國立臺灣大學電機資訊學院電信工程學研究所
博士論文
Graduate Institute of Communication Engineering
College of Electrical Engineering and Computer Science
National Taiwan University
Doctoral Dissertation

量子資訊理論中的錯誤率分析
Error Exponent Analysis in Quantum Information Theory

鄭皓中
Hao-Chung Cheng

指導教授：葉丙成博士
Advisor: Ping-Cheng Yeh, Ph.D.

中華民國 107 年 1 月
January, 2018

國立臺灣大學博士學位論文
口試委員會審定書

量子資訊理論中的錯誤率分析
Error Exponent Analysis in Quantum Information Theory

本論文係鄭皓中君（學號 F99942118）在國立臺灣大學電信工程學研究所完成之碩（博）士學位論文，於民國 107 年 1 月 9 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

（簽名）

（指導教授）

所　　長　　（簽名）

doi:10.6342/NTU201800597

Acknowledgement (致謝)

I would like to thank my supervisor – Professor Ping-Cheng Yeh – for his valuable guidance and being my life beacon. From my undergraduate in NTUEE, Prof. Yeh not only supervised me how to do research, but also taught me to be a decent person with integrity. I would also like to thank my co-supervisor – Professor Min-Hsiu Hsieh – for teaching me all the knowledge in quantum information theory and leading me to academics. I would like to thank all the staffs in NTUEE and GICE for your constant helps. I would like to thanks all the members in BL515. So glad to be able to study, play, and do research with you all.

I'm honored to be a part of National Taiwan University Chinese Orchestra (NTUCO). Thanks all the members in NTUCO every year. It was a great time to play music with you. Thanks for supporting me for my three concertos of Bamboo flute, Souna, and Guan. Thanks for supporting me in semester 2012-2013 for my conducting. Thanks for Little Giant Chamber Orchestra (小巨人絲竹樂團), Taipei Youth Chinese Orchestra (臺北青年國樂團), and Youth Orchestra in Taipei Chinese Orchestra (臺北市立國樂團附設青年團). You make my undergraduate and PhD life more colorful.

I would like to sincerely thank my family – both my parents and my senior brother. Thank you for letting me pursue what I want. I will be eternally grateful for all of your supports.

Lastly, I would like to thank all the funding that support me through my study: Ministry of Science and Technology Overseas Project for Post Graduate Research with Grant 105-2917-I-002-028 and 104-2221-E-002-072; Hua Gu scholarship (財團法人華固教育基金會獎助學金), E-Sun Bank scholarship (玉山銀行培育傑出管理人才獎學金), Hsing Tian Kong scholarship (財團法人行天宮資優生長期培育), and Ho Foundation (臺大何宜慈博士紀念獎學金).

doi:10.6342/NTU201800597

中文摘要

資訊理論中最基本的問題之一是刻劃三個重要參數的取捨—資訊處理的品質優劣、錯誤更正碼的區塊長度、以及傳輸率。錯誤率指數分析即為一個強大且有效的方法來研究當傳輸率固定時錯誤概率如何隨著編碼區塊增大進而指數遞減。在本論文中，我們討論兩個重要的量子資訊處理協定—經由量子資訊的協助來壓縮經典數據、以及經典數據經由量子信道傳輸—之錯誤率指數分析。

我們首先證明錯誤指數函數的諸多重要性質，使我們得以更深刻理解量子資訊協定的錯誤率行為模式。第二、在有限的區塊編碼長度下我們對研究的兩種量子資訊協定求得精確的錯誤率分析，為次世代量子資訊科技的設計提供了更佳的品質估計準則。最後，我們研究當傳輸率趨近一些重要的閾值時的錯誤概率行為—當壓縮率緩慢逼近條件滴值被時壓縮的經典數據得以被完美恢復、以及當傳輸率緩慢逼近信道容量時數據傳輸得以無暇傳輸。

此論文呈現方式力求以經典資訊理論的架構來撰寫，因此者不限於具有量子資訊理論的學者。工程司、科技設計者以及任何對量子資訊理論有興趣者皆能藉由閱讀此論文來探索此豐富且深邃的研究課題。

關鍵字：錯誤指數分析、中偏差分析、大偏差分析、量子資訊理論、經典量子信道、量子輔助資訊、可靠度函數、矩陣分析

doi:10.6342/NTU201800597

Abstract

One of the fundamental problems in information theory is to clarify the trade-offs between the performance of an information task, the size of the coding scheme, and the coding rate that determines the efficiency of the task. Error exponent analysis was proposed as a powerful methodology to study how rapidly the error probability exponentially decays with an increase of coding blocklength when the rate is fixed. In this thesis, we give an exposition of error exponent analysis to two important quantum information processing protocols—classical data compression with quantum side information, and classical communications over quantum channels.

We first prove substantial properties of various exponent functions, which allow us to better characterize the error behaviors of the tasks. Second, we establish accurate achievability and optimality finite blocklength bounds for the optimal error probability, providing useful and measurable benchmarks for future quantum information technology design. Finally, we study the error probability under the scenario that the coding rate converges to certain limits, a research topic known as moderate deviation analysis. In other words, we show that the data recovery can be perfect when the compression rate approaches the conditional entropy slowly, and the reliable communication over a classical-quantum channel is possible as the transmission rate approaches channel capacity slowly.

The audience of this thesis are not restricted to researchers with backgrounds in quantum information theory. Engineers, technology providers, and people who interest in information processing are welcome to explore the frontiers along this line of research.

Keywords: error exponent analysis, moderate deviation analysis, quantum information theory, classical-quantum channel, Slepian-Wolf coding, quantum side information, reliability function, large deviation theory, matrix analysis.

doi:10.6342/NTU201800597

Contents

doi:10.6342/NTU201800597

doi:10.6342/NTU201800597

List of Figures

doi:10.6342/NTU201800597

List of Tables

doi:10.6342/NTU201800597

Chapter 1

Introduction

Information processing and transmission with the assistance of quantum mechanics has emerged as a promising technology in the forthcoming future. For example, Bennett and and Brassard [1, 2] proposed a quantum key distribution protocol, which provides us a secure way for sharing secret keys between two parties. The task of quantum teleportation is to noiselessly transfer a quantum state to a remote user [3] and it has become a key ingredient of quantum computation [4, 5]. In view of the latest and most significant achievements of communicating quantumly from a launched satellite with base stations [6], it is generally believed that the laboratory testing of novel quantum communication experiments will soon be complete.

To practically implement such quantum information processing (QIP) technologies, it would require universal quantum computation as the principle component (e.g. to perform the decoding strategies). Nevertheless, the state-of-the-art quantum computers, at least for the near future, is limited to around 50 qubits. Thus, evaluating how well a QIP system in practical domains only with finite resources becomes a pressing matter [7, 8]. The goal of this thesis is to investigate two fundamental QIP tasks and characterize their performance benchmarks, providing invaluable guidance to the design of the next-generation quantum technology.

In this thesis, we are interested in the QIP protocols that benefits and advances current information processing systems. Namely, we study the problems of (1) *information storage with a quantum helper*, and (2) *information transmission over a quantum channel*. Due to the non-cloning and probabilistic nature of quantum mechanics, the processing errors are inevitable. Therefore, our ultimate goal is to provide an accurate error analysis for theses QIP protocols. In Section 1.1, we give the backgrounds and literature review of this research topic. In Section 1.2, we introduce the mathematical formalisms of the studied QIP protocols. Our contributions are listed in Section 1.3. Lastly, we illustrate the structure of the thesis in Section 1.4.

1.1 Backgrounds

One of the core purposes in information theory is to protect the information when compressing and transmitting. In Shannon's seminal work [9], it was shown that the reliable communication over a channel is possible, provided that the transmission rate is below the *channel capacity* C, and an arbitrary large coding scheme is given. On the other hand, Sleipian and Wolf [10] studied a source compression scenario with an assistance of the side information. Let X denote the random variable

doi:10.6342/NTU201800597

of the source and Y be that of the side information. They showed that the perfect source recovery is feasible as long as the compression rate is above the *conditional entropy* $H(X|Y)$ and an arbitrary large block code is provided.

Therefore, investigating the interplay between the compression/transmission rate, coding blocklength and the probability of error is one of the fundamental problems in information theory. Based on different ranges of the error probability, analysis of the information processing performance roughly falls into the following three categories: (i) *large error probability* or *non-vanishing error probability* regime; (ii) *medium error probability* regime; and (iii) *small error probability* regime.

In the non-vanishing error probability regime, the largest code rate, given a coding length n and an error probability no more than ϵ, is one of the main research focuses. Strassen [11] first demonstrated that the maximum size of an n-blocklength code through a discrete memoryless channel (DMC) $\mathcal{W}$, denoted by $M^*(\mathcal{W}^n, \epsilon)$, yields an asymptotic expansion to the order $\sqrt{n}$, and hence this is called *second-order analysis*:

$$\log M^*(\mathcal{W}^n, \epsilon) = nC + \sqrt{nV}\,\Phi^{-1}(\epsilon) + O(\log n), \tag{1.1}$$

where the quantities C and V denote the capacity [9] and the dispersion [12] of the channel, and Φ is the cumulative distribution function of a standard normal random variable. Equivalently, Eq. (1.1) yields the following relationship between the optimal decoding error with blocklength n and rate $C - A/\sqrt{n}$ for any constant A:

$$\lim_{n\to+\infty} \epsilon^*\left(n, C - A/\sqrt{n}\right) = \Phi\left(\frac{A}{\sqrt{V}}\right). \tag{1.2}$$

Strassen's result relied on the *Gaussian approximation* or the *central limit theorem (CLT)*, and is also called the *small deviation regime*. His work was latter refined by Hayashi [13], Polyanskiy *et al.* [12], and extended to quantum channels [14, 15, 16, 7]. The results for higher-order asymptotics are referred to Refs. [17, 18, 19].

In the *small error probability* regime, Shannon [20] introduced the *reliability function* $E(R)$ as the optimal error exponent:

$$\lim_{n\to+\infty} -\frac{1}{n}\log \epsilon^*\left(n, R\right) = E(R), \tag{1.3}$$

for rate R below the channel capacity[1] C. The quantity $E(R)$ then provides a measure of how rapidly the error probability approaches zero with an increase in blocklength. This characterization of the reliability function is hence called the *reliability function analysis* or the *error exponent analysis*. This seminal work entails the analysis of a broad class of channels [22, 21, 23, 24, 25, 26]. The exponential decay of the error probability in Eq. (1.3) is a consequence of the *large deviation principle (LDP)* [27]. In summary, the errors in Eqs. (1.2) and (1.3), respectively, fall into the CLT regime and large-deviation regime.

Given a classical channel, lower bounds for the reliability function (termed *achievability*), can be established by random coding arguments [28, 22, 29, 21]. However, upper bounds (also called

[1] To the best of our knowledge, the reliability function $E(R)$ is only known in the high rate regime, i.e. at rates above a *critical rate* (see e.g. [21, p. 160]).

doi:10.6342/NTU201800597

optimality) require different techniques since the code-dependent bounds on the error probability need to be optimized over all codebooks. The first result—the *sphere-packing bound* $E(R) \leq E_{\mathrm{sp}}(R)$—was developed by Shannon, Gallager, and Berlekamp [30]. The *sphere-packing exponent* $E_{\mathrm{sp}}(R)$ is defined as

$$E_{\mathrm{sp}}(R) := \sup_{s \geq 0} \left\{ \max_{P} E_0(s, P) - sR \right\}, \tag{1.4}$$

where P is maximized over all probability distributions on the input alphabet, and $E_0(s, P)$ is the *auxiliary function* or *Gallager's function* [29]. Unlike Shannon-Gallager-Berlekamp's technique which relates channel coding to binary hypothesis testing, Haroutunian [31, 24] employed a combinatorial method and obtained an upper bound for the reliability function in terms of the following expression

$$\widetilde{E}_{\mathrm{sp}}(R) := \max_{P} \min_{\bar{\mathcal{W}}} \left\{ D\left(\bar{\mathcal{W}} \| \mathcal{W} | P\right) : I(P, \bar{\mathcal{W}}) \leq R \right\}, \tag{1.5}$$

where $\bar{\mathcal{W}}$ is minimized over all dummy channels with the same output alphabet as $\mathcal{W}$, $D(\bar{\mathcal{W}} \| \mathcal{W} | P)$ is the conditional relative entropy between the dummy channel $\bar{\mathcal{W}}$ and the true channel $\mathcal{W}$, and $I(P, \bar{\mathcal{W}})$ is the mutual information of the channel $\bar{\mathcal{W}}$ (the detailed definitions are given in Chapter 3). It was later realized that the two quantities in Eqs. (1.4) and (1.5) are equivalent: they are related by convex program duality [32, 33, 25]. Therefore, these two expressions, Eqs. (1.4) or (1.5), are both called sphere-packing exponents.

