, we obtain the claim of the lemma. $\qquad\square$

As in the classical case, we can adopt the information geometry perspective. For $\hat{\rho} > 0$, the quantum Fisher information metric on the operator space is given by a map $G_{\hat{\rho}} : \mathscr{L}(\mathscr{H}) \times \mathscr{L}(\mathscr{H}) \to \mathbb{C}$ defined as

$$G_{\hat{\rho}}(\hat{X}, \hat{Y}) := \langle \hat{X}, \mathscr{K}_{\hat{\rho}}^{-1}(\hat{Y}) \rangle_{\text{HS}} = \text{tr}[\hat{X}^\dagger \mathscr{K}_{\hat{\rho}}^{-1}(\hat{Y})]. \tag{A.75}$$

The quantum Fisher information matrix is then represented as

$$K_{\hat{\rho}(\theta),kl} = G_{\hat{\rho}(\theta)}(\partial_k \hat{\rho}(\theta), \partial_l \hat{\rho}(\theta)). \tag{A.76}$$

As is the classical case, the quantum Fisher information also satisfies the monotonicity, whose proof invokes a similar technique to the proof of the monotonicity of divergences (Theorem A.3).

Theorem A.5 (Monotonicity of the quantum Fisher information, Theorem 3 of [19]) *Let $\mathscr{E}$ be 2-positive and TP. Suppose that $\hat{\rho}$, $\mathscr{E}(\hat{\rho})$ are (normalized) positive-definite states. For any $\hat{X} \in \mathscr{L}(\mathscr{H})$,*

$$G_{\hat{\rho}}(\hat{X}, \hat{X}) \geq G_{\mathscr{E}(\hat{\rho})}(\mathscr{E}(\hat{X}), \mathscr{E}(\hat{X})). \tag{A.77}$$

In particular, the quantum Fisher information matrix satisfies

$$K_{\hat{\rho}(\theta)} \geq K_{\mathscr{E}(\hat{\rho}(\theta))}. \tag{A.78}$$

Proof It is sufficient to prove that

$$\mathscr{K}_{\hat{\rho}}^{-1} \geq \mathscr{E}^{\dagger} \mathscr{K}_{\mathscr{E}(\hat{\rho})}^{-1} \mathscr{E}. \tag{A.79}$$

This is equivalent to

$$\mathscr{I} \geq \mathscr{K}_{\hat{\rho}}^{1/2} \mathscr{E}^{\dagger} \mathscr{K}_{\mathscr{E}(\hat{\rho})}^{-1} \mathscr{E} \mathscr{K}_{\hat{\rho}}^{1/2} \quad \Leftrightarrow \quad \mathscr{I} \geq \mathscr{K}_{\mathscr{E}(\hat{\rho})}^{-1/2} \mathscr{E} \mathscr{K}_{\hat{\rho}} \mathscr{E}^{\dagger} \mathscr{K}_{\mathscr{E}(\hat{\rho})}^{-1/2}, \tag{A.80}$$

where $\mathscr{I}$ is the identity. Thus, what we will prove is that

$$\mathscr{K}_{\mathscr{E}(\hat{\rho})} \geq \mathscr{E} \mathscr{K}_{\hat{\rho}} \mathscr{E}^{\dagger}. \tag{A.81}$$

By using a contraction $\mathscr{V}$ in Lemma A.3 by replacing $\hat{\sigma}$ with $\hat{\rho}$, we have

$$\mathscr{E} \mathscr{K}_{\hat{\rho}} \mathscr{E}^{\dagger} = \mathscr{R}_{\mathscr{E}(\hat{\rho})}^{1/2} \mathscr{V}^{\dagger} f(\mathscr{D}_{\hat{\rho},\hat{\rho}}) \mathscr{V} \mathscr{R}_{\mathscr{E}(\hat{\rho})}^{1/2}. \tag{A.82}$$

We apply inequality (A.59), but now f is operator concave and operator monotone with $f(0) \geq 0$:

$$\mathscr{V}^{\dagger} f(\mathscr{D}_{\hat{\rho},\hat{\rho}}) \mathscr{V} \leq f(\mathscr{D}_{\mathscr{E}(\hat{\rho}),\mathscr{E}(\hat{\rho})}). \tag{A.83}$$

From inequalities (A.82) and (A.83), we finally obtain inequality (A.81). $\qquad\square$

To relate the quantum Fisher information to parameter estimation, we consider quantum measurement on $\hat{\rho}(\theta)$, where the measurement itself is assumed to be independent of θ. Let $p(\theta)$ be the classical probability distribution obtained by the measurement, and let $\hat{p}(\theta) := \sum_i p(\theta) \hat{P}_i$ be the corresponding diagonal density operator with a fixed basis represented by projectors $\{\hat{P}_i\}$. Let $J_{p(\theta)}$ be the classical Fisher information matrix of $p(\theta)$, which is equivalent to the quantum Fisher information matrix of $\hat{p}(\theta)$ from Lemma A.4 (here we take normalization $f(1) = 1$). Since $\hat{\rho}(\theta) \mapsto \hat{p}(\theta)$ is a CPTP map, the monotonicity (A.78) implies

$$K_{\hat{\rho}(\theta)} \geq J_{p(\theta)}, \tag{A.84}$$

which is sometimes called the quantum Cramer-Rao inequality. By combining this with the classical Cramer-Rao bound (2.48), we obtain a bound of the accuracy of unbiased estimation of θ from quantum measurement:

$$\mathrm{Cov}_{\theta}(\theta_{\mathrm{est}}) \geq K_{\hat{\rho}(\theta)}^{-1}, \tag{A.85}$$

where the left-hand side depends on the choice of quantum measurement, but the right-hand side does not.

We finally note the quantum analogue of the Chentsov's theorem (Theorem 2.1), which is referred to as the Petz's theorem. The quantum monotone metric is defined as follows.

Definition A.4 (*Quantum monotone metric*) Suppose that $G_{\hat{\rho}} : \mathscr{L}(\mathscr{H}) \times \mathscr{L}(\mathscr{H}) \to \mathbb{C}$ is defined for positive-definite states $\hat{\rho}$. We call $G_{\hat{\rho}}$ a quantum monotone metric on the operator space, if it satisfies the following.

- $G_{\hat{\rho}}$ is sesquilinear (as is the case for the ordinary complex inner product).
- $G_{\hat{\rho}}(\hat{X}, \hat{X}) \geq 0$ holds for any $\hat{\rho}$, where the equality is achieved if and only if $\hat{X} = 0$.
- $\hat{\rho} \mapsto G_{\hat{\rho}}(\hat{X}, \hat{X})$ is continuous for any $\hat{X}$.
- The monotonicity (A.77) holds for any CPTP $\mathscr{E}$ and for any $\hat{\rho}$, $\hat{X}$.

The quantum Fisher information metric (A.75) is a quantum monotone metric. The Petz's theorem states the converse:

Theorem A.6 (Petz's theorem, Theorem 5 of [19]) *Any quantum monotone metric is a quantum Fisher information metric with some f.*

We finally note that a metric is called symmetric, if $G_{\hat{\rho}}(\hat{X}, \hat{Y}) = G_{\hat{\rho}}(\hat{Y}^{\dagger}, \hat{X}^{\dagger})$. It is known that the quantum Fisher information metric is symmetric if and only if f satisfies $f(x) = xf(x^{-1})$ (Theorem 7 of Ref. [19]). In this case, the corresponding quantum Fisher information matrix is a symmetric matrix. For example, the SLD Fisher information is symmetric. Moreover, it is known that there are the smallest and the largest symmetric Fisher information metrics (Corollary 9 of Ref. [19]); In particular, the SLD metric is the smallest one.

Appendix B
Hypothesis Testing

As discussed in Chap. 7, hypothesis testing is an important tool to analyze the smooth divergences and their asymptotic limit. In this Appendix, we briefly overview the basic concepts of quantum hypothesis testing. In Sect. B.1, we introduce the hypothesis testing divergence and discuss its relation to semidefinite programming. In Sect. B.2, we discuss a most fundamental theorem regarding quantum hypothesis testing: the quantum Stein's lemma, which is another representation of the quantum relative AEP discussed in Sect. 7.3.

The basic task of quantum hypothesis testing can be stated as follows: we want to distinguish two quantum states $\hat{\rho}$ and $\hat{\sigma}$ with $\hat{\sigma}$ being the false null hypothesis, and minimize the error probability of the second kind given by $\text{tr}[\hat{\sigma}\hat{Q}]$ with $0 \leq \hat{Q} \leq \hat{I}$, while keeping $\text{tr}[\hat{\rho}\hat{Q}] \geq \eta$ for a given $0 < \eta < 1$.

B.1 Hypothesis Testing Divergence

The divergence-like quantity related to the above-mentioned error probability is the hypothesis testing divergence [45, 153, 179]. Let $\hat{\rho}$ be a (normalized) state and $\hat{\sigma} \geq 0$ be an unnormalized state. For $0 < \eta < 1$, the hypothesis testing divergence is defined as

$$S_{\text{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}) := -\ln\left(\frac{1}{\eta}\min_{0 \leq \hat{Q} \leq \hat{I}, \text{tr}[\hat{\rho}\hat{Q}] \geq \eta} \text{tr}[\hat{\sigma}\hat{Q}]\right). \tag{B.1}$$

We note that $S_{\text{H}}^{\eta=1}(\hat{\rho}\|\hat{\sigma}) = S_0(\hat{\rho}\|\hat{\sigma})$ as shown in Eq. (5.41), while we only focus on $0 < \eta < 1$ in this Appendix.

Because the argument of the logarithm in Eq. (B.1) can be rewritten as $\min_{0 \leq \hat{Q}' \leq \hat{I}/\eta, \text{tr}[\hat{\rho}\hat{Q}'] \geq 1} \text{tr}[\hat{\sigma}\hat{Q}']$ with $\hat{Q}' := \hat{Q}/\eta$, we have

$$S_{\text{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}) \geq S_{\text{H}}^{\eta'}(\hat{\rho}\|\hat{\sigma}) \quad \text{for} \quad \eta \leq \eta'. \tag{B.2}$$

We note the scaling property:

$$S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}/Z) = S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}) + \ln Z. \tag{B.3}$$

Moreover, if $\hat{\sigma} \leq \hat{\sigma}'$,

$$S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}') \leq S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}). \tag{B.4}$$

A particularly important property of the hypothesis testing divergences in our context is that $S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma})$ with $\eta \simeq 0$ and $\eta \simeq 1$ are essentially equivalent to the smooth ∞- and 0-divergences, respectively:

Proposition B.1 (Lemma 40 of Ref. [45]; Proposition 1 of Ref. [129]) *For* $0 < \varepsilon < 1/2$,

$$S_{\mathrm{H}}^{1-\varepsilon^2/6}(\hat{\rho}\|\hat{\sigma}) - \ln \frac{1-\varepsilon^2/6}{\varepsilon^2/6} \leq S_0^{\varepsilon}(\hat{\rho}\|\hat{\sigma}) \leq S_{\mathrm{H}}^{1-\varepsilon}(\hat{\rho}\|\hat{\sigma}) - \ln(1-\varepsilon), \tag{B.5}$$

$$S_{\mathrm{H}}^{2\varepsilon}(\hat{\rho}\|\hat{\sigma}) - \ln 2 \leq S_{\infty}^{\varepsilon}(\hat{\rho}\|\hat{\sigma}) \leq S_{\mathrm{H}}^{\varepsilon^2/2}(\hat{\rho}\|\hat{\sigma}) - \ln(1-\varepsilon). \tag{B.6}$$

See Ref. [45] for the proof. We note that Ref. [45, 129] adopted different ways of smoothing, and thus inequality (B.6) is slightly different (see also Sect. 7.4).

We next discuss the intuitive meaning of Proposition B.1 in a non-rigorous manner. For the case of $S_0^{\varepsilon}(\hat{\rho}\|\hat{\sigma})$, by noting Eq. (5.41), we have

$$S_{\mathrm{H}}^{\eta\simeq1}(\hat{\rho}\|\hat{\sigma}) \simeq -\ln\left(\min_{0\leq\hat{Q}\leq\hat{I},\ \mathrm{tr}[\hat{\rho}\hat{Q}]\simeq1}\mathrm{tr}[\hat{\sigma}\hat{Q}]\right) \simeq S_0^{\varepsilon\simeq0}(\hat{\rho}\|\hat{\sigma}). \tag{B.7}$$

For the case of $S_{\infty}^{\varepsilon}(\hat{\rho}\|\hat{\sigma})$, we invoke the dual expression of Eq. (B.1), which is given by Eq. (B.21) at the end of this section. By letting $\eta \simeq 0$, $-\mathrm{tr}[\hat{X}]/\eta$ in Eq. (B.21) would take a very large negative value, and thus $\mathrm{tr}[\hat{X}] \simeq 0$ should hold for the maximum. Then,

$$S_{\mathrm{H}}^{\eta\simeq0}(\hat{\rho}\|\hat{\sigma}) \simeq -\ln\left(\max_{\mu\geq0,\mu\hat{\rho}\lesssim\hat{\sigma}}\mu\right) \simeq \ln\|\hat{\sigma}^{-1/2}\hat{\rho}\hat{\sigma}^{-1/2}\|_{\infty} \simeq S_{\infty}^{\varepsilon\simeq0}(\hat{\rho}\|\hat{\sigma}). \tag{B.8}$$

We note some general properties of the hypothesis testing divergences. First, $S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma})$ satisfies the monotonicity in the following form:

Proposition B.2 *For* $0 < \eta < 1$ *and any positive and trace-nonincreasing map* $\mathscr{E}$,

$$S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}) \geq S_{\mathrm{H}}^{\eta}(\mathscr{E}(\hat{\rho})\|\mathscr{E}(\hat{\sigma})). \tag{B.9}$$

Proof The proof goes completely parallel to that of Theorem 5.4 (b):

$$\min_{0\leq\hat{Q}\leq\hat{I},\mathrm{tr}[\mathscr{E}(\hat{\rho})\hat{Q}]\geq\eta} \mathrm{tr}[\mathscr{E}(\hat{\sigma})\hat{Q}] = \min_{0\leq\hat{Q}\leq\hat{I},\mathrm{tr}[\hat{\rho}\mathscr{E}^{\dagger}(\hat{Q})]\geq\eta} \mathrm{tr}[\hat{\sigma}\mathscr{E}^{\dagger}(\hat{Q})] \quad \text{(B.10)}$$

$$\geq \min_{0\leq\hat{Q}'\leq\hat{I},\mathrm{tr}[\hat{\rho}\hat{Q}']\geq\eta} \mathrm{tr}[\hat{\sigma}\hat{Q}'], \quad \text{(B.11)}$$

which implies inequality (B.9). $\qquad\square$

Lemma B.1 (Inequality (22) of [129]) *Let $\varepsilon \geq 0$ and $D(\hat{\rho}, \hat{\tau}) \leq \varepsilon$. For any $0 < \eta < 1 - \varepsilon$,*

$$S_{\mathrm{H}}^{\eta+\varepsilon}(\hat{\tau}\|\hat{\sigma}) \leq S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}) + \ln\left(\frac{\eta+\varepsilon}{\eta}\right). \quad \text{(B.12)}$$

Proof There exists $\hat{\Delta}$ such that $\hat{\rho} + \hat{\Delta} \geq \hat{\tau}$, $\hat{\Delta} \geq 0$, and $\mathrm{tr}[\hat{\Delta}] \leq \varepsilon$. If $0 \leq \hat{Q} \leq \hat{I}$, we have $\mathrm{tr}[\hat{\tau}\hat{Q}] \leq \mathrm{tr}[(\hat{\rho} + \hat{\Delta})\hat{Q}] \leq \mathrm{tr}[\hat{\rho}\hat{Q}] + \varepsilon$. Now let $\hat{Q}$ be the optimal choice for the left-hand side of (B.12). Then, $\mathrm{tr}[\hat{\rho}\hat{Q}] \geq \eta$ holds and thus

$$\frac{1}{\eta+\varepsilon}\mathrm{tr}[\hat{\sigma}\hat{Q}] \geq \frac{\eta}{\eta+\varepsilon}\frac{1}{\eta}\min_{0\leq\hat{Q}'\leq\hat{I},\mathrm{tr}[\hat{\rho}\hat{Q}']\geq\eta}\mathrm{tr}[\hat{\sigma}\hat{Q}']. \quad \text{(B.13)}$$