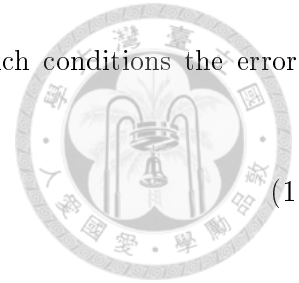
Error exponent analysis in classical-quantum (c-q) channels is more challenging because of the noncommutative nature of quantum mechanics. Burnashev and Holevo [34] introduced a quantum version of the auxiliary function [35, 36] and initialized the study of reliability functions in c-q channels. However, the random coding bound (i.e. achievability) for c-q channels is still unsolved. Winter [37] derived a sphere-packing bound (i.e. optimality) for c-q channels in the form of $\widetilde{E}_{\mathrm{sp}}(R)$ in Eq. (1.5), generalizing Haroutunian's idea [31]. Dalai [38] employed Shannon-Gallager-Berlekamp's approach [30] to establish a sphere-packing bound with Gallager's exponent in Eq. (1.4). In the follow-up work [39], Dalai and Winter pointed out that these two exponents are not equal in c-q channels. We remark that both Dalai and Winter's results are asymptotic and not finite blocklength.

The Slepian-Wolf coding with quantum side information (QSI) was studied by Devetak and Winter [40]. They generalized Slepian and Wolf's result [10] to the quantum case: the optimal probability of error asymptotically vanishes as the compression rate is above the *quantum conditional entropy* $H(X|B)_\rho$, where B denotes the quantum system. Similar to the role of channel capacity in channel coding, we term $H(X|B)_\rho$ the *Slepian-Wolf limit*. The non-vanishing error probability regime was later studied by Renes and Renner [41], and Tomamichel and Hayashi [14]. A second-order asymptotics similar Eq. (1.1) was established.

The most paragraph of this thesis will focus on the error exponent analysis for both Slepian-Wolf coding with QSI and classical-quantum channel coding. We especially focus on the finite blocklegnth characterizations of the optimal error probability. In Chapters 6 and 7, we establish finite blocklength bounds for Slepian-Wolf coding with QSI. In Chapters 10 and 11, we review the best-to-date achievability bound for c-q channel coding, and prove a tight sphere-packing bound in finite blocklengths.

The study of the medium error probability regime was pioneered by Altuğ and Wagner [42, 43]. They investigated the asymptotic behavior of the optimal decoding error when the coding rate con-

doi:10.6342/NTU201800597



verges to capacity sufficiently slowly. Specifically, they studied under which conditions the error is asymptotically equal to[2]

$$\epsilon^*\left(n, C - a_n\right) \sim \Phi\left(\frac{\sqrt{n}a_n}{\sqrt{v}}\right) \sim \mathrm{e}^{\frac{-na_n^2}{2v}}, \tag{1.6}$$

where the sequence of positive numbers $(a_n)_{n\in\mathbb{N}}$ satisfies

$$\begin{aligned} &\text{(i)} \lim_{n\to+\infty} a_n = 0; \\ &\text{(ii)} \lim_{n\to+\infty} a_n\sqrt{n} = +\infty. \end{aligned} \tag{1.7}$$

Evidently, the transmission rate in Eq. (1.6) approaches capacity slower than $1/\sqrt{n}$. A DMC with errors satisfying Eq. (1.6) possesses a *moderate deviation property (MDP)* [27, Section 3.7], and hence it is also called the *moderate deviation* regime. The constant v in Eq. (1.6) equals the channel dispersion V when both the limit in Eq. (1.2) and MDP hold [44, Theorem 1]. We refer the interested readers to Refs. [44, 45, 46, 47, 43] for further results in classical channel coding.

As an application of our established error exponent bounds, we extend our techniques to the moderate deviation regime. In Chapters 8 and 12, we demonstrate that the optimal error probability of the both two QIP tasks vanishes when the compression rate approaches the Slepian-Wolf limit and the transmission rate approaches the channel capacity, respectively. Specifically, we show that

$$\lim_{n\to+\infty} \frac{\log \epsilon^*(n, H(X|B)_\rho + a_n)}{na_n^2} = -\frac{1}{2V}; \tag{1.8}$$

$$\lim_{n\to+\infty} \frac{\log \epsilon^*(n, C - a_n)}{na_n^2} = -\frac{1}{2V}, \tag{1.9}$$

where $(a_n)_{n\in\mathbb{N}}$ is any sequence satisfying Eq. (1.7).

We remark that these error probability regime described above—(i), (ii), and (iii)—all have theoretical significance and practical value. The non-vanishing error probability regime, (i), has been relatively well studied in the quantum scenario, while the small and medium error probability, (ii) and (iii), are rarely explored, which is the ultimate goal and purpose of this thesis. We summarize the error behaviors in these three regimes in Table 1.1.

Our methodology contains a varieties of matrix inequalities and matrix calculus. Moreover, we employ the sharp concentration inequalities—Bahadur-Ranga Rao's concentration inequality [48] and Chaganty-Sethuraman's concentration inequality [49]—in strong large deviation theory to establish our finite blocklength bounds. We collect the mathematical tools of matrix analysis and large deviation theory in Chapter 2.

1.2 Quantum Information Processing Protocols

In the following, we introduce two quantum information processing protocols studied in this thesis—(1) information storage with a quantum helper, and (2) information transmission over a quantum channel. The interested readers can refer to the books [5, 50] for more detailed and various quantum information processing protocols.

[2] We denote $f_n \sim g_n$ if and only if $\lim_{n\to+\infty} \frac{f_n}{g_n} = 1$.

doi:10.6342/NTU201800597

Error Regimes	Concentration Phenomena	Hypothesis Testing	Source \ Channel Coding
Large Error	CLT: $\Pr\left(S_n \geq \sqrt{n}x\right) \to 1-\Phi\left(\frac{x}{\sqrt{v}}\right)$	$\widehat{\alpha}_{\exp\left\{-n\left[D-\frac{A}{\sqrt{n}}\right]\right\}} \to \Phi\left(\frac{A}{\sqrt{V}}\right)$	$\varepsilon^*\left(n, H+\frac{A}{\sqrt{n}}\right) \to \Phi\left(\frac{A}{\sqrt{V}}\right)$ $\varepsilon^*\left(n, C-\frac{A}{\sqrt{n}}\right) \to \Phi\left(\frac{A}{\sqrt{V}}\right)$
Medium Error	MDP: $\Pr\left(S_n \geq na_n x\right) = \mathrm{e}^{-\frac{na_n^2}{2v}x+o(na_n^2)}$	$\widehat{\alpha}_{\exp\{-n[D-a_n]\}} = \mathrm{e}^{-\frac{na_n^2}{2V}+o(na_n^2)}$	$\varepsilon^*(n, H+a_n) = \mathrm{e}^{-\frac{na_n^2}{2V}+o(na_n^2)}$ $\varepsilon^*(n, C-a_n) = \mathrm{e}^{-\frac{na_n^2}{2V}+o(na_n^2)}$
Small Error	LDP: $\Pr\left(S_n \geq nx\right) = \mathrm{e}^{-n\Lambda^*(x)+o(n)}$	$\widehat{\alpha}_{\exp\{-nr\}} = \mathrm{e}^{-n\phi(r)+o(n)}$	$\varepsilon^*(n, R) = \mathrm{e}^{-nE(R)+o(n)}$

Table 1.1: This table compares the asymptotic error behaviors of quantum hypothesis testing and classical-quantum channel coding in three error probability regimes: (i) large error (central limit theorem), (ii) medium error (moderate deviation principle), and (iii) small error (large deviation principle). The quantity S_n denotes the sum of n independent and identically distributed random variables with zero mean and variance v. The exponent Λ^* is the Legendre-Fenchel transform of the normalized cumulant generating function of S_n [27]. The error $\widehat{\alpha}_{\exp\{-nr\}}$ is defined as the minimum type-I error with the type-II error smaller than $\exp\{-nr\}$. The quantities D and V in the hypothesis testing column denote the quantum relative entropy and the relative entropy variance, respectively. The optimal error probability with blocklength n and rate R is denoted by $\epsilon^*(n, R)$. The quantities C and V in the channel coding column indicate the channel capacity and the channel dispersion, respectively. The sequence $(a_n)_{n\in\mathbb{N}}$ satisfies Eq. (1.7). The quantity $E(R)$ is the reliability function of the classical-quantum channel [35], and has not been fully characterized yet.

1.2.1 Information Storage with a Quantum Helper (Source Coding)

We consider a source of classical information which is produced with some quantum side information. That is, for some finite alphabet $\mathcal{X}$, with some probability $p(x)$, the source produces the classical information $x \in \mathcal{X}$, along with a quantum state ρ_B^x on a finite-dimensional Hilbert space $\mathcal{H}_B$. Such a source is characterized by a classical-quantum (c-q) state

$$\rho_{XB} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x| \otimes \rho_B^x. \tag{1.10}$$

where $\{|x\rangle\}_{x\in\mathcal{X}}$ is an orthonormal basis of a Hilbert space $\mathcal{H}_X$ of dimension $|\mathcal{X}|$. We note that the quantum state ρ_B on a finite Hilbert space $\mathcal{H}_B$ can be characterized by a *density operator* (or *density matrix*) such that ρ_B is positive semidefinite $\rho_B \geq 0$ and has unit trace $\mathrm{Tr}[\rho_B]$ [5, 51, 50].

The task is to compress the classical information produced by the source to a smaller index set $\mathcal{I}$ and to later decompress the information with the assistance of the quantum side information as a helper. For convenience, we also term this task *Slepian-Wolf coding with quantum side information*.

A deterministic encoder is map $\mathcal{E} : \mathcal{X} \to \mathcal{I}$ where the alphabet $\mathcal{I}$ has size $|\mathcal{I}|$. A decoder, denoted by $\mathcal{D}$, receives the compressed symbol $\mathcal{E}(x)$ along with the quantum state ρ_B^x, and produces $\hat{x} \in \mathcal{X}$, aiming to achieve $\hat{x} = x$.

Thus, the decoding is a map

$$\mathcal{I} \times \mathcal{S}(B) \ni (w, \rho_B) \to \mathcal{D}(w, \rho_B) \in \mathcal{X}. \tag{1.11}$$

If we fix the first argument as $w \in \mathcal{W}$, we have that the decoder $\mathcal{D}(w, \cdot)$ is a map from $\mathcal{S}(B) \to \mathcal{X}$,

doi:10.6342/NTU201800597

i.e. is a positive operator-valued measurement (POVM), and we denote by $\mathcal{S}(B)$ the set of quantum states on Hilbert space $\mathcal{H}_B$. Thus, we can represent the decoding by a collection of POVMs $\{\mathcal{P}_w\}_{w\in\mathcal{W}}$, where $\mathcal{P}_w = \{\Pi_{\hat{x}}^{(w)}\}_{\hat{x}\in\mathcal{X}}$ with $\Pi_{\hat{x}}^{(w)} \geq 0$ and $\sum_{\hat{x}\in\mathcal{X}} \Pi_{\hat{x}}^{(w)} = \mathbb{1}$, for each $w \in \mathcal{I}$. That is, if the message x is sent, the decoder receives $\mathcal{E}(x)$, and measures the state ρ_B^x with the POVM $\{\Pi_{\hat{x}}^{(\mathcal{E}(x))}\}_{\hat{x}\in\mathcal{X}}$.

A random encoding F from $\mathcal{X}$ to $\mathcal{W}$ is one which maps x to w with some probability $p(w|x)$. We can see the random encoding therefore as applying the deterministic encoding

$$(x_1, \ldots, x_{|\mathcal{X}|}) \mapsto (i_1, \ldots, i_{|\mathcal{X}|}) \tag{1.12}$$

with probability $p(i_1|x_1)p(i_2|x_2)\cdots p(i_{|\mathcal{X}|}|x_{|\mathcal{X}|})$. Let us write $\mathcal{F} =: \{\mathcal{E}_j : j = 1, \ldots, |\mathcal{F}|\}$ for the collection of deterministic encoders. Then a random encoding $\mathcal{F}$ applies $\mathcal{E}_j$ with some probability P_j.