$\qquad\square$

Lemma B.2 *Let $\hat{\sigma}_k \geq 0$ and define $\hat{\sigma} := \sum_k r_k\hat{\sigma}_k \otimes |k\rangle\langle k|$ with $\{|k\rangle\}$ being orthonormal and $r_k > 0$. Then,*

$$S_{\mathrm{H}}^{\eta}(\hat{\rho}\otimes|k\rangle\langle k|\|\hat{\sigma}) = S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}_k) - \ln r_k. \quad \text{(B.14)}$$

Proof Let $\hat{Q}$ be the optimal choice for $S_{\mathrm{H}}^{\eta}(\hat{\rho}\otimes|k\rangle\langle k|\|\hat{\sigma})$. Then $\langle k|\hat{Q}|k\rangle$ is a candidate for $S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}_k)$ because $\mathrm{tr}[\hat{\rho}\otimes|k\rangle\langle k|\hat{Q}] = \mathrm{tr}[\hat{\rho}\langle k|\hat{Q}|k\rangle]$. From this and $\mathrm{tr}[\hat{\sigma}\hat{Q}] \geq \mathrm{tr}[\hat{\sigma}_k\langle k|\hat{Q}|k\rangle]r_k$, we have $S_{\mathrm{H}}^{\eta}(\hat{\rho}\otimes|k\rangle\langle k|\|\hat{\sigma}) \leq S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}_k) - \ln r_k$.

Conversely, let $\hat{Q}'$ be the optimal choice for $S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}_k)$. Then $\hat{Q}' \otimes |k\rangle\langle k|$ is a candidate for $S_{\mathrm{H}}^{\eta}(\hat{\rho}\otimes|k\rangle\langle k|\|\hat{\sigma})$. From this and $\mathrm{tr}[\hat{\sigma}_k\hat{Q}'] = \mathrm{tr}[\hat{\sigma}\hat{Q}'\otimes|k\rangle\langle k|]r_k^{-1}$, we obtain $S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}_k) \leq S_{\mathrm{H}}^{\eta}(\hat{\rho}\otimes|k\rangle\langle k|\|\hat{\sigma}) + \ln r_k$. $\qquad\square$

We note that the above lemma also holds for S_0 and S_∞. However, it does not hold for S_0^ε and S_∞^ε with $\varepsilon > 0$, which is the reason why we used the hypothesis testing divergence in Theorem 8.2 instead of directly addressing S_0^ε and S_∞^ε.

We next briefly overview semidefinite programing in line with Refs. [45, 153, 180], from which we obtain the dual expression of the hypothesis testing divergence. (See Ref. [181] for a comprehensive textbook of convex optimization.) We note that our terminologies "primal" and "dual" are the same as in Ref. [180]. On the other

hand, "dual" and "primal" are exchanged in some papers (e.g., Ref. [153]), where the standard hypothesis testing becomes the "primal" program.

Let $\hat{C}$ and $\hat{D}$ be Hermitian matrices and $\mathscr{E}$ be a Hermitian-preserving map. The primal program is the optimization problem given by

$$\sup_{\hat{X} \geq 0, \mathscr{E}(\hat{X}) \leq \hat{D}} \mathrm{tr}[\hat{C}\hat{X}] =: \alpha, \tag{B.15}$$

while the dual program is given by

$$\inf_{\hat{Y} \geq 0, \mathscr{E}^\dagger(\hat{Y}) \geq \hat{C}} \mathrm{tr}[\hat{D}\hat{Y}] =: \beta. \tag{B.16}$$

Without any further assumption, we have

$$\alpha \leq \beta, \tag{B.17}$$

which is called the weak duality. The proof is straightforward: For $\hat{X} \geq 0$ with $\mathscr{E}(\hat{X}) \leq \hat{D}$ and $\hat{Y} \geq 0$ with $\mathscr{E}^\dagger(\hat{Y}) \geq \hat{C}$, we have

$$\mathrm{tr}[\hat{C}\hat{X}] \leq \mathrm{tr}[\mathscr{E}^\dagger(\hat{Y})\hat{X}] = \mathrm{tr}[\hat{Y}\mathscr{E}(\hat{X})] \leq \mathrm{tr}[\hat{Y}\hat{D}], \tag{B.18}$$

and then take the supremum over $\hat{X}$ and the infimum over $\hat{Y}$.

Moreover, it is known that the *strong duality* holds under certain conditions, such as the Slater's conditions (e.g., Theorem 2.2 of [180]):

- If β is finite and there exists $\hat{X} > 0$ such that $\mathscr{E}(\hat{X}) < \hat{D}$, then $\alpha = \beta$ holds and there exists $\hat{Y}$ such that $\hat{Y} \geq 0$, $\mathscr{E}^\dagger(\hat{Y}) \geq \hat{C}$ and $\mathrm{tr}[\hat{D}\hat{Y}] = \beta$.
- If α is finite and there exists $\hat{Y} > 0$ such that $\mathscr{E}^\dagger(\hat{Y}) > \hat{C}$, then $\alpha = \beta$ holds and there exists $\hat{X}$ such that $\hat{X} \geq 0$, $\mathscr{E}(\hat{X}) \leq \hat{D}$ and $\mathrm{tr}[\hat{C}\hat{X}] = \alpha$.

Our hypothesis testing is a dual program in the above definition, where $\hat{D} := \hat{\sigma}$, $\hat{Y} := \hat{Q}$, and

$$\mathscr{E}^\dagger(\hat{Y}) := \begin{bmatrix} -\hat{Y} & 0 \\ 0 & \mathrm{tr}[\hat{\rho}\hat{Y}] \end{bmatrix}, \quad \hat{C} := \begin{bmatrix} -\hat{I} & 0 \\ 0 & \eta \end{bmatrix}. \tag{B.19}$$

We note that $e^{-S_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma})} = \eta^{-1}\beta$. The corresponding primal program is given by

$$\mathscr{E}(\hat{X}) = \mu\hat{\rho} - \hat{X}_{11}, \quad \text{where} \quad \hat{X} = \begin{bmatrix} \hat{X}_{11} & \hat{X}_{12} \\ \hat{X}_{21} & \mu \end{bmatrix} \quad \text{with } \mu \in \mathbb{R}. \tag{B.20}$$

The condition $\mathscr{E}(\hat{X}) \leq \hat{D}$ then gives $\mu\hat{\rho} - \hat{X}_{11} \leq \hat{\sigma}$, and $\mathrm{tr}[\hat{C}\hat{X}] = \eta\mu - \mathrm{tr}[X_{11}]$. From the weak duality (B.17), we have $e^{-S_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma})} \geq \eta^{-1}\alpha$. Moreover, it is easy to see that there exists $\hat{Y} > 0$ such that $\hat{Y} < \hat{I}$ and $\mathrm{tr}[\hat{\rho}\hat{Y}] > \eta$. Thus, the condition of the strong duality is satisfied, and we have $e^{-S_{\mathrm{H}}^\eta(\hat{\rho}\|\hat{\sigma})} = \eta^{-1}\alpha$ and we can replace sup

by max for α. Then, by rewriting $\hat{X}_{11}$ as just $\hat{X}$ and noting that $\hat{X}_{12}$ and $\hat{X}_{21}$ are irrelevant to the optimization, we obtain [45, 153]

$$S_{\mathrm{H}}^{\eta}(\hat{\rho}\|\hat{\sigma}) = -\ln\left(\max_{\mu \geq 0, \hat{X} \geq 0, \mu\hat{\rho} \leq \hat{\sigma} + \hat{X}} \left\{\mu - \frac{\mathrm{tr}[\hat{X}]}{\eta}\right\}\right). \tag{B.21}$$

B.2 Quantum Stein's Lemma

We consider the asymptotic behavior of the hypothesis testing divergence, which is characterized by the quantum Stein's lemma. See also Appendix C for the classical Stein's lemma.