An $(1, R)$-*Slepian-Wolf code* for the c-q state ρ_{XB} is an ordered pair $\mathcal{C} = (\mathcal{F}, \mathcal{D})$ consisting of a (possibly random) encoder $\mathcal{F}$ and decoder $\mathcal{D}$, such that the alphabet $\mathcal{I}$ has size $R = \log|\mathcal{W}|$. R is called the *compression rate* of the code $\mathcal{C}$.Using the above notation, the *probability of success of* $\mathcal{C}$ is given by

$$\mathrm{P_s}(\mathcal{C}) = \sum_{x\in\mathcal{X}} p(x) \sum_{j=1}^{|\mathcal{F}|} P_j \operatorname{Tr}[\rho_B^x \Pi_x^{(\mathcal{E}_j(x))}] \tag{1.13}$$

where $\mathcal{C} := (\mathcal{F}, \mathcal{D})$ for the possibly random encoding $\mathcal{F}$ which gives the deterministic encoding $\mathcal{E}_j$ with probability P_j, and decoding $\mathcal{D}$ which is defined via the collection of POVMs $\{\mathcal{P}_w\}_{w\in\mathcal{W}}$, and $\mathcal{P}_w = \{\Pi_x^{(w)}\}_{x\in\mathcal{X}}$. We may likewise define the probability of error of the code $\mathcal{C}$ by

$$\mathrm{P_e}(\mathcal{C}) := 1 - \mathrm{P_s}(\mathcal{C}) \tag{1.14}$$

In the following, we define the *optimal one-shot compression rate*:

$$R^*(1,\varepsilon) = \inf\left\{R : \text{for some } R' \leq R,\ \exists\,(1,R')\text{-Slepian Wolf code } \mathcal{C} \text{ for } \rho_{XB} \text{ s.t. } \mathrm{P_e}(\mathcal{C}) \leq \varepsilon\right\} \tag{1.15}$$

Similarly, the *optimal one-shot probability of error* for ρ_{XB} is defined as:

$$\varepsilon^*(1,R) := \inf\left\{\mathrm{P_e}(\mathcal{C}) : \mathcal{C} \text{ is an } (1,R')\text{-Slepian Wolf code for } \rho_{XB} \text{ for some } R' \leq R\right\}. \tag{1.16}$$

The Slepian-Wolf coding can be easily applied to the n-shot case when the underlying c-q state $\rho_{XB} \in \mathcal{S}(XB)$ has an independent an identically distributed extension $\rho_{X^nB^n} = \rho_{XB}^{\otimes n}$. In this case, an (n, R)-Slepian Wolf code for the state ρ_{XB} is defined as a $(1, nR)$-Slepian Wolf code for the state $\rho_{XB}^{\otimes n}$. We define the *optimal n-shot probability of error* for ρ_{XB} as

$$\varepsilon^*(n,R) := \inf\left\{\mathrm{P_e}(\mathcal{C}) : \mathcal{C} \text{ is an } (n,R')\text{-Slepian Wolf code for } \rho_{XB} \text{ for some } R' \leq R\right\}, \tag{1.17}$$

and likewise the *optimal n-shot compression rate* for ρ_{XB} as

$$R^*(n,\varepsilon) = \inf\left\{R : \text{for some } R' \leq R,\ \exists\,(n,R')\text{-Slepian Wolf code } \mathcal{C} \text{ for } \rho_{XB} \text{ s.t. } \mathrm{P_e}(\mathcal{C}) \leq \varepsilon\right\}. \tag{1.18}$$

We illustrate the protocol of Slepian-Wolf coding with QSI Figure 1.1 below.

doi:10.6342/NTU201800597

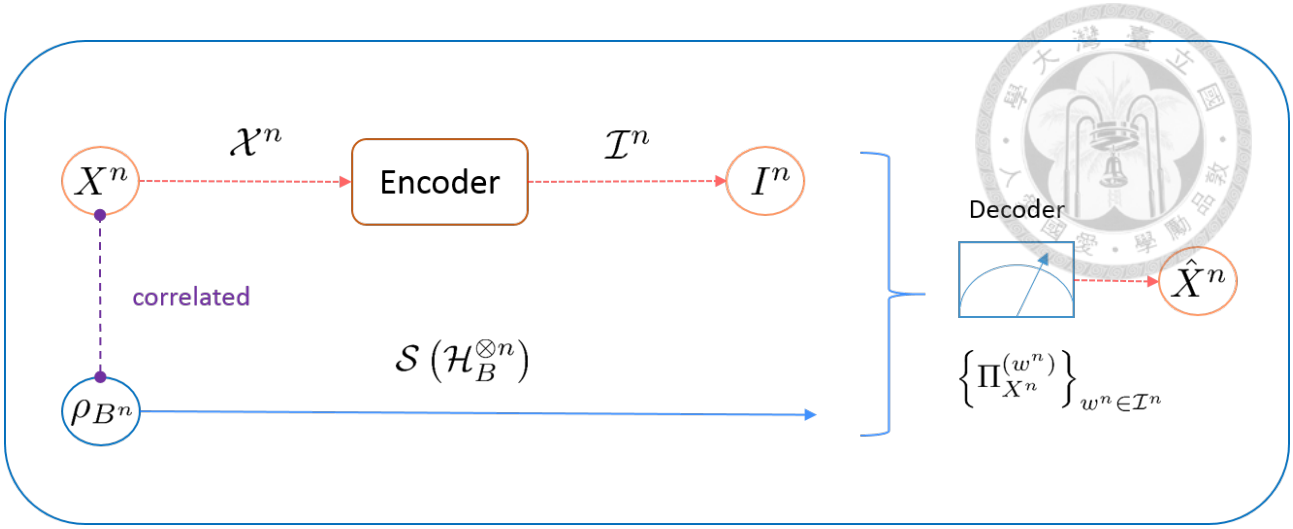


Figure 1.1: We are given n copies of a classical source X which is correlated with a quantum system B. We compress the source into an index set $\mathcal{I}_n$ via the encoder $\mathcal{E}_n$, and then perform a decoding via $\mathcal{D}_n$ which has access to the side information ρ_{B^n}. This yields the output $\hat{X}_n$ with associated alphabets $\mathcal{X}^n$. The decoder $\mathcal{D}_n$ here is a family of positive operator-valued measurement (POVM) $\left\{\Pi_{X^n}^{(w^n)}\right\}_{w^n\in\mathcal{I}^n}$. The red-dotted lines indicate classical information, while the blue-solid lines stand for quantum information.

1.2.2 Information Transmission over a Quantum Channel (Channel Coding)

Let $\mathcal{M}$ be a finite alphabetical set with size $M = |\mathcal{M}|$. An (n-blocklength) *encoder* is a map $\mathcal{F}_n : \mathcal{M} \to \mathcal{X}^n$ that encodes each message $m \in \mathcal{M}$ to a codeword $\mathbf{x}^n(m) := x_1(m)x_2(m)\dots x_n(m) \in \mathcal{X}^n$. Here, we assume that the input alphabet $\mathcal{X}$ is finite. The codeword $\mathbf{x}^n(m)$ is then mapped to a state $\rho^n_{\mathbf{x}^n(m)}$ in the n-fold of Hilbert space $\mathcal{H}$. The *decoder* is described by a POVM $\Pi_n = \{\Pi_{n,1}, \dots, \Pi_{n,M}\}$ on $\mathcal{H}^{\otimes n}$, where $\Pi_{n,i} \geq 0$ and $\sum_{i=1}^M \Pi_{n,i} = \mathbb{1}$. Throughout this thesis, we assume that the channel output state $\rho^n_{\mathbf{x}^n(m)}$ has a tensor product structure. That is, $\rho^n_{\mathbf{x}^n(m)}$ can be presented as

$$W^{\otimes n}_{\mathbf{x}^n(m)} = W_{x_1(m)} \otimes W_{x_2(m)} \otimes \cdots \otimes W_{x_n(m)} \in \mathcal{S}(\mathcal{H}^{\otimes n}). \tag{1.19}$$

Then, this protocol is equivalent to a c-q channel coding with a c-q channel $\mathcal{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$. We leave the scenarios of classical message communications over general quantum channels as future work; see also the open problems in Chapter 13.

The pair $(\mathcal{F}_n, \Pi_n) =: \mathcal{C}_n$ is called a *code* with *coding rate* (or called *transmission rate*) $R = \frac{1}{n}\log|\mathcal{C}_n| = \frac{1}{n}\log M$. The error probability of sending a message m with the code $\mathcal{C}_n$ is given by the Born rule $\varepsilon_m(\mathcal{C}_n) := 1 - \mathrm{Tr}\left(\Pi_{n,m} W^{\otimes n}_{\mathbf{x}^n(m)}\right)$. We use $\varepsilon_{\max}(\mathcal{C}_n) = \max_{m\in\mathcal{M}} \varepsilon_m(\mathcal{C}_n)$ and $\bar{\varepsilon}(\mathcal{C}_n) = \frac{1}{M}\sum_{m\in\mathcal{M}} \varepsilon_m(\mathcal{C}_n)$ to denote the *maximal* error probability and the *average* error probability, respectively. Given a sequence $\mathbf{x}^n \in \mathcal{X}^n$, we denote by

$$P_{\mathbf{x}^n}(x) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{x = x_i\} \tag{1.20}$$

the empirical distribution of $\mathbf{x}^n$, where x_i is the i-th position of $\mathbf{x}^n$. A constant composition code with a composition $P_{\mathbf{x}^n}$ refers to a codebook whose codewords all have the same distribution $P_{\mathbf{x}^n}$.

Denote by $\varepsilon^*(n, R)$ the smallest average probability of error among all the coding strategies with a

doi:10.6342/NTU201800597

blocklengh n and coding rate R. Our goal in this thesis is then to characterize $\varepsilon^*(n, R)$ as a function of (n, R); see Part II. Figure 1.2 below depicts the protocol of c-q channel coding.

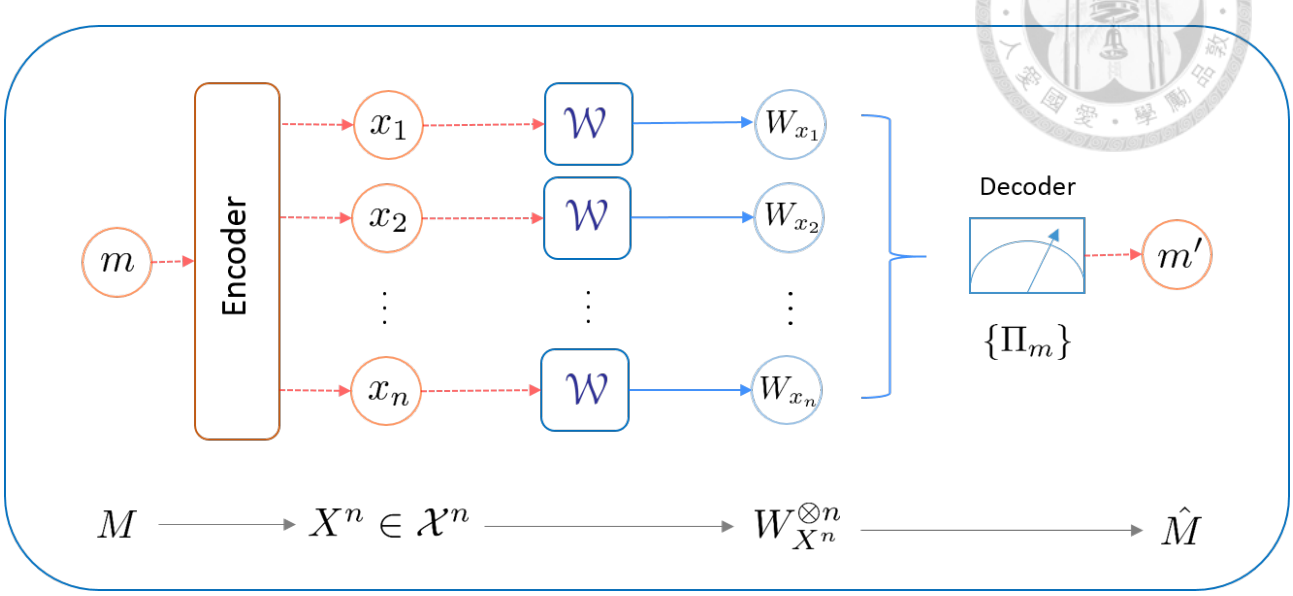


Figure 1.2: We encode the (classical) message m to an n-blocklength sequence x^n. Then, input sequence will be mapped to an n-product channel output state $W_{\mathbf{x}^n}^{\otimes n}$. Lastly, the decoder, a positive operator-valued measurement (POVM), measures the channel output state to obtain the estimated message $\hat{m}$. The red-dotted lines indicate classical information, while the blue-solid lines stand for quantum information.

doi:10.6342/NTU201800597

1.3 Main Contributions

Although the aim of this thesis is to give an exposition to the current development of the error exponent analysis in quantum information theory, we list our contributions in the following. The results can also be found in the papers [36, 26, 52, 53].