First, we define the hypothesis testing divergence rate for sequences $\widehat{P} = \{\hat{\rho}_n\}_{n=1}^{\infty}$ and $\widehat{\Sigma} = \{\hat{\sigma}_n\}_{n=1}^{\infty}$ of (normalized) quantum states:

$$S_{\mathrm{H}}^{\eta}(\widehat{P}\|\widehat{\Sigma}) := \lim_{n \to \infty} \frac{1}{n} S_{\mathrm{H}}^{\eta}(\hat{\rho}_n\|\hat{\sigma}_n). \tag{B.22}$$

We note that the limit does not necessarily exist, while we can always define the upper and lower limits, which are equivalent to the lower and upper spectral divergence rates introduced in Sect. 7.2. In fact, from Proposition B.1:

$$\lim_{\varepsilon \to +0} \limsup_{n \to \infty} \frac{1}{n} S_{\mathrm{H}}^{\varepsilon}(\hat{\rho}_n\|\hat{\sigma}_n) = \overline{S}(\widehat{P}\|\widehat{\Sigma}), \tag{B.23}$$

$$\lim_{\varepsilon \to +0} \liminf_{n \to \infty} \frac{1}{n} S_{\mathrm{H}}^{1-\varepsilon}(\hat{\rho}_n\|\hat{\sigma}_n) = \underline{S}(\widehat{P}\|\widehat{\Sigma}). \tag{B.24}$$

The quantum Stein's lemma states that the minimized error probability of the second kind is asymptotically characterized by the KL divergence rate. More precisely: *For any $0 < \eta < 1$,*

$$S_{\mathrm{H}}^{\eta}(\widehat{P}\|\widehat{\Sigma}) = S_1(\widehat{P}\|\widehat{\Sigma}) \tag{B.25}$$

holds.

From Eq. (B.1), we see that Eq. (B.25) implies that $\mathrm{tr}[\hat{\sigma}_n \hat{Q}_n] \sim e^{-nS_1(\widehat{P}\|\widehat{\Sigma})}$. That is, the quantum Stein's lemma states that the KL divergence rate characterizes the large deviation property of the error probability in hypothesis testing.

The quantum Stein's lemma can be equivalently rephrased in terms of the spectral divergence rates from Eqs. (B.23), (B.24), and inequality (B.2):

Proposition B.3 *The quantum Stein's lemma (B.25) holds, if and only if*

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) = \overline{S}(\widehat{P}\|\widehat{\Sigma}) = S_1(\widehat{P}\|\widehat{\Sigma}). \tag{B.26}$$

Proof This is obvious, but we here prove the "if" part. For $0 < \eta \leq 1/2$, we have
$S_1(\widehat{P}\|\widehat{\Sigma}) = \lim_{\varepsilon \to +0} \lim\sup_{n \to \infty} S_H^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n)/n \geq \lim\sup_{n \to \infty} S_H^\eta(\hat{\rho}_n\|\hat{\sigma}_n)/n \geq$
$\lim\inf_{n \to \infty} S_H^\eta(\hat{\rho}_n\|\hat{\sigma}_n)/n \geq \lim\inf_{n \to \infty} S_H^{1-\eta}(\hat{\rho}_n\|\hat{\sigma}_n)/n \geq \lim_{\varepsilon \to +0} \lim\inf_{n \to \infty}$
$S_H^{1-\varepsilon}(\hat{\rho}_n\|\hat{\sigma}_n)/n = S_1(\widehat{P}\|\widehat{\Sigma})$. The proof is almost the same for $1/2 \leq \eta < 1$. $\square$

Equation (B.26) is exactly the same as the relative quantum AEP (7.39). Moreover, from Proposition B.1, we have the following [45].

Proposition B.4 *The quantum Stein's lemma (B.25) holds, if and only if for any* $0 < \varepsilon < 1/2$,

$$\lim_{n \to \infty} \frac{1}{n} S_0^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n) = \lim_{n \to \infty} \frac{1}{n} S_\infty^\varepsilon(\hat{\rho}_n\|\hat{\sigma}_n) = S_1(\widehat{P}\|\widehat{\Sigma}). \tag{B.27}$$

To summarize, in order to prove the collapse of $\overline{S}(\widehat{P}\|\widehat{\Sigma})$ and $\underline{S}(\widehat{P}\|\widehat{\Sigma})$ to $S_1(\widehat{P}\|\widehat{\Sigma})$, it is necessary and sufficient to prove the quantum Stein's lemma. In other words, the quantum Stein's lemma is regarded as a representation of the quantum relative AEP (see also Appendix C for the classical case). In fact, if $\widehat{P}$ is translation invariant and ergodic and $\widehat{\Sigma}$ is the Gibbs state of a local and translation invariant Hamiltonian, then the quantum Stein's lemma (B.25) holds, which is equivalent to Theorem 7.10 of Sect. 7.3 (Theorem 3 of Ref. [129]).

The simplest case is i.i.d. with $\widehat{P} := \{\hat{\rho}^{\otimes n}\}$ and $\widehat{\Sigma} := \{\hat{\sigma}^{\otimes n}\}$. In this case, the quantum Stein's lemma states that for any $0 < \eta < 1$,

$$S_H^\eta(\widehat{P}\|\widehat{\Sigma}) = S_1(\hat{\rho}\|\hat{\sigma}), \tag{B.28}$$

which has been proved in Refs. [149, 150]. Given Proposition B.3, this is equivalent to Theorem 7.7 (Theorem 2 of Ref. [61]).

Appendix C
Classical Asymptotic Equipartition Properties

In this Appendix, we consider the classical asymptotic equipartition properties (AEP) of stochastic processes, which is the classical counterpart of Sect. 7.3. Specifically, we discuss the Shannon-McMillan theorem and the classical Stein's lemma.

Let $\{x_l\}_{l \in \mathbb{Z}}$ be a two-sided stochastic process, where $x_l \in B$ with B being a finite set. This is equivalent to a one-dimensional classical spin system. Formally, the stochastic process is given by a measure μ over $K := B^{\mathbb{Z}}$. We introduce a shift operator $\mathscr{T} : K \to K$ such that $(\mathscr{T}x)_l := x_{l+1}$.

Let $X_n := (x_{-l}, x_{-l+1}, \ldots, x_l)$ with $n := 2l + 1$. We consider sequences of probability distributions $\widehat{P} := \{\rho_n(X_n)\}$ and $\widehat{\Sigma} := \{\sigma_n(X_n)\}$. The spectral divergence rates of these sequences can be defined in a completely parallel manner to the quantum case discussed in Chap. 7. We sometimes loosely identify $\{x_l\}_{l \in \mathbb{Z}}$ and $\widehat{P}$.

In the classical case, we can directly show that if $\widehat{P}$ and $\widehat{\Sigma}$ satisfy a relative version of the AEP, the lower and the upper divergence rates coincide and further equal the KL divergence rate. We start with the definition of the AEP in the form of convergence in probability.

Definition C.1 (*Classical AEP*) $\widehat{P}$ satisfies the AEP, if the Shannon entropy rate $S_1(\widehat{P})$ exists, and $-\frac{1}{n} \ln \rho_n(X_n)$ converges to $S_1(\widehat{P})$ in probability by sampling X_n according to ρ_n.

This definition is equivalent to a statement with *typical set* as described in the following proposition, which can be shown in the same manner as in Theorem 3.1.2 of Ref. [3]. Let $\mathscr{X}_n := B^n$ be the totality of all possible X_n's.

Proposition C.1 $\widehat{P}$ *satisfies the AEP, if and only if for any $\varepsilon > 0$ and $\delta > 0$, there exists a typical set $Q_n^{\varepsilon,\delta} \subset \mathscr{X}_n$, satisfying the following properties for sufficiently large n:*

(a) *For any $X_n \in Q_n^{\varepsilon,\delta}$, $\exp(-n(S_1(\widehat{P}) + \varepsilon)) \leq \rho_n(X_n) \leq \exp(-n(S_1(\widehat{P}) - \varepsilon))$.*
(b) *$\rho_n[Q_n^{\varepsilon,\delta}] > 1 - \delta$, where $\rho_n[Q_n^{\varepsilon,\delta}]$ is the probability of $Q_n^{\varepsilon,\delta}$ according to the distribution ρ_n.*

(c) $(1 - \delta) \exp(n(S_1(\widehat{P}) - \varepsilon)) \leq |Q_n^{\varepsilon,\delta}| \leq \exp(n(S_1(\widehat{P}) + \varepsilon))$*, where* $|Q_n^{\varepsilon,\delta}|$ *describes the number of elements of* $Q_n^{\varepsilon,\delta}$*.*

This formulation of the AEP clearly represents the meaning of "equipartition"; Almost all events (i.e., events in the typical set) have almost the same probabilities.