(I) We prove major properties of error exponent functions and auxiliary functions for both Slepian-Wolf coding with QSI (Chapter 5) and classical-quantum channel coding (Chapter 9). Specifically,

(a) We show that the error exponent functions introduced by Blahut [32, 23], Haroutunian [31, 24], and Csiszár-Körner [54, 55, 25] have variational representations (Theorems 5.1 and 9.1). Theses representations are equivalent to the Gallager's expressions [29, 21, 56] in the classical case. However, in the quantum case, they are expressed by the log-Euclidean Rényi divergence [57, 58], while Gallager's expressions correspond to Petz's Rényi divergence [59]. As a consequence of the Golden-Thompson inequality [60, 61], the variational representations are weaker than Gallager's expressions in the optimality part, i.e. the converse (see Theorem 9.1). Nevertheless, they have applications in the strong converse domain[3] [58, 53] and the moderate deviation analysis (see Section 12.2)

(b) Since the error exponent functions are the *Lengendre-Fenchel transform* of the auxiliary functions , the properties of the auxiliary functions immediately characterize that of the error exponent functions. We prove the concavity properties, which solves an open problem addressed by Holevo [35], and the first-order/second-order derivatives (Propositions 5.1, 5.2, 9.1, 9.2, and 9.3).

(c) We prove the continuity and the saddle-point property of the error exponent functions, which is one of the crucial steps of establishing finite blocklength results (Propositions 5.3 and 9.5).

(d) An asymptotic expansion of error exponent functions when the compression rate (resp. transmission rate) approaches the Slepian-Wolf limit (resp. channel capacity) is shown in Propositions 8.1 and 12.2. This property results in the moderate deviation analysis (see Chapters 8 and 12).

(II) We establish a finite blocklength achievability bound of the Slepian-Wolf coding with QSI (Theorem 6.1), which has the following applications:

(a) recovering Devetak and Winter's asymptotic achievability result, i.e. any compression rate larger than the Slepian-Wolf limit is achievable;

(b) achievability of the moderate deviation analysis (Theorem 8.1);

(c) proof ingredient in the achievability of strong converse domain [53].

(III) For the optimality part, we establish a series of following results:

[3] The strong converse domain means that the compression rate (resp. transmission rate) is smaller (resp. larger) than the Slepian-Wolf limit (resp. channel capacity). In this case, the optimal probability of success exponentially decays [37, 62, 63, 57, 58, 64, 53]. This thesis does not include contents of the strong converse part.

doi:10.6342/NTU201800597

(a) a sharp converse Hoeffding bound (see Theorem 4.4 and Corollary 4.1) for binary quantum hypothesis testing, which is the main ingredient for the finite blocklength error exponent analysis and moderate deviation analysis in quantum information theory;

(b) By proving an one-shot converse bound to relate the source coding problem to hypothesis testing (Proposition 7.1), we employ the above sharp converse Hoeffding bound to show the finite blocklength sphere-packing bound of Slepian-Wolf coding with QSI (Theorem 7.1).

(c) With an one-shot converse bound reducing the channel coding problem to hypothesis testing (Proposition 11.3), we prove the finite blocklength sphere-packing bound of classical-quantum channel coding. Under the assumption of using constant composition codes, i.e. the composition for each codeword in the codebook is the same, we derive the exact prefactor (see Theorem 11.1). For general codes, the obtained prefactor is significantly improved from the previous result of subexponential [38] to polynomial (see Corollary 11.1). We remark that the exact prefactor for general codes remains open even in the classical case.

(IV) For the moderate deviation regime, we discuss the trade-offs between the rate, optimal probability of error, and the blocklength.

(a) When the exponential decaying rate of the type-II error in quantum hypothesis testing approaches the relative entropy from below with the speed not faster than $O(1/\sqrt{n})$, we show that the optimal type-I error vanishes asymptotically (Theorems 4.5 and 4.6):

$$\lim_{n\to+\infty} \frac{1}{na_n^2} \log \widehat{\alpha}_{\exp\{-n[D(\rho\|\sigma)-a_n]\}} \left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) = -\frac{1}{2V(\rho\|\sigma)}, \tag{1.21}$$

where $\widehat{\alpha}_\mu$ denotes the smallest type-I error when the type-II error does not exceed μ; $D(\rho\|\sigma)$ and $V(\rho\|\sigma)$ denote the relative entropy and relative variance of ρ and σ, respectively.

(b) When the compression rate approaches the Slepian-Wolf limit from above with the speed not faster than $O(1/\sqrt{n})$, we show that the optimal probability vanishes asymptotically (Theorem 8.1):

$$\lim_{n\to+\infty} \frac{\log \epsilon^*(n, H(X|B)_\rho + a_n)}{na_n^2} = -\frac{1}{2V}. \tag{1.22}$$

(c) When the transmission rate approaches the channel capacity from below with the speed not faster than $O(1/\sqrt{n})$, we show that the optimal probability vanishes asymptotically (Theorems 12.1 and 12.2):

$$\lim_{n\to+\infty} \frac{\log \epsilon^*(n, C - a_n)}{na_n^2} = -\frac{1}{2V}. \tag{1.23}$$

1.4 Structure of the Thesis

Organization.

The thesis is divided into three parts. *Part I: Fundamentals* collects the necessary mathematical tools—matrix analysis and large deviation theory (Chapter 2), the notation of all quantum entropic quantities and their properties (Chapter 3), and the error exponent analysis for quantum hypothesis

doi:10.6342/NTU201800597

testing (Chapter 4). The two quantum information tasks investigated in this thesis are presented in Parts II and III, respectively.

Part II: Information Storage with a Quantum Helper discusses the source coding scenario—the error exponent analysis for Slepian-Wolf coding with QSI. We introduce the error exponent functions in Chapter 5 and prove their properties. The achievability and optimality are studied in Chapters 6 and 7. Next, we move on to the moderate deviation regime in Chapter 8, which heavily relies on the established results in achievability and optimality.

Part III: Information Transmission over a Quantum Channel investigates the channel coding scheme—the error exponent analysis for communications over classical-quantum channels. The organization is similar to Part II: the error exponent function, achievability, optimality, and the moderate deviation analysis are presented in Chapters 9, 10, 11, and 12, respectively. Lastly, we conclude this thesis in Chapter 13 and provide open problems for future study.

Structure.
The structure of the thesis is depicted in Figure 1.3. The matrix mathematics provided in Chapter 2.1 will be useful in proving properties of the quantum entropic quantities in Chapter 3, properties of error exponent functions in Chapters 5 and 9, and the achievability in Chapters 4, 6 and 10. The techniques of large deviation theory in chapter 2.2 will be applied in the optimality part in Chapters 4, 7, and 11. The optimality in Chapters 7 and 11 requires the sharp converse bound of quantum hypothesis testing in Chapter 4. In either Part II or Part III, the moderate deviation analysis (Chapters 8 and 12) relies on the properties of error exponent functions (Chapters 5 and 9), achievability (Chapters 6 and 10), and optimality (Chapters 7 and 11).

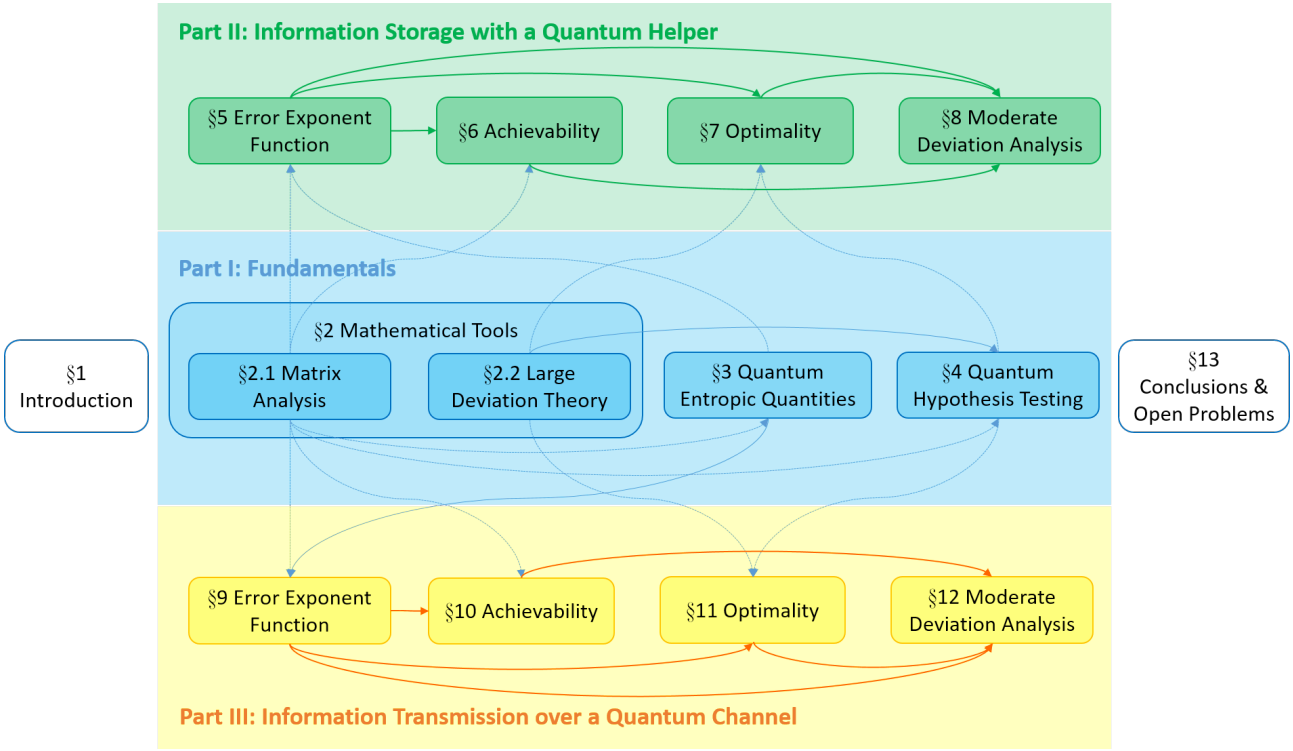


Figure 1.3: Structure of the thesis.

doi:10.6342/NTU201800597

Part I

Fundamentals

doi:10.6342/NTU201800597

Chapter 2

Mathematical Tools

We provide preliminaries mathematical tools in this Chapter. The introductory matrix analysis is given in Section 2.1. In Section 2.2, we present the backgrounds of large deviation theory.

2.1 Matrix Analysis

In this section, we provide backgrounds of matrix analysis. For a general treatment of this topic, interested readers can refer to [65, Section 2.1], [66, Chapter 17], [67, Section X.4], [68, Section 5.3], and [69, Chapter 3].

We denote by $\mathbb{M}^{\mathrm{sa}}$ the set of self-adjoint operators, and by $\mathbb{M}_d^{\mathrm{sa}}(I)$ the set of Hermitian $d \times d$ matrices with eigenvalues contained in I. Similarly, let $\mathbb{M}_d^{+}$ and $\mathbb{M}_d^{++}$ be the set of $d \times d$ positive semi-definite matrices and positive definite matrices, respectively.