We next define the relative AEP, which is the classical counterpart of our quantum formulation in Sect. 7.3.

Definition C.2 (*Relative AEP*) $\widehat{P}$ and $\widehat{\Sigma}$ satisfy the relative AEP, if the KL divergence rate $S_1(\widehat{P}\|\widehat{\Sigma})$ exists, and $\frac{1}{n} \ln \frac{\rho_n(X_n)}{\sigma_n(X_n)}$ converges to $S_1(\widehat{P}\|\widehat{\Sigma})$ in probability by sampling X_n according to ρ_n.

The following proposition can be shown in the same manner as Theorem 11.8.2 of Ref. [3].

Proposition C.2 $\widehat{P}$ *and* $\widehat{\Sigma}$ *satisfy the relative AEP, if and only if for any* $\varepsilon > 0$ *and* $\delta > 0$*, there exists a relative typical set* $Q_n^{\varepsilon,\delta} \subset \mathscr{X}_n$*, satisfying the following properties for sufficiently large n:*

(a) For any $X_n \in Q_n^{\varepsilon,\delta}$*,*

$$\exp(n(S_1(\widehat{P}\|\widehat{\Sigma}) - \varepsilon)) \leq \frac{\rho_n(X_n)}{\sigma_n(X_n)} \leq \exp(n(S_1(\widehat{P}\|\widehat{\Sigma}) + \varepsilon)). \qquad \text{(C.1)}$$

(b) $\rho_n[Q_n^{\varepsilon,\delta}] > 1 - \delta$*.*
(c) $(1 - \delta) \exp(-n(S_1(\widehat{P}\|\widehat{\Sigma}) + \varepsilon)) \leq \sigma_n[Q_n^{\varepsilon,\delta}] \leq \exp(-n(S_1(\widehat{P}\|\widehat{\Sigma}) - \varepsilon))$*.*

Here, $\rho_n[Q_n^{\varepsilon,\delta}]$ *and* $\sigma_n[Q_n^{\varepsilon,\delta}]$ *represent the probabilities of* $Q_n^{\varepsilon,\delta}$ *according to distributions* ρ_n *and* σ_n*, respectively.*

Next, we define classical ergodicity.

Definition C.3 (*Classical ergodicity*) $\{x_l\}_{l \in \mathbb{Z}}$ is translation invariant if μ is invariant under the shift operator $\mathscr{T}$. Moreover, a translation-invariant process $\{x_l\}_{l \in \mathbb{Z}}$ is ergodic, if any subset of K that is invariant under $\mathscr{T}$ has measure 0 or 1.

The above definition is a special case of quantum ergodicity (Definition 7.8) for the situation where the algebra of observables is commutative, which is highlighted by the following lemma.

Lemma C.1 *A two-sided stochastic process is translation invariant and ergodic, if and only if it is an extreme point of the set of translation-invariant stochastic processes.*

We now state the classical Shannon-McMillan(-Breiman) theorem. The weaker statement with convergence in probability is enough for our purpose in the following argument.

Theorem C.1 (Classical Shannon-McMillan theorem) *If $\widehat{P}$ is translation invariant and ergodic, $-\frac{1}{n}\ln\rho_n(X_n)$ converges to $S_1(\widehat{P})$ almost surely (and thus in probability, implying the AEP).*

In the case of i.i.d., the Shannon-McMillan theorem immediately follows from the law of large numbers, because $-\ln\rho(X_{2n+1}) = -\sum_{l=-n}^{n}\ln\rho(x_l)$. In general, the Birkhoff's ergodic theorem plays a crucial role in the proof of the Shannon-McMillan theorem (see, for example, Theorem 16.8.1 of [3]).

Theorem C.2 (Birkhoff's ergodic theorem) *Let f be a (measurable) function from K to $\mathbb{R}$. If $\{x_l\}_{l\in\mathbb{Z}}$ is translation invariant and ergodic, then for almost every $\omega\in K$,*

$$\lim_{n\to\infty}\frac{1}{2n+1}\sum_{l=-n}^{n}f(\mathcal{T}^l\omega) = \int d\mu(\omega)f(\omega). \tag{C.2}$$

The Shannon-McMillan theorem can be generalized to a relative version (see, e.g., Ref. [182]), if $\widehat{P}$ is ergodic and $\widehat{\Sigma}$ is Markovian (of finite order). Here, $\widehat{\Sigma}$ is Markovian of order m, if for any n the conditional probability satisfies $\sigma(x_n|x_{n-1}, x_{n-2}, \ldots) = \sigma(x_n|x_{n-1}, \ldots, x_{n-m})$ almost surely (e.g., Ref. [183]).

Proposition C.3 (Classical relative Shannon-McMillan theorem) *If $\widehat{P}$ is translation invariant and ergodic and $\widehat{\Sigma}$ is translation invariant and Markovian (of finite order), then by sampling X_n according to ρ_n, $\frac{1}{n}\ln\frac{\rho_n(X_n)}{\sigma_n(X_n)}$ converges to $S_1(\widehat{P}\|\widehat{\Sigma})$ almost surely (and thus in probability, implying that $\widehat{P}$ and $\widehat{\Sigma}$ satisfy the relative AEP).*

Proof We only need to consider the term of $\ln\sigma_n(X_n)$. From the assumption of Markovian, this term is decomposed into the sum of the logarithm of conditional probabilities almost surely, except for the surface term that does not contribute in the limit. We can then apply the Birkhoff's ergodic theorem.   □

Now, the following is the classical counterpart of Theorem 7.10.

Proposition C.4 *If $\widehat{P}$ and $\widehat{\Sigma}$ satisfy the relative AEP, we have*

$$\underline{S}(\widehat{P}\|\widehat{\Sigma}) = \overline{S}(\widehat{P}\|\widehat{\Sigma}) = S_1(\widehat{P}\|\widehat{\Sigma}). \tag{C.3}$$

We remark that the Gibbs state of a local and translation-invariant Hamiltonian in one dimension is translation invariant and Markovian (of finite order). This is essentially the Hammersley-Clifford theorem [184] (see also Ref. [185]), and the case of infinite systems (as considered here) is proved in Lemma 19 of Ref. [129].

We can directly prove Proposition C.4 from the definition of the spectral divergence rates (see Proposition 21 of Ref. [129]). Here, we instead present a proof of the classical Stein's lemma, from which Proposition C.4 follows by using Proposition B.3. The following proof is a slight modification of Theorem 11.8.3 of Ref. [3].

Proposition C.5 (Classical Stein's lemma) *If $\widehat{P}$ and $\widehat{\Sigma}$ satisfy the relative AEP, then for any $0 < \eta < 1$,*

$$S_{\mathrm{H}}^{\eta}(\widehat{P}\|\widehat{\Sigma}) = S_1(\widehat{P}\|\widehat{\Sigma}). \tag{C.4}$$

Proof We first note that the classical hypothesis testing divergence is given by

$$S_{\mathrm{H}}^{\eta}(\rho_n\|\sigma_n) := -\ln\left(\frac{1}{\eta}\min_{Q_n:\rho_n[Q_n]\geq\eta}\sigma_n[Q_n]\right). \tag{C.5}$$

Take $0 < \varepsilon < 1 - \eta$. Let Q_n^{ε} be defined by Proposition C.2 as $Q_n^{\varepsilon} := Q_n^{\varepsilon,\varepsilon}$. From Proposition C.2 (b), we have $\rho_n[Q_n^{\varepsilon}] > 1 - \varepsilon > \eta$, and thus Q_n^{ε} is a candidate for the minimization in Eq. (C.5). From this and Proposition C.2 (c), we obtain

$$S_{\mathrm{H}}^{\eta}(\rho_n\|\sigma_n) \geq -\ln\left(\frac{1}{\eta}\sigma_n[Q_n^{\varepsilon}]\right) \geq n\left(S_1(\widehat{P}\|\widehat{\Sigma}) - \varepsilon\right) + \ln\eta. \tag{C.6}$$