Let $\mathcal{U}, \mathcal{W}$ be real Banach spaces. The *Fréchet derivative* of a function $\boldsymbol{f} : \mathcal{U} \to \mathcal{W}$ at a point $\boldsymbol{X} \in \mathcal{U}$, if it exists[1], is a unique linear mapping $\mathsf{D}\boldsymbol{f}[\boldsymbol{X}] : \mathcal{U} \to \mathcal{W}$ such that

$$\frac{\|\boldsymbol{f}(\boldsymbol{X}+\boldsymbol{E}) - \boldsymbol{f}(\boldsymbol{X}) - \mathsf{D}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E})\|_{\mathcal{W}}}{\|\boldsymbol{E}\|_{\mathcal{U}}} \to 0 \quad \text{as } \boldsymbol{E} \in \mathcal{U} \text{ and } \|\boldsymbol{E}\|_{\mathcal{U}} \to \boldsymbol{0},$$

or, equivalently,

$$\|\boldsymbol{f}(\boldsymbol{X}+\boldsymbol{E}) - \boldsymbol{f}(\boldsymbol{X}) - \mathsf{D}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E})\|_{\mathcal{W}} = o(\|\boldsymbol{E}\|_{\mathcal{U}}),$$

where $\|\cdot\|_{\mathcal{U}(\mathcal{W})}$ is a norm in $\mathcal{U}$ (resp. $\mathcal{W}$). The notation $\mathsf{D}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E})$ then is interpreted as "the Fréchet derivative of $\boldsymbol{f}$ at $\boldsymbol{X}$ in the direction $\boldsymbol{E}$". Furthermore, the Fréchet derivative implies the Gâteaux derivative such that the differentiation of $\boldsymbol{f}(\boldsymbol{X}+t\boldsymbol{E})$ with respect to the real variable t is

$$\frac{\boldsymbol{f}(\boldsymbol{X}+t\boldsymbol{E}) - \boldsymbol{f}(\boldsymbol{X})}{t} \to \mathsf{D}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E}) \quad \text{as } t \to 0.$$

For example, if the operator-valued function is the inversion $\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{X}^{-1}$ for each invertible matrix $\boldsymbol{X}$, then (see e.g. [67, Example X.4.2])

$$\mathsf{D}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{Y}) = -\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{X}^{-1}. \tag{2.1}$$

[1] We assume the functions considered in the paper are Fréchet differentiable. The readers can refer to, e.g. [70, 71], for conditions for when a function is Fréchet differentiable.

doi:10.6342/NTU201800597

The Fréchet derivative also satisfies several properties similar to conventional derivatives of real-valued functions (see e.g. [69, Theorem 3.4]):

Proposition 2.1 (Properties of Fréchet Derivatives). *Let $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ be real Banach spaces.*

1. *(Sum Rule) If $\boldsymbol{f}_1 : \mathcal{U} \to \mathcal{W}$ and $\boldsymbol{f}_2 : \mathcal{U} \to \mathcal{W}$ are Fréchet differentiable at $\boldsymbol{A} \in \mathcal{U}$, then so is $\boldsymbol{f} = \alpha\boldsymbol{f}_1 + \beta\boldsymbol{f}_2$ and $\mathsf{D}\boldsymbol{f}[\boldsymbol{A}](\boldsymbol{E}) = \alpha \cdot \mathsf{D}\boldsymbol{f}_1[\boldsymbol{A}](\boldsymbol{E}) + \beta \cdot \mathsf{D}\boldsymbol{f}_2[\boldsymbol{A}](\boldsymbol{E})$.*
2. *(Product Rule) If $\boldsymbol{f}_1 : \mathcal{U} \to \mathcal{W}$ and $\boldsymbol{f}_2 : \mathcal{U} \to \mathcal{W}$ are Fréchet differentiable at $\boldsymbol{A} \in \mathcal{U}$ and assume the multiplication is well-defined in $\mathcal{W}$, then so is $\boldsymbol{f} = \boldsymbol{f}_1 \cdot \boldsymbol{f}_2$ and $\mathsf{D}\boldsymbol{f}[\boldsymbol{A}](\boldsymbol{E}) = \mathsf{D}\boldsymbol{f}_1[\boldsymbol{A}](\boldsymbol{E}) \cdot \boldsymbol{f}_2(\boldsymbol{A}) + \boldsymbol{f}_1(\boldsymbol{A}) \cdot \mathsf{D}\boldsymbol{f}_2[\boldsymbol{A}](\boldsymbol{E})$.*
3. *(Chain Rule) Let $\boldsymbol{f}_1 : \mathcal{U} \to \mathcal{V}$ and $\boldsymbol{f}_2 : \mathcal{V} \to \mathcal{W}$ be Fréchet differentiable at $\boldsymbol{A} \in \mathcal{U}$ and $\boldsymbol{f}_1(\boldsymbol{A})$ respectively, and let $\boldsymbol{f} = \boldsymbol{f}_2 \circ \boldsymbol{f}_1$ (i.e. $\boldsymbol{f}(\boldsymbol{A}) = \boldsymbol{f}_2(\boldsymbol{f}_1(\boldsymbol{A}))$. Then $\boldsymbol{f}$ is Fréchet differentiable at $\boldsymbol{A}$ and $\mathsf{D}\boldsymbol{f}[\boldsymbol{A}](\boldsymbol{E}) = \mathsf{D}\boldsymbol{f}_2[\boldsymbol{f}_1(\boldsymbol{A})]\left(\mathsf{D}\boldsymbol{f}_1[\boldsymbol{A}](\boldsymbol{E})\right)$.*

Similarly, the m-th Fréchet derivative $\mathsf{D}^m\boldsymbol{f}[\boldsymbol{X}]$ is a unique multi-linear map from $\mathcal{U}^m \triangleq \mathcal{U} \times \cdots \times \mathcal{U}$ (m times) to $\mathcal{W}$ that satisfies

$$\begin{aligned}\|\mathsf{D}^{m-1}\boldsymbol{f}[\boldsymbol{X}+\boldsymbol{E}_m](\boldsymbol{E}_1,\ldots,\boldsymbol{E}_{m-1}) - \mathsf{D}^{m-1}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E}_1,\ldots,\boldsymbol{E}_{m-1}) \\ - \mathsf{D}^m\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E}_1,\ldots,\boldsymbol{E}_m)\|_{\mathcal{W}} = o(\|\boldsymbol{E}_m\|_{\mathcal{U}})\end{aligned}$$

for each $\boldsymbol{E}_i \in \mathcal{U}, i = 1,\ldots,m$. If $\mathsf{D}^m\boldsymbol{f}[\boldsymbol{X}]$ is continuous at $\boldsymbol{X}$, then the m-th Fréchet derivative can be expressed as a mixed partial derivative [72, Section 9] (see also [73, Theorem 2.3.1]).

$$\mathsf{D}^m\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E}_1,\ldots,\boldsymbol{E}_m) = \frac{\partial}{\partial s_1}\cdots\frac{\partial}{\partial s_m}\bigg|_{s_1=\ldots=s_m=0} \boldsymbol{f}\left(\boldsymbol{X} + s_1\boldsymbol{E}_1 + \cdots + s_m\boldsymbol{E}_m\right).$$

We can observe, from the above equation, that the m-th Fréchet derivative is symmetric about all $\boldsymbol{E}_i$; see [74, Theorem 8], [67, p. 313], and [75, Theorem 4.3.4]. We refer to Refs. [76, Section 8.12], [66, Chapter 17], [75, Section 4.3], and [77] for further information about higher order Fréchet derivatives.

The following proposition relates the second order Fréchet derivative with the convexity of a matrix-valued function, i.e. $\boldsymbol{f}(t\boldsymbol{A}) + \boldsymbol{f}((1-t)\boldsymbol{B}) \preceq \boldsymbol{f}(t\boldsymbol{A} + (1-t)\boldsymbol{B})$ for all $0 \le t \le 1$.

Proposition 2.2 (Convexity of twice Fréchet differentiable matrix functions [78, Proposition 2.2]). *Let U be an open convex subset of a real Banach space $\mathcal{U}$, and $\mathcal{W}$ is also a real Banach space. Then a twice Fréchet differentiable function $\boldsymbol{f} : U \to \mathcal{W}$ is convex if and only if $\mathsf{D}^2\boldsymbol{f}(\boldsymbol{X})(\boldsymbol{h},\boldsymbol{h}) \succeq \boldsymbol{0}$ for each $\boldsymbol{X} \in U$ and $\boldsymbol{h} \in \mathcal{U}$.*

The *partial Fréchet derivative* of multivariate functions can be defined as follows [68, Section 5.3]. Let $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ be real Banach spaces, $\boldsymbol{f} : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$. For a fixed $\boldsymbol{v}_0 \in \mathcal{V}$, $\boldsymbol{f}(\boldsymbol{u}, \boldsymbol{v}_0)$ is a function of $\boldsymbol{u}$ whose derivative at $\boldsymbol{u}_0$, if it exists, is called the partial Fréchet derivative of $\boldsymbol{f}$ with respect to $\boldsymbol{u}$, and is denoted by $\mathsf{D}_{\boldsymbol{u}}\boldsymbol{f}[\boldsymbol{u}_0, \boldsymbol{v}_0]$. The partial Fréchet derivative $\mathsf{D}_{\boldsymbol{v}}\boldsymbol{f}[\boldsymbol{u}_0, \boldsymbol{v}_0]$ is defined similarly.

The Fréchet derivative and the partial Fréchet derivative can be related as follows.

Proposition 2.3 (Partial Fréchet derivative [68, Proposition 5.3.15]). *If $\boldsymbol{f} : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$ is Fréchet differentiable at $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{U} \times \mathcal{V}$, then the partial Fréchet derivatives $\mathsf{D}_{\boldsymbol{X}}\boldsymbol{f}[\boldsymbol{X}, \boldsymbol{Y}]$ and $\mathsf{D}_{\boldsymbol{Y}}\boldsymbol{f}[\boldsymbol{X}, \boldsymbol{Y}]$ exist, and*

$$\mathsf{D}\boldsymbol{f}[\boldsymbol{X}, \boldsymbol{Y}](\boldsymbol{h}, \boldsymbol{k}) = \mathsf{D}_{\boldsymbol{X}}\boldsymbol{f}[\boldsymbol{X}, \boldsymbol{Y}](\boldsymbol{h}) + \mathsf{D}_{\boldsymbol{Y}}\boldsymbol{f}[\boldsymbol{X}, \boldsymbol{Y}](\boldsymbol{k}).$$

doi:10.6342/NTU201800597

Now let $\boldsymbol{f}:\mathcal{U}^n\to\mathcal{W}$ and assume it is a holomorphic function (i.e. Fréchet differential in a neighborhood of every point in its domain), then the *Taylor expansion* $\boldsymbol{f}\left(\underline{\boldsymbol{X}}+\underline{\boldsymbol{E}}\right)$ at $\underline{\boldsymbol{X}}\triangleq(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$, $\underline{\boldsymbol{E}}\triangleq(\boldsymbol{E}_1,\ldots,\boldsymbol{E}_n)\in\mathcal{U}^n$ can be expressed as

$$\begin{aligned}\boldsymbol{f}\left(\underline{\boldsymbol{X}}+\underline{\boldsymbol{E}}\right)&=\boldsymbol{f}\left(\underline{\boldsymbol{X}}\right)+\sum_{k=1}^{\infty}\frac{1}{k!}\mathsf{D}^k\boldsymbol{f}\left[\underline{\boldsymbol{X}}\right]\left(\underbrace{\boldsymbol{E},\ldots,\boldsymbol{E}}_{k}\right)\\&=\boldsymbol{f}\left(\underline{\boldsymbol{X}}\right)+\sum_{j=1}^{n}\mathsf{D}_{\boldsymbol{X}_j}\boldsymbol{f}\left[\underline{\boldsymbol{X}}\right]\left(\boldsymbol{E}_j\right)+\frac{1}{2!}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathsf{D}^2_{\boldsymbol{X}_j\boldsymbol{X}_k}(\boldsymbol{E}_j,\boldsymbol{E}_k)+\text{Remaining terms.}\end{aligned}\tag{2.2}$$

For any map $\boldsymbol{f}:\mathcal{U}\to\mathcal{W}$ and an operator $\boldsymbol{X}\in\mathcal{U}$, we define the induced norm of the Fréchet derivative $\mathsf{D}\boldsymbol{f}[\boldsymbol{X}]$ as

$$\|\mathsf{D}\boldsymbol{f}[\boldsymbol{X}]\|\triangleq\sup_{\boldsymbol{E}\neq\boldsymbol{0}}\frac{\|\mathsf{D}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E})\|}{\|\boldsymbol{E}\|},\tag{2.3}$$

where the norm can be any consistent norm (e.g. $\|\mathsf{D}\boldsymbol{f}[\boldsymbol{X}]\|_2=\sup_{\boldsymbol{E}\neq\boldsymbol{0}}\|\mathsf{D}\boldsymbol{f}[\boldsymbol{X}](\boldsymbol{E})\|_2/\|\boldsymbol{E}\|_2$).