By taking the limit, $S_{\mathrm{H}}^{\eta}(\widehat{P}\|\widehat{\Sigma}) := \lim_{n\to\infty}\frac{1}{n}S_{\mathrm{H}}^{\eta}(\rho_n\|\sigma_n) \geq S_1(\widehat{P}\|\widehat{\Sigma}) - \varepsilon$. Since $\varepsilon > 0$ can be taken arbitrarily small, we have

$$S_{\mathrm{H}}^{\eta}(\widehat{P}\|\widehat{\Sigma}) \geq S_1(\widehat{P}\|\widehat{\Sigma}). \tag{C.7}$$

On the other hand, let Q_n be the optimal candidate for the minimization in Eq. (C.5) satisfying $\rho_n[Q_n] \geq \eta$. For any $0 < \varepsilon < \eta$, we can show in the same manner as Lemma 11.8.1 of Ref. [3] that

$$\sigma_n[Q_n] \geq \sigma_n[Q_n \cap Q_n^{\varepsilon}] \geq \rho_n[Q_n \cap Q_n^{\varepsilon}]\exp\left(-n(S_1(\widehat{P}\|\widehat{\Sigma}) + \varepsilon)\right) \tag{C.8}$$

$$\geq \left(\rho_n[Q_n] - (1 - \rho_n[Q_n^{\varepsilon}])\right)\exp\left(-n(S_1(\widehat{P}\|\widehat{\Sigma}) + \varepsilon)\right) \tag{C.9}$$

$$> (\eta - \varepsilon)\exp\left(-n(S_1(\widehat{P}\|\widehat{\Sigma}) + \varepsilon)\right), \tag{C.10}$$

where Q_n^{ε} is defined as above and we used Proposition C.2 (a) and (b). Therefore,

$$S_{\mathrm{H}}^{\eta}(\rho_n\|\sigma_n) = -\ln\left(\frac{1}{\eta}\sigma_n[Q_n]\right) < n\left(S_1(\widehat{P}\|\widehat{\Sigma}) + \varepsilon\right) + \ln\frac{\eta}{\eta - \varepsilon}. \tag{C.11}$$

By taking the limit, $S_{\mathrm{H}}^{\eta}(\widehat{P}\|\widehat{\Sigma}) \leq S_1(\widehat{P}\|\widehat{\Sigma}) + \varepsilon$. Since $\varepsilon > 0$ can be taken arbitrarily small, we have

$$S_{\mathrm{H}}^{\eta}(\widehat{P}\|\widehat{\Sigma}) \leq S_1(\widehat{P}\|\widehat{\Sigma}). \tag{C.12}$$

From inequalities (C.7) and (C.12), we obtain Eq. (C.4). $\square$

References

1. Callen, H.B.: Thermodynamics and an Introduction to Thermostatistics, 2nd Edition. John Wiley and Sons, New York (1985)
2. Lieb, E.H., Yngvasonb, J.: Phys. Rep. **314**, 669 (1999)
3. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd Edition. Wiley-Interscience (2006)
4. Nielsen, M. A., Chuang, I. L.: Quantum Computation and Quantum Information. Cambridge Univ. Press, Cambridge (2000)
5. Ohya, M., Petz, D.: Quantum entropy and its use. Springer, Berlin, 2nd edition (2004)
6. Hayashi, M.: Quantum Information: An Introduction. Springer-Verlag, Berlin (2006)
7. Bhatia, R.: Matrix Analysis. Springer (1991)
8. Hiai, F., Interdiscip. Inform. Sci. **16**, 139–248 (2010)
9. Shannon, C.: Bell System Technical Journal **27**, 379–423, 623–656 (1948)
10. von Neumann, J.: Mathematische Grundlagen der Quantumechanik. Springer, Berlin (1932) [Eng. trans. Beyer, R. T.: Mathematical Foundations of Quantum Mechanics. Prinston University Press, Princeton (1955)]
11. Kullback, S., Leibler, R. A.: Ann. Math. Stat. **22**, 79 (1951)
12. Umegaki, H.: Ködai Math. Sem. Rep. **14**, 59 (1962)
13. Rényi, A.: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability **1**, 547–561 (1961)
14. Ali, S.M., Silvey, S.D.: J. Roy. Stat. Soc. Ser. B **28**, 131–142 (1966)
15. Csiszár, I.: Studia Scientiarum Mathematicarum Hungarica, **2**, 299–318 (1967)
16. Petz, D.: Publ. RIMS. Kyoto Univ. **21**, 781–800 (1985)
17. Petz, D.: Rep. Math. Phys. **23**, 57–65 (1986)
18. Fisher, R.A.: Proc. Camb. Phil. Soc. **22**, 700 (1925)
19. Petz, D.: Linear Algebra and its Applications **244**, 81–96 (1996)
20. Amari, S., Nagaoka, H.: Methods of Information Geometry. Oxford University Press, New York (2000), American Mathematical Society, Providence (2007)
21. Toyabe, S., Sagawa, T., Ueda, M., Muneyuki, E., Sano, M.: Nat. Phys. **6**, 988 (2010)
22. Bérut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R., Lutz, E.: Nature **483**, 187–189 (2012)
23. Ribezzi-Crivellari, M., Ritort, F.: Nat. Phys. **15**, 660–664 (2019)
24. Koski, J.V., Maisi, V.F., Sagawa, T., Pekola, J.P.: Phys. Rev. Lett. **113**, 030601 (2014)

25. Masuyama, Y., *et al.*: Nat. Commu. **9**,1291 (2018)
26. Camati, P. A. *et al.*: Phys. Rev. Lett. **117**, 240502 (2016)
27. Evans, D.J., Cohen, E.G.D., Morriss, G.P.: Phys. Rev. Lett. **71**, 2401 (1993)
28. Jarzynski, C.: Phys. Rev. Lett. **78**, 2690 (1997)
29. Crooks, G.E.: Phys. Rev. E **60**, 2721 (1999)
30. Seifert, U.: Phys. Rev. Lett. **95**, 040602 (2005)
31. Jarzynski, C.: Annu. Rev. Condens. Matter Phys. **2**, 329 (2011)
32. Seifert, U.: Rep. Prog. Phys. **75**, 126001 (2012)
33. Esposito, M., Harbola, U., Mukamel, S.: Rev. Mod. Phys. **81**, 1665 (2009)
34. Sagawa, T.: arXiv:1202.0983v3 (2012) [Chapter of: Lectures on Quantum Computing, Thermodynamics and Statistical Physics. Kinki University Series on Quantum Computing, World Scientific (2012)]
35. Funo, K., Ueda, M., Sagawa, T.: Quantum Fluctuation Theorems; [Chapter of: Binder, F., Correa, L., Gogolin, C., Anders, J., Adesso G. (eds): Thermodynamics in the Quantum Regime. Fundamental Theories of Physics, **195**. Springer, Cham (2018)]
36. Sagawa, T.: Prog. Theor. Phys. **127**, 1 (2012)
37. Parrondo, J. M. R., Horowitz, J. M., Sagawa, T.: Nat. Phys. **11**, 131 (2015)
38. Sagawa, T.: Second law, entropy production, and reversibility in thermodynamics of information; [Chapter of: Snider G. *et al.* (eds.): Energy Limits in Computation: A Review of Landauer's Principle, Theory and Experiments. pp. 101–139, Springer (2019)]
39. Leff, H. S., Rex, A.F. (eds.): Maxwell's demon 2: Entropy, Classical and Quantum Information, Computing. Princeton University Press, New Jersey (2003)
40. Brandão, F.G.S.L., Horodecki, M., Oppenheim, J., Renes, J.M., Spekkens, R.W.: Phys. Rev. Lett. **111**, 250404 (2013)
41. Horodecki, M., Oppenheim, J.: Nat. Commun. **4**, 2059 (2013)
42. Åberg, J.: Nat. Commu. **4**, 1925 (2013)
43. Brandão, F.G.S.L., Horodecki, M., Ng, N.H.Y., Oppenheim, J., Wehner, S.: Proc. Natl. Acad. Sci. **112**, 3275 (2015)
44. Weilenmann, M., Kraemer, L., Faist, P., Renner, R.: Phys. Rev. Lett. **117**, 260601 (2016)
45. Faist, P., Renner, R.: Phys. Rev. X **8**, 021011 (2018)
46. Gour, G., Müller, M. P., Narasimhachar, V., Spekkens, R.W., Yunger Halpern, N.: Phys. Rep. **583** 1–58 (2015)
47. Goold, J., Huber, M., Riera, A., del Rio, L., Skrzypzyk, P.: J. Phys. A: Math. Theor. **49**, 143001 (2016)
48. Lostaglio, M.: Rep. Prog. Phys. **82**, 114001 (2019)
49. Alhambra, Á.M., Lostaglio, M., Perry, C.: Quantum **3**, 188 (2019)
50. Yunger Halpern, N., Limmer, D.T.: Phys. Rev. A **101**, 042116 (2020)
51. Chitambar E., Gour, G.: Rev. Mod. Phys. **91**, 025001 (2019)
52. Nielsen, M.A.: Phys. Rev. Lett. **83**, 436 (1999)
53. Takagi R., Regula, B.: Phys. Rev. X **9**, 031053 (2019)
54. Liu, Z.-W., Bu, K., Takagi, R.: Phys. Rev. Lett. **123**, 020401 (2019)
55. Marshall, A.W., Olkin, I., Arnold, B.C.: Inequalities: Theory of Majorization and Its Applications. Springer (2011)
56. Blackwell, D.: The Annals of Mathematical Statistics **24**, 265 (1953)
57. Torgersen, E.: Comparison of Statistical Experiments. Encyclopedia of Mathematics and its Applications, Cambridge University Press (1991)
58. Ruch, E., Schranner, R., Seligman, T.H.: J. Chem. Phys. **69** 386–392 (1978)
59. Han, T.S., Verdú, S.: IEEE Trans. on Inf. Theory **39**, 752–772 (1993)
60. Han, T.S.: Information-Spectrum Methods in Information Theory. Springer (2003) [The original version is in Japanese published from Baifukan-Press, Tokyo (1998)]
61. Nagaoka, H., Hayashi, M.: IEEE Trans. on Inf. Theory **53**, 534–549 (2007)
62. Bowen G., Datta, N.: Proc. IEEE Int. Symp. Info. Theory, page 451 (2006)
63. Bowen G., Datta, N.: arXiv:quant-ph/0610003 (2006)
64. Datta, N., Renner, R.: IEEE Trans. on Inf. Theory **55**, 2807-2815 (2009)