The norm of the Fréchet derivative is closely related to the condition numbers, which measure the sensitivity of an operator-valued function to perturbations in the variables. Hence, the *absolute condition number* is defined by

$$\mathrm{cond}_{\mathrm{abs}}(\boldsymbol{f},\boldsymbol{X})\triangleq\lim_{\varepsilon\to0}\sup_{\|\boldsymbol{E}\|\le\varepsilon}\frac{\|\boldsymbol{f}(\boldsymbol{X}+\boldsymbol{E})-\boldsymbol{X}\|}{\varepsilon}.\tag{2.4}$$

Then the norm of the Fréchet derivative can be expressed by the absolute condition number [79]

$$\mathrm{cond}_{\mathrm{abs}}(\boldsymbol{f},\boldsymbol{X})=\|\mathsf{D}\boldsymbol{f}[\boldsymbol{X}]\|.$$

We note that there are several algorithms and software packages that can compute the absolute condition number; see [69, Section 3], [80] and references therein.

Next, we introduce the *standard matrix functions*. For each self-adjoint and bounded operator $\boldsymbol{A}\in\mathbb{M}^{\mathrm{sa}}$ with the spectrum $\sigma(\boldsymbol{A})$ and the spectral measure $\boldsymbol{E}$, the *spectral decomposition* is given as

$$\boldsymbol{A}=\int_{\lambda\in\sigma(\boldsymbol{A})}\lambda\,\mathrm{d}\boldsymbol{E}(\lambda).\tag{2.5}$$

Hence, each scalar function can be extended to a standard matrix function as follows.

Definition 2.1 (Standard Matrix Function)**.** Let $f:I\to\mathbb{R}$ be a real-valued function on an interval I of the real line. Suppose that $\boldsymbol{A}\in\mathbb{M}^{\mathrm{sa}}(I)$ has the spectral decomposition (2.5). Then

$$f(\boldsymbol{A})\triangleq\int_{\lambda\in\sigma(\boldsymbol{X})}f(\lambda)\,\mathrm{d}\boldsymbol{E}(\lambda).$$

From this equation, it is clear that $\sigma(f(\boldsymbol{A}))=f(\sigma(\boldsymbol{A}))$, which is called the *spectral mapping theorem*.

A function $f:I\to\mathbb{R}$ is called *operator convex* if for each $\boldsymbol{A},\boldsymbol{B}\in\mathbb{M}^{\mathrm{sa}}(I)$ and $0\le t\le1$,

$$f(t\boldsymbol{A})+f((1-t)\boldsymbol{B})\le f(t\boldsymbol{A}+(1-t)\boldsymbol{B}).$$

doi:10.6342/NTU201800597

Similarly, a function $f : I \to \mathbb{R}$ is called *operator monotone* if for each $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{M}^{\mathrm{sa}}(I)$,

$$\boldsymbol{A} \leq \boldsymbol{B} \;\Rightarrow\; f(\boldsymbol{A}) \leq f(\boldsymbol{B}).$$

It is remarkable that not all convex (resp. monotone) functions are operator convex (resp. monotone). For example, the exponential function is not operator convex nor operator monotone on $[0, \infty)$; the power functions that are operator convex are $f(x) = x^p$ for $p \in [-1, 0] \cup [1, 2]$ and $f(x) = -x^p$ for $p \in [0, 1]$. However, the trace function on $\mathbb{M}^{\mathrm{sa}}$ given by $\boldsymbol{A} \to \mathrm{Tr}[f(\boldsymbol{A})]$ preserves the convexity or monotonicity.

Proposition 2.4 (Convexity and Monotonicity for Trace Functions [81, Section 2.2]). *Consider a real-valued function $f : I \to \mathbb{R}$. If f is convex (resp. monotone) on $U \subseteq I$, then the function $\boldsymbol{A} \to \mathrm{Tr}[f(\boldsymbol{A})]$ is convex (resp. monotone) on $\mathbb{M}^{sa}(U)$.*

We refer the readers to Refs. [73] and [82] for general expositions to operator convex and monotone functions.

If the scalar function is continuously differentiable, then it is convenient to introduce the following two properties for the trace function of Fréchet derivatives.

Proposition 2.5 ([82, Theorem 3.23]). *Let $\boldsymbol{A}, \boldsymbol{X} \in \mathbb{M}^{sa}$ and $t \in \mathbb{R}$. Assume $f : I \to \mathbb{R}$ is a continuously differentiable function defined on interval I and assume that the eigenvalues of $\boldsymbol{A}+t\boldsymbol{X} \subset I$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Tr}\, f(\boldsymbol{A} + t\boldsymbol{X})\Big|_{t=t_0} = \mathrm{Tr}[\boldsymbol{X} f'(\boldsymbol{A} + t_0\boldsymbol{X})].$$

In the following, we collect necessary matrix inequalities that will be employed later in this thesis. Let $x := (x_1, \ldots, x_d) \in \mathbb{R}^d$ be a d-dimensional vector with positive elements. Denote by $x^{\downarrow} := (x_1^{\downarrow}, \ldots, x_d^{\downarrow})$ the *decreasing arrangement* of x, i.e. $x_1^{\downarrow} \geq \cdots \geq x_d^{\downarrow}$. We say that x is *weak majorized* by y, denoted by $x \prec_{\mathrm{w}} y$, if

$$\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow},\ 1 \leq k \leq d. \tag{2.6}$$

The *weak log-majorization* $x \prec_{\mathrm{wlog}} y$ is defined when $\log x \prec_{\mathrm{w}} \log y$, where we denote by $\log x$ the vector whose components equal to the logarithm of the components of x. It is well-known that $x \prec_{\mathrm{wlog}} y$ implies $x \prec_{\mathrm{w}} y$ [67, Example II.3.5]. Let $\lambda(X)$ denote the vector of eigenvalues of the matrix X. For two positive semi-definite matrices A and B, the weak majorization $\lambda(A) \prec_{\mathrm{w}} \lambda(B)$ is equivalent to $|||A||| \leq |||B|||$ for all unitarily-invariant norm $|||\cdot|||$ [82, Theorem 6.23].

Lemma 2.1 ([83, Theorem 2.10]). *For any $A, B \in \mathbb{M}_d^{++}$, and $0 \leq \tau \leq 1$. Then*

$$\lambda(A \#_\tau B) \prec_{\mathrm{wlog}} \lambda\left(A^{1-\tau} B^{\tau}\right). \tag{2.7}$$

Lemma 2.2 (Araki-Lieb-Thirring Inequality [84]; see also [67, Theorem IX.2.10]). *Let $A, B \in \mathbb{M}_d^{+}$. Then, we have*

$$\lambda\left(B^t A^t B^t\right) \prec_{\mathrm{w}} \lambda\left((BAB)^t\right),\ \textit{for } t \in [0, 1], \tag{2.8}$$

$$\lambda\left(B^t A^t B^t\right) \succ_{\mathrm{w}} \lambda\left((BAB)^t\right),\ \textit{for } t \geq 1. \tag{2.9}$$

doi:10.6342/NTU201800597

Lemma 2.3 ([67, Example II.3.5]). *Let $x, y \in \mathbb{R}^d_{\geq 0}$ (the set of d-dimensional vectors of non-negative real numbers). Then*

$$x \prec_{\mathrm{w}} y \quad \textit{implies} \quad x^t \prec_{\mathrm{w}} y^t \tag{2.10}$$

for all $t \geq 1$.

Lemma 2.4 (See, e.g. [81, Section 2.2]). *Let f be a monotonically increasing function on the real line. Then $A \preceq B$ implies*

$$\mathrm{Tr}\left[f(A)\right] \leq \mathrm{Tr}\left[f(B)\right]. \tag{2.11}$$

Lemma 2.5 (Matrix Hölder's Inequality [67, Corollary IV.2.6]). *Let $A, B \in \mathbb{M}_d^+$. Then*

$$\mathrm{Tr}\left[AB\right] \leq \left(\mathrm{Tr}\left[A^{\frac{1}{\theta}}\right]\right)^{\theta} \left(\mathrm{Tr}\left[B^{\frac{1}{1-\theta}}\right]\right)^{1-\theta} \tag{2.12}$$

for all $0 \leq \theta \leq 1$.

Lemma 2.6. *Let $A, B \in \mathbb{M}_d^{++}$. Then, for every $t \geq 1$ and $0 \leq \tau \leq 1$, we have*

$$\mathrm{Tr}\left[(A \#_\tau B)^t\right] \leq \mathrm{Tr}\left[A^{t(1-\tau)} B^{t\tau}\right]. \tag{2.13}$$

Proof of Lemma 2.6. From Lemma 2.1, we have

$$\lambda\left(A \#_\tau B\right) \prec_{\mathrm{w}} \lambda\left(A^{1-\tau} B^{\tau}\right) \tag{2.14}$$

$$= \lambda\left(A^{\frac{1-\tau}{2}} B^{\tau} A^{\frac{1-\tau}{2}}\right) \tag{2.15}$$

$$\prec_{\mathrm{w}} \lambda\left(\left(A^{\frac{t(1-\tau)}{2}} B^{t\tau} A^{\frac{t(1-\tau)}{2}}\right)^{\frac{1}{t}}\right), \tag{2.16}$$

where we employ the fact that $\lambda(XY) = \lambda(YX)$ for any two square matrices X, Y in Eq. (2.15) (see e.g. [82, Example 1.19]). The last inequality (2.16) follows from Eq. (2.8) in Lemma 2.2. Next, applying Lemma 2.3 on the above inequality yields

$$\lambda\left((A \#_\tau B)^t\right) \prec_{\mathrm{w}} \lambda\left(A^{\frac{t(1-\tau)}{2}} B^{t\tau} A^{\frac{t(1-\tau)}{2}}\right). \tag{2.17}$$

Finally, since the trace function is the summation of eigenvalues, the weak majorization in Eq. (2.17) implies the trace norm inequality in Eq. (2.13). □

Lemma 2.7 (Golden-Thompson Inequality [60, 61]). *For any two For any two operators $A, B \geq 0$, it follows that*

$$\mathrm{Tr}\left[\mathrm{e}^{A+B}\right] \leq \mathrm{Tr}\left[\mathrm{e}^{A}\, \mathrm{e}^{b}\right]. \tag{2.18}$$

Moreover,

doi:10.6342/NTU201800597

Lemma 2.8 ([85, Theorem 1], [86, Theorem 2]). *For any two operators $A, B \geq 0$, and $t \in [0,1]$, we have*

$$\begin{aligned} \mathrm{Tr}\left[A^t B^{1-t}\right] &\geq \mathrm{Tr}\left[\{A-B>0\}B\right] + \mathrm{Tr}\left[\{A-B\leq 0\}A\right] && (2.19)\\ &= \mathrm{Tr}\left[A+B-|A-B|\right]/2. && (2.20)\end{aligned}$$

Lemma 2.9 (Hayashi-Nagaoka Inequality [87, Lemma 2]). *For operators $0 \leq S \leq \mathbb{1}$, and $T \geq 0$, we have*

$$I - (S+T)^{-\frac{1}{2}} S (S+T)^{-\frac{1}{2}} \leq 2(\mathbb{1}-S) + 4T. \tag{2.21}$$

Lemma 2.10 ([88, Lemma 1]). *For any two operators $A, B \geq 0$, and $t \in [0, 1/2]$, we have*

$$\mathrm{Tr}\left[A^t B^{1-t}\right] \geq \mathrm{Tr}\left[\left\{A^{1-t} - B^{1-t} > 0\right\} B\right] + \mathrm{Tr}\left[\left\{A^{1-t} - B^{1-t} \leq 0\right\} A\right]. \tag{2.22}$$

Lemma 2.11 ([82, Theorem 3.23]). *Let $\boldsymbol{A}, \boldsymbol{X}$ be $d \times d$ Hermitian matrices, and $t \in \mathbb{R}$. Assume $f : I \to \mathbb{R}$ is a continuously differentiable function. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Tr}\, f(\boldsymbol{A} + t\boldsymbol{X})\Big|_{t=t_0} = \mathrm{Tr}[\boldsymbol{X} f'(\boldsymbol{A} + t_0 \boldsymbol{X})].$$

Lemma 2.12. *For all positive semi-definite operators A, B $s \in [0,1]$ and $\gamma > 0$,*

$$\|(A+\gamma\mathbb{1})^s - (B+\gamma\mathbb{1})^s\|_\infty \leq (\|A-B\|_\infty + \gamma)^s - \gamma^s. \tag{2.23}$$