65. Datta, N.: IEEE Trans. on Inf. Theory **55**, 2816–2826 (2009)
66. Renner, R.: PhD thesis, ETH Zurich (2005)
67. Renner R., Wolf, S.: International Symposium on Information Theory, Proceedings. pp. 233 (2004)
68. Van Erven, T., Harremoës, P.: arXiv:1001.4448 (2010)
69. Van Erven, T., Harremoës, P.: IEEE Trans. on Inf. Theory **60**, 3797-3820 (2014)
70. Liese F., Vajda, I.: IEEE Trans. on Inf. Theory **52**, 4394–4412, (2006)
71. Ito, S., Dechant, A.: Phys. Rev. X **10**, 021056 (2020)
72. Chentsov (Čencov), N.N.: Statistical Decision Rules and Optimal Inference. American Mathematical Society, Rhode Island, U.S.A. (1982) [Originally published in Russian, Nauka, Moscow (1972)].
73. Renes, J. M.: J. Math. Phys. **57**, 122202 (2016)
74. Shiraishi, N.: J. Phys. A: Math. Theor. **53**, 425301 (2020)
75. Jonathan, D., Plenio, M.B.: Phys. Rev. Lett. **83**, 3566–3569 (1999)
76. Turgut, S.: J. Phys. A: Math. Theor. **40**, 12185–12212 (2007)
77. Klimesh, M.: arXiv:0709.3680 (2007)
78. Aubrun, G., Nechita, I.: Commu. Math. Phys. **278**, 133–144 (2008)
79. van Dam, W., Hayden, P.: Phys. Rev. A **67**, 060302 (2003)
80. Müller, M.P., Pastena, M.: IEEE Trans. on Inf. Theory **62**, 1711–1720 (2016)
81. Lostaglio, M., Müller, M.P., Pastena, M.: Phys. Rev. Lett. **115**, 150402 (2015)
82. Gallego, R.; Eisert, J.; Wilming, H.: New J. Phys. **18**, 103017 (2016)
83. Wilming, H., Gallego, R., Eisert, J.: Entropy **19**, 241 (2017)
84. Müller, M.P.: Phys. Rev. X **8**, 041051 (2018)
85. Rethinasamy, S., Wilde, M.M.: Phys. Rev. Research **2**, 033455 (2020)
86. Lostaglio, M., Müller, M.P.: Phys. Rev. Lett. **123**, 020403 (2019)
87. Shiraishi, N., Sagawa, T.: Phys. Rev. Lett. **126**, 150502 (2021)
88. Ryff, J. V.: Transactions of the American Mathematical Society, **117**, 92–100 (1965)
89. Ryff, J. V.: Pacific J. Math. **13**, 1379–1386 (1963)
90. Esposito, M., Van den Broeck, C.: Europhys. Lett. **95**, 40004 (2011)
91. Kondepudi, D. and Prigogine, I.: From Heat Engines to Dissipative Structures. John Wiley and Sons, New York (1998)
92. Hatano, T., Sasa, S.-i.: Phys. Rev. Lett. **86**, 3463 (2001)
93. Esposito, M., Van den Broeck, C.: Phys. Rev. Lett. **104**, 090601 (2010)
94. Horowitz, J.M., Parrondo, J.M.R.: New J. Phys. **15**, 085028 (2013)
95. Horowitz, J.M., Sagawa, T.: J. Stat. Phys. **156**, 55–65 (2014)
96. Landauer, R.: IBM J. Res. Dev. **5**, 183 (1961)
97. Sagawa, T., Ueda, M.: Phys. Rev. Lett. **109**, 180602 (2012)
98. Sagawa, T., Ueda, M.: New J. Phys. **15** 125012 (2013)
99. Jacobs, K.: Phys. Rev. A **80**, 012322 (2009)
100. Peres, A.: Phys. Rev. Lett. **77**, 1413 (1996)
101. Plenio, M.B.: Phys. Rev. Lett. **95**, 090503 (2005)
102. Choi, M.-D.: Linear Algebra and its Applications **10**, 285–290 (1975)
103. Lieb, E.H., Ruskai, M.B.: Phys. Rev. Lett. **30**, 434 (1973)
104. Lieb, E.H., Ruskai, M.B.: J. Math. Phys. **14**, 1938 (1973)
105. Lieb, E.H.: Advances in Mathematics, **11**, 267 (1973)
106. Ruskai, M.B.: J. Math. Phys. **43**, 4358 (2002)
107. Lindblad, G.: Commun. Math. Phys. **39**, 111 (1974)
108. Lindblad, G.: Commun. Math. Phys. **40**, 147 (1975)
109. Uhlmann, A.: Commun. Math. Phys. **54**, 21 (1977)
110. Petz, D.: Rev. Math. Phys. **15**, 79 (2003)
111. Nielsen, M.A., Petz, D.: Quantum Inf. Comput. **5**, 507 (2005)
112. Hiai, F., Mosonyi, M., Petz, D., Bény, C.: Rev. Math. Phys. **23**, 691–747 (2011)
113. Hiai, F.: Quantum f-Divergences in von Neumann Algebras: Reversibility of Quantum Operations. Springer (2021) Operator Algebras and Quantum Statistical Mechanics 1: C^*- and W^*-algebras. Symmetry Groups, Decomposition of States. Springer, New York (1979)

114. Tomamichel, M.: PhD thesis, ETH Zurich; arXiv:1203.2142 (2012)
115. Wilde, M. M., Winter, A., Yang D.: Commu. Math. Phys. **331**, 593–622 (2014)
116. Muller-Lennert, M., Dupuis, F., Szehr, O., Fehr, S., Tomamichel, M.: J. Math. Phys. **54**, 122203 (2013)
117. Frank, R. L., Lieb, E. H., arXiv:1306.5358 (2013)
118. Beigi, S.: J. Math. Phys. **54**, 122202 (2013)
119. Mitsuhashi, Y.: Private communication (2020)
120. Landau, L.J., Streater, R.F.: Linear Algebra and its Applications **193**, 107–127 (1993)
121. Shor, P.W.: Structure of Unital Maps and the Asymptotic Quantum Birkhoff Conjecture. Presentation (2010) http://science-visits.mccme.ru/doc/steklov-talk.pdf
122. Haagerup U., Musat, M.: Commun. Math. Phys. **303**, 555 (2011)
123. Mendl, C.B., Wolf, M.M.: Commun. Math. Phys. **289**, 1057 (2009)
124. Ji, Z., Natarajan, A., Vidick, T., Wright, J., Yuen, H.: arXiv:2001.04383 (2020)
125. Buscemi, F., Gour, G.: Phys. Rev. A **95**, 012110 (2017)
126. Gour, G., Jennings, D., Buscemi, F., Duan, R., Marvian, I.: Nat. Commu. **9**, 5352 (2018)
127. Oppenheim, J., Horodecki, M., Horodecki, P., Horodecki, R.: Phys. Rev. Lett. **89**, 180402 (2002)
128. Horodecki, M., Oppenheim, J., Horodecki, R.: Phys. Rev. Lett. **89**, 240403 (2002)
129. Sagawa, T., Faist, P., Kato, K., Matsumoto, K., Nagaoka, H., Brandão, F. G. S. L.: J. Phys. A: Math. Theor. **54**, 495303 (2021)
130. Heinosaari, T., Jivulescu, M.A., Reeb, D., Wolf, M.M.: J. Math. Phys. **53**, 102208 (2012)
131. Matsumoto, K.: arXiv:1409.5658 (2014)
132. Ito, K., Kumagai, W., Hayashi, M.: Phys. Rev. A **92**, 052308 (2015)
133. Kumagai, W., Hayashi, M.: arXiv:1306.4166v4 (2013)
134. Hayashi, M.: IEEE Trans. on Inf. Theory **54**, 4619–4637 (2008)
135. Hayashi, M.: IEEE Trans. on Inf. Theory **55**, 4947–4966 (2009)
136. Datta, N., Leditzky, F.: IEEE Trans. on Inf. Theory **61**, 582–608 (2015)
137. Tomamichel, M., Berta, M., Renes, J.M.: Nat. Commun. **7**, 11419 (2016)
138. Matsumoto, K.: arXiv:1010.1030 (2010)
139. Jiao, Y., Wakakuwa, E., Ogawa, T.: arXiv:1701.09050 (2017)
140. Bratteli, O., Robinson, D.W.: Operator Algebras and Quantum Statistical Mechanics 1: C^*- and W^*-algebras. Symmetry Groups, Decomposition of States. Springer, New York (1979)
141. Bratteli, O., Robinson, D.W.: Operator Algebras and Quantum Statistical Mechanics 2: Equilibrium States. Models in Quantum Statistical Mechanics. Springer, New York (1981)
142. Ruelle, D.: Statistical Mechanics: Rigorous Results. World Scientific, Singapore (1999)
143. Araki, H.: Commun. Math. Phys. **14**, 120–157 (1969)
144. Tasaki, H.: J. Stat. Phys. **172**, 905 (2018)
145. Bjelaković, I., Krueger, T., Siegmund-Schultze, R., Szkola, A.: Inventiones mathematicae **155**, 203–222 (2004)
146. Bjelaković, I., Szkola, A.: Quantum Information Processing **4**, 49–63 (2005)
147. Ogata, Y.: Lett. Math. Phys. **103**, 1367–1376 (2013)
148. Schoenmakers, B., Tjoelker, J., Tuyls, P., Verbitskiy, E.: Proc. ISIT, 256–260 (2007)
149. Hiai, F., Petz, D.: Commun. Math. Phys. **143**, 99 (1991)
150. Ogawa, T., Nagaoka, H.: IEEE Trans. on Inf. Theory **46**, 2428 (2000)
151. Bjelaković, I., Siegmund-Schultze, R.: Commun. Math. Phys.**247** 697–712 (2004)
152. Faist, P., Sagawa, T., Kato, K., Nagaoka, H., Brandão, F. G. S. L.: Phys. Rev. Lett. **123**, 250601 (2019)
153. Dupuis, F., Kraemer, L., Faist, P., Renes, J. M., Renner, R.: Proceedings of the XVIIth International Congress on Mathematical Physics, 134–153 (2013)
154. Mosonyi, M., Ogawa, T.: Commun. Math. Phys. **334**, 1617–1648 (2015)
155. Pusz, W., Woronowicz, S.L.: Commu. Math. Phys. **58**, 273–290 (1978)
156. Lenard, A.: J. Stat. Phys. **19**, 575–586 (1978)
157. Breuer, H.-P., Petruccione, F.: The Theory of Open Quantum Systems. Oxford University Press (2002)

158. Faist, P., Oppenheim, J., Renner, R.: New J. Phys. **17**, 043003 (2015)
159. Åberg, J.: Phys. Rev. Lett. **113**, 150402 (2014)
160. Skrzypczyk, P., Short, A. J., Popescu, S.: Nat. Commu. **5**, 4185 (2014)
161. Malabarba, A. S. L., Short, A. J., Kammerlander, P.: New J. Phys. **17**, 045027 (2015)
162. Korzekwa, K., Lostaglio, M., Oppenheim, J., Jennings, D.: New J. Phys. **18** 023045 (2016)
163. Woods, M. P., Silva, R., Oppenheim, J.: Annales Henri Poincare **20**, 125–218 (2019)
164. Mitsuhashi, Y., Kaneko, K., Sagawa, T.: arXiv:2103.06060, to appear in Phys. Rev. X (2022)
165. Frenzel, M. F., Jennings, D., Rudolph, T.: Phys. Rev. E **90**, 052136 (2014)
166. Tajima, H., Shiraishi, N., Saito, K.: Phys. Rev. Lett. **121**, 110403 (2018)
167. Alhambra, Á.M., Masanes, L., Oppenheim, J., Perry, C.: Phys. Rev. X **6**, 041017 (2016)
168. Hayashi, M., Tajima, H.: Phys. Rev. A **95**, 032132 (2017)
169. Skrzypczyk, P., Short, A.J., Popescu, S.: arXiv:1302.2811 (2013)
170. Dahlsten, O.C.O., Renner, R., Rieper, E., Vedral, V.: New J. Phys. **13** 053015 (2011)
171. del Rio, L., Åberg, J., Renner, R., Dahlsten, O., Vedral, V.: Nature **474**, 61 (2011)
172. Faist, P., Dupuis, F., Oppenheim, J., Renner, R.: Nat. Commu. **6**, 7669 (2015)
173. Loomis, S.P., Crutchfield, J.P.: Phys. Rev. Lett. **125**, 020601 (2020)
174. Brandão, F. G.S.L., Gour, G.: Phys. Rev. Lett. **115**, 070503 (2015)
175. Choi, M.-D.: Illinois J. Math. **18**, 565 (1974)
176. Hansen, F., Pedersen , G.K.: Mathematische Annalen **258**, 229–241(1982)
177. Muller-Hermes, A., Reeb, D.: Annales Henri Poincare **18**, 1777–1788 (2017)
178. Hayashi, M. (ed.): Asymptotic Theory of Quantum Statistical Inference. World Scientific (2005)
179. Wang, L., Renner, R.: Phys. Rev. Lett. **108**, 200501 (2012)
180. Watrous, J.: Theory of Computing **5**, 217–238 (2009)
181. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge University Press (2004)
182. Algoet, P. H., Cover, T. M.: Annals of Prob. **16**, 899–909 (1988)
183. Doob, J.L.: Stochastic Processes. Wiley-Interscience; Revised edition (1990)
184. Hammersley, J.M., Clifford, P.E.: Markov field on finite graphs and lattices. (1971) http://www.statslab.cam.ac.uk/?grg/books/hammfest/hamm-cliff.pdf
185. Kato, K., Brandão, F.G.S.L.: Commu. Math. Phys. **370**, 117–149 (2019)