Proof of Lemma 2.12. The proof follows similar argument in Ref. [89]. Since the claim holds trivially for $s \in \{0,1\}$, we only prove the case of $s \in (0,1)$. Recall the integral representation: for $s \in (0,1)$,

$$\begin{aligned} (A+\gamma\mathbb{1})^s &= \frac{\sin s\pi}{\pi} \int_0^\infty t^{s-1} \frac{A+\gamma\mathbb{1}}{A+(\gamma+t)\mathbb{1}}\, \mathrm{d}t, && (2.24)\\ &= \frac{\sin s\pi}{\pi} \int_0^\infty t^{s-1} \left[\mathbb{1} - \frac{t\mathbb{1}}{A+(\gamma+t)\mathbb{1}}\right] \mathrm{d}t. && (2.25)\end{aligned}$$

Then, it suffices to prove

$$\left\| \frac{t\mathbb{1}}{B+(\gamma+t)\mathbb{1}} - \frac{t\mathbb{1}}{A+(\gamma+t)\mathbb{1}} \right\|_\infty \leq \frac{t}{\gamma+t} - \frac{t}{\|A-B\|_\infty + \gamma + t}. \tag{2.26}$$

We first show Eq. (2.26) with the assumption $A - B =: C \geq 0$. Replacing B and C by B/t and C/t respectively and denoting $x := 1 + \gamma/t$, Eq. (2.26) is equivalent to

$$\begin{aligned} \left\|(B+x\mathbb{1})^{-1} - (B+C+x\mathbb{1})^{-1}\right\|_\infty &\leq x^{-1} - (\|C\|_\infty + x)^{-1} && (2.27)\\ &= \left\|x^{-1}\mathbb{1} - (C+x\mathbb{1})^{-1}\right\|_\infty. && (2.28)\end{aligned}$$

Since

$$\begin{aligned} &(B+x\mathbb{1})^{-1} - (B+C+x\mathbb{1})^{-1} \\ &= (B+x\mathbb{1})^{-\frac{1}{2}} \left(\mathbb{1} - \left[(B+x\mathbb{1})^{-\frac{1}{2}} C (B+x\mathbb{1})^{-\frac{1}{2}} + \mathbb{1}\right]\right)^{-1} (B+x\mathbb{1})^{-\frac{1}{2}}, \end{aligned} \tag{2.29}$$

doi:10.6342/NTU201800597

the sub-multiplicativity of operator norm thus yields

$$\left\|(B+x\mathbb{1})^{-1}-(B+C+x\mathbb{1})^{-1}\right\|_\infty \tag{2.30}$$
$$\leq \left\|(B+x\mathbb{1})^{-\frac{1}{2}}\right\|_\infty \left\|\mathbb{1}-\left((B+x\mathbb{1})^{-\frac{1}{2}}C\,(B+x\mathbb{1})^{-\frac{1}{2}}+\mathbb{1}\right)^{-1}\right\|_\infty \left\|(B+x\mathbb{1})^{-\frac{1}{2}}\right\|_\infty \tag{2.31}$$
$$\leq x^{-1}\left\|\mathbb{1}-\left((B+x\mathbb{1})^{-\frac{1}{2}}C\,(B+x\mathbb{1})^{-\frac{1}{2}}+\mathbb{1}\right)^{-1}\right\|_\infty . \tag{2.32}$$

Further, the fact

$$C^{\frac{1}{2}}\,(B+x\mathbb{1})^{-1}\,C^{\frac{1}{2}}\leq x^{-1}C. \tag{2.33}$$

implies that

$$x^{-1}\left\|\mathbb{1}-\left((B+x\mathbb{1})^{-\frac{1}{2}}C\,(B+x\mathbb{1})^{-\frac{1}{2}}+\mathbb{1}\right)^{-1}\right\|_\infty = x^{-1}\left\|\mathbb{1}-\left(C^{\frac{1}{2}}\,(B+x\mathbb{1})^{-1}\,C^{\frac{1}{2}}+\mathbb{1}\right)^{-1}\right\|_\infty \tag{2.34}$$
$$\leq x^{-1}\left\|\mathbb{1}-\left(x^{-1}C+\mathbb{1}\right)^{-1}\right\|_\infty \tag{2.35}$$
$$= \left\|x^{-1}\mathbb{1}-(C+x\mathbb{1})^{-1}\right\|_\infty , \tag{2.36}$$

which establishes Eq. (2.28).

Lastly, we consider the general case $A,B\geq 0$. Denoting by $f(u):=(u+\gamma)^s$. It is clearly that $u\mapsto f(u)$ is an operator monotone function (see e.g. [67, Theorem V.1.9]). Then, the inequality $0\leq A\leq B+(A-B)_+$ implies

$$f(A)-f(B)\leq f(B+(A-B)_+)-f(B), \tag{2.37}$$

which in turn yields

$$\left\|(f(A)-f(B))_+\right\|_\infty \leq \left\|f(B+(A-B)_+)-f(B)\right\|_\infty . \tag{2.38}$$

On the other hand, the established Eq. (2.23) with the pair $\{B+(A-B)_+,B\}$ leads to

$$\left\|f(B+(A-B)_+)-f(B)\right\|_\infty \leq f(\left\|(A-B)_+\right\|_\infty)-f(0) \tag{2.39}$$
$$= \left\|f((A-B)_+)-f(0)\mathbb{1}\right\|_\infty . \tag{2.40}$$

Combing Eqs. (2.38) and (2.40) gives

$$\left\|(f(A)-f(B))_+\right\|_\infty \leq \left\|f((A-B)_+)-f(0)\mathbb{1}\right\|_\infty . \tag{2.41}$$

Exchanging the role of A and B, we have

$$\left\|(f(B)-f(A))_+\right\|_\infty \leq \left\|f((B-A)_+)-f(0)\mathbb{1}\right\|_\infty . \tag{2.42}$$

doi:10.6342/NTU201800597

Then, our claim holds as follows:

$$\begin{aligned}
\|f(A)-f(B)\|_\infty &= \| \, |f(A)-f(B)| \, \|_\infty && (2.43)\\
&= \left\| (f(A)-f(B))_+ \oplus (f(B)-f(A))_+ \right\|_\infty && (2.44)\\
&\le \| [f((A-B)_+)-f(0)\mathbb{1}] \oplus [f((B-A)_+)-f(0)\mathbb{1}] \|_\infty && (2.45)\\
&= \| f(|A-B|)-f(0)\mathbb{1} \|_\infty && (2.46)\\
&= f\left(\|A-B\|_\infty\right)-f(0). && (2.47)
\end{aligned}$$

□

Lemma 2.13. *[90, Corollary 3.6] Let A_i be $m \times m$ positive semi-definite matrix and Z_i be $n \times m$ matrix for $i = 1, \ldots, k$. Then, for all unitarily invariant norms $\|\cdot\|$ and $\gamma > 0$, the map*

$$(p,t) \mapsto \left\| \left(\sum_{i=1}^{k} Z_i^* A_i^{t/p} Z_i \right)^{\gamma p} \right\| \tag{2.48}$$

is jointly log-convex on $(0,+\infty) \times (-\infty,+\infty)$.

2.2 Large Deviation Theory

In this section, we will see that the Lengendre-Fenchel transform is closely related to the error-exponent function of hypothesis testing and channel coding. Consider the following binary classical hypotheses:

$$\begin{aligned}
\mathsf{H}_0 &: p^n := p_{x_1} \otimes p_{x_2} \otimes \cdots p_{x_n},\\
\mathsf{H}_1 &: q^n := q_{x_1} \otimes q_{x_2} \otimes \cdots q_{x_n},
\end{aligned} \tag{2.49}$$

where p_{x_i}, q_{x_i} are probability mass functions; and x_i belongs to some finite alphabet $\mathcal{X}$ and $n \in \mathbb{N}$ be fixed. Given any $r \ge 0$, recall the definition of the error-exponent function in Eq. (4.7):

$$\phi_n(r) = \phi_n(r|p^n\|q^n) = \sup_{\alpha \in (0,1]} \left\{ \frac{1-\alpha}{\alpha} \left(\frac{1}{n} D_\alpha \left(p^n \| q^n\right) - r \right) \right\}. \tag{2.50}$$

Without loss of generality, we assume that $p^n \ll q^n$ have the same support since elements of q_{x_i}, that do not lie in the support of p_{x_i}, do not contribute to $\phi_n(r)$.

Let Z be a random variable with probability measure μ. Further, we assume Z is finite on $\mathtt{supp}(\mu)$. The cumulant generating function (c.g.f.) of Z is defined as

$$\Lambda(t) := \log \mathbb{E}_\mu \left[\mathrm{e}^{tZ} \right], \quad t \in \mathbb{R}. \tag{2.51}$$

The *Lengendre-Fenchel transform* of $\Lambda(t)$ is

$$\Lambda^*(z) := \sup_{t \in \mathbb{R}} \left\{ zt - \Lambda(t) \right\}. \tag{2.52}$$

Such a transform plays a significant role in concentration inequalities, convex analysis, and large deviation theory [27].

doi:10.6342/NTU201800597

Let $P_{\mathbf{x}^n}$ be the empirical distribution of the sequence $\mathbf{x}^n = x_1 x_2 \dots x_n$. Let $Z_0 = \log \frac{q^n}{p^n}$ with probability measure p^n, $Z_1 = \log \frac{p^n}{q^n}$ with probability measure q^n, and denote

$$\begin{aligned} \Lambda_{0,P_{\mathbf{x}^n}}(t) &:= \frac{1}{n} \log \mathbb{E}_{p^n}\left[\mathrm{e}^{tZ_0}\right] = \sum_{x\in\mathcal{X}} P_{\mathbf{x}^n}(x)\Lambda_{0,x_i}(t), \\ \Lambda_{1,P_{\mathbf{x}^n}}(t) &:= \frac{1}{n} \log \mathbb{E}_{q^n}\left[\mathrm{e}^{tZ_1}\right] = \sum_{x\in\mathcal{X}} P_{\mathbf{x}^n}(x)\Lambda_{1,x_i}(t); \end{aligned} \tag{2.53}$$

where

$$\Lambda_{0,x_i}(t) := \log \mathbb{E}_{p_{x_i}}\left[\mathrm{e}^{t\log\frac{q_{x_i}}{p_{x_i}}}\right], \quad \Lambda_{1,x_i}(t) := \log \mathbb{E}_{q_{x_i}}\left[\mathrm{e}^{t\log\frac{p_{x_i}}{q_{x_i}}}\right]. \tag{2.54}$$

Rewrite the right-hand side of Eq. (2.50) with $\alpha = \frac{1}{1+s}$, and observe that

$$\sum_{x\in\mathcal{X}} P_{\mathbf{x}^n}(x) s D_{\frac{1}{1+s}}(p_x\|q_x) = -(1+s)\Lambda_{0,P_{\mathbf{x}^n}}\left(\frac{s}{1+s}\right) \tag{2.55}$$

$$=: E_0^{(2)}(s, P_{\mathbf{x}^n}). \tag{2.56}$$

Then the error-exponent function in Eq. (2.50) can also be viewed as a Lengendre-Fenchel transform of $E_0^{(2)}(s, P_{\mathbf{x}^n})$:

$$\phi_n(r) = \sup_{s\geq 0}\left\{E_0^{(2)}(s, P_{\mathbf{x}^n}) - sr\right\}. \tag{2.57}$$

The following lemma relates $\phi_n(r)$ to $\Lambda^*_{j,P_{\mathbf{x}^n}}(z)$, the Lengendre-Fenchel transform of Eq. (2.53):

$$\Lambda^*_{j,P_{\mathbf{x}^n}}(z) := \sup_{t\in\mathbb{R}}\left\{tz - \Lambda_{j,P_{\mathbf{x}^n}}(t)\right\}, \quad j\in\{0,1\}. \tag{2.58}$$

Lemma 2.14 (Regularity). *Let p^n and q^n, $n\in\mathbb{N}$, be described as above. Assume $r > \frac{1}{n}D_0(p^n\|q^n)$ and $\phi_n(r) > 0$. The following hold:*

(a) $\Lambda''_{0,P_{\mathbf{x}^n}}(t) > 0$ *for all* $t\in[0,1]$.

(b) $\Lambda^*_{0,P_{\mathbf{x}^n}}(\phi_n(r) - r) = \phi_n(r)$.

(c) $\Lambda^*_{1,P_{\mathbf{x}^n}}(r - \phi_n(r)) = r$.

(d) Let $t^\star := t^\star_{r,P_{\mathbf{x}^n}}$ *be the optimizer of* $\Lambda^*_{0,P_{\mathbf{x}^n}}(z)$ *in Eq. (2.58), and* $s^\star := s^\star_{r,P_{\mathbf{x}^n}}$ *be the optimizer of* $\phi_n(r)$ *in Eq. (2.57). The optimizer* $t^\star\in(0,1)$ *is unique, and satisfies* $\Lambda'_{0,P_{\mathbf{x}^n}}(t^\star) = \phi_n(r) - r$. *In particular, one has* $t^\star = \frac{s^\star}{1+s^\star}$; $s^\star = -\frac{\partial\phi_n(r)}{\partial r}$; *and* $\frac{\partial^2\phi_n(r)}{\partial r^2} = -\left(\frac{\partial^2 E_0^{(2)}(s,P_{\mathbf{x}^n})}{\partial s^2}\Big|_{s=s^\star}\right)^{-1} = \frac{\left(1+s^\star_{r,P_{\mathbf{x}^n}}\right)^3}{\Lambda''_{0,P_{\mathbf{x}^n}}(t^\star)} > 0$.

doi:10.6342/NTU201800597

Before proving Lemma 2.14, we will need the following partial derivatives with respect to t:

$$\Lambda'_{0,x_i}(t) = \mathbb{E}_{\hat{q}_{x_i,t}}\left[\log\frac{q_{x_i}}{p_{x_i}}\right], \quad \Lambda'_{1,x_i}(t) = \mathbb{E}_{\hat{q}_{x_i,1-t}}\left[\log\frac{p_{x_i}}{q_{x_i}}\right]; \tag{2.59}$$

$$\Lambda''_{0,x_i}(t) = \mathrm{Var}_{\hat{q}_{x_i,t}}\left[\log\frac{q_{x_i}}{p_{x_i}}\right], \quad \Lambda''_{1,x_i}(t) = \mathrm{Var}_{\hat{q}_{x_i,1-t}}\left[\log\frac{p_{x_i}}{q_{x_i}}\right], \tag{2.60}$$

where we denote the *tilted distributions* for every $i \in [n]$ and $t \in [0,1]$ by

$$\hat{q}_{x_i,t}(\omega) := \frac{p_{x_i}(\omega)^{1-t} q_{x_i}(\omega)^t}{\sum_{\omega\in\mathtt{supp}(p_{x_i})} p_{x_i}(\omega)^{1-t} q_{x_i}(\omega)^t}, \quad \omega \in \mathtt{supp}(p_{x_i}). \tag{2.61}$$

It is also easy to verify that

$$\Lambda_{0,x_i}(t) = \Lambda_{1,x_i}(1-t), \quad \Lambda'_{0,x_i}(t) = -\Lambda'_{1,x_i}(1-t), \quad \Lambda''_{0,x_i}(t) = \Lambda''_{1,x_i}(1-t). \tag{2.62}$$

This lemma closely follows Ref. [91, Lemma 9]; however, the major difference is that we prove the claim using $\phi_n(r|\rho^n\|\sigma^n)$ in Eq. (4.7) instead of the discrimination function: $\min\{D(\tau\|\rho) : D(\tau\|\sigma) \le r\}$ in Eq. (9.20). This expression is crucial to obtaining the sphere-packing bound in Theorem 11.1 in the strong from, cf. Eq. (1.4), instead of the weak form, cf. Eq. (1.5).

Proof of Lemma 2.14.

(2.14-(a)) We will prove this statement by contradiction. Let $t \in [0,1]$, Assuming that $\Lambda''_{0,P_{\mathbf{x}^n}}(t) = 0$, implies $\Lambda''_{0,x}(t) = 0$, $\forall x \in \mathtt{supp}(P_{\mathbf{x}^n})$. Recall from Eq. (2.60)

$$0 = \Lambda''_{0,x}(t) = \mathrm{Var}_{\hat{q}_{x,t}}\left[\log\frac{q_x}{p_x}\right], \tag{2.63}$$

which is equivalent to

$$p_x(\omega) = q_x(\omega)\cdot \mathrm{e}^{-\Lambda'_{0,x}(t)}, \quad \forall\omega \in \mathtt{supp}(p_x). \tag{2.64}$$

Summing both sides of Eq. (2.64) over $\omega \in \mathtt{supp}(p_x)$ gives

$$1 = \mathrm{Tr}\left[p_x^0 q_x\right]\mathrm{e}^{-\Lambda'_{0,x}(t)}. \tag{2.65}$$

Then, Eqs. (2.64) and (2.65) imply that

$$\phi_n(r) = \sup_{0<\alpha\le 1}\frac{\alpha-1}{\alpha}\left(r - \sum_{x\in\mathcal{X}} P_{\mathbf{x}^n}(x) D_\alpha\left(p_x\|q_x\right)\right) \tag{2.66}$$

$$= \sup_{0<\alpha\le 1}\frac{\alpha-1}{\alpha}\left(r + \sum_{x\in\mathcal{X}} P_{\mathbf{x}^n}(x)\log\mathrm{Tr}\left[p_x^0 q_x\right]\right) \tag{2.67}$$

$$= 0, \tag{2.68}$$

where Eq. (2.68) follows since $r > \frac{1}{n}D_0(p^n\|q^n) = -\frac{1}{n}\sum_{x\in\mathcal{X}} P_{\mathbf{x}^n}(x)\log\mathrm{Tr}\left[p_x^0 q_x\right]$ by assumption. However, this contradicts with the assumption $\phi_n(r) > 0$. Hence, we conclude item (a).

doi:10.6342/NTU201800597

(2.14-(b)) Observe that $E_0^{(2)}(s, P_{\mathbf{x}^n}) - sr$ in Eq. (2.57) is strictly concave in $s \in \mathbb{R}_{\geq 0}$ since

$$\frac{\partial^2 E_0^{(2)}(s, P_{\mathbf{x}^n})}{\partial s^2} = -\frac{1}{(1+s)^3}\Lambda''_{0,P_{\mathbf{x}^n}}\left(\frac{s}{1+s}\right) < 0, \tag{2.69}$$

owing to Eqs. (2.56), (2.60), and Lemma (a). Moreover, $s = 0$ cannot be an optimum in Eq. (2.57); otherwise, it will violate the assumption $\phi_n(r) \geq 0$. Thus a unique maximizer $s^\star \in \mathbb{R}_{>0}$ exists such that

$$\phi_n(r) = -s^\star r + E_0^{(2)}(s^\star, P_{\mathbf{x}^n}) \tag{2.70}$$

$$= \frac{s^\star}{1+s^\star}\Lambda'_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right) - \Lambda_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right). \tag{2.71}$$

where in the second equality we use Eq. (2.56) and

$$r = \left.\frac{\partial E_0^{(2)}(s, P_{\mathbf{x}^n})}{\partial s}\right|_{s=s^\star} \tag{2.72}$$

$$= -\frac{1}{1+s^\star}\Lambda'_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right) - \Lambda_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right). \tag{2.73}$$

Comparing Eq. (2.71) with (2.73) gives

$$\Lambda'_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right) = \phi_n(r) - r, \tag{2.74}$$

which is exactly the optimum solution to $\Lambda^*_{0,P_{\mathbf{x}^n}}(z)$ in Eq. (2.58) with

$$t^\star = \frac{s^\star}{1+s^\star} \in (0,1), \tag{2.75}$$

$$z = \phi_n(r) - r. \tag{2.76}$$

Hence, we obtain

$$\Lambda^*_{0,P_{\mathbf{x}^n}}(\phi_n(r) - r) = t^\star z - \Lambda_{0,P_{\mathbf{x}^n}}(t^\star) \tag{2.77}$$

$$= \frac{s^\star}{1+s^\star}(\phi_n(r) - r) - \Lambda_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right) \tag{2.78}$$

$$= \frac{s^\star}{1+s^\star}\Lambda'_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right) - \Lambda_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right) \tag{2.79}$$

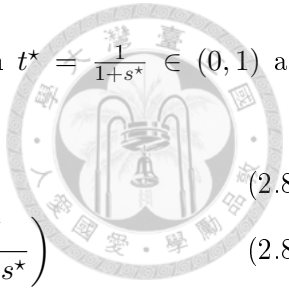
$$= \phi_n(r), \tag{2.80}$$

where Eqs. (2.74) and (2.71) are used in the third and last equalities.

(2.14-(c)) This proof follows from similar arguments in item (b) and Eq. (2.62). Eqs. (2.74) and (2.62) lead to

$$\Lambda'_{1,P_{\mathbf{x}^n}}\left(\frac{1}{1+s^\star}\right) = r - \phi_n(r), \tag{2.81}$$

doi:10.6342/NTU201800597



which satisfies the optimum solution to $\Lambda_{1,P_{\mathbf{x}^n}}(z)$ in Eq. (2.58) with $t^\star = \frac{1}{1+s^\star} \in (0,1)$ and $z = r - \phi_n(r)$. Then,

$$\begin{aligned}
\Lambda^*_{1,P_{\mathbf{x}^n}}(r-\phi_n(r)) &= t^\star z - \Lambda_{1,P_{\mathbf{x}^n}}(t^\star) && (2.82)\\
&= \frac{1}{1+s^\star}(r-\phi_n(r)) - \Lambda_{1,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right) && (2.83)\\
&= \frac{1}{1+s^\star}\Lambda'_{1,P_{\mathbf{x}^n}}\left(\frac{1}{1+s^\star}\right) - \Lambda_{1,P_{\mathbf{x}^n}}\left(\frac{1}{1+s^\star}\right) && (2.84)\\
&= r, && (2.85)
\end{aligned}$$

where the third equality is due to Eq. (2.81), and the last equality follows from Eqs. (2.62) and (2.73).

(2.14-(d)) The fact that a unique optimizer $t^\star \in (0,1)$ exists such that $\Lambda'_{0,P_{\mathbf{x}^n}}(t^\star) = \phi_n(r) - r$ follows directly from Eqs. (2.74), (2.75) and $\Lambda''_{0,P_{\mathbf{x}^n}}(t) > 0$, for $t \in [0,1]$.

Moreover, Eqs. (2.70), (2.72), and (2.69) yield

$$\begin{aligned}
-\frac{\partial\phi_n(r)}{\partial r} &= s^\star, && (2.86)\\
\frac{\partial^2\phi_n(r)}{\partial r^2} &= -\frac{\partial s^\star}{\partial r} = -\left(\frac{\partial^2 E_0^{(2)}(s,P_{\mathbf{x}^n})}{\partial s^2}\right)^{-1}\Bigg|_{s=s^\star} = \frac{(1+s^\star)^3}{\Lambda_{0,P_{\mathbf{x}^n}}\left(\frac{s^\star}{1+s^\star}\right)}, && (2.87)
\end{aligned}$$

which completes the claim in item (d).

□

Let $(Z_i)_{i=1}^n$ be a sequence of independent, real-valued random variables with probability measures $(\mu_i)_{i=1}^n$. Let $\Lambda_i(t) := \log \mathbb{E}\left[\mathrm{e}^{tZ_i}\right]$ and define the Legendre-Fenchel transform of $\frac{1}{n}\sum_{i=1}^n \Lambda_i(\cdot)$ to be:

$$\Lambda^*_n(z) := \sup_{t\in\mathbb{R}}\left\{zt - \frac{1}{n}\sum_{i=1}^n \Lambda_i(t)\right\}, \quad \forall z \in \mathbb{R}. \tag{2.88}$$

Then there exists a real number $t^\star \in (0,1]$ for every $z \in \mathbb{R}$ such that

$$\begin{aligned}
z &= \frac{1}{n}\sum_{i=1}^n \Lambda'_i(t^\star); && (2.89)\\
\Lambda^*_n(z) &= zt^\star - \frac{1}{n}\sum_{i=1}^n \Lambda_i(t^\star). && (2.90)
\end{aligned}$$

Define the probability measure $\tilde{\mu}_i$ via

$$\frac{\mathrm{d}\tilde{\mu}_i}{\mathrm{d}\mu_i}(z_i) := \mathrm{e}^{t^\star z_i - \Lambda_i(t^\star)}, \tag{2.91}$$

and let $\bar{Z}_i := Z_i - \mathbb{E}_{\tilde{\mu}_i}[Z_i]$. Furthermore, define $m_{2,n} := \sum_{i=1}^n \mathrm{Var}_{\tilde{\mu}_i}\left[\bar{Z}_i\right]$, $m_{3,n} := \sum_{i=1}^n \mathbb{E}_{\tilde{\mu}_i}\left[\left|\bar{Z}_i\right|^3\right]$, and $K_n(t^\star) := \frac{15\sqrt{2\pi}m_{3,n}}{m_{2,n}}$. With these definitions, we can now state the following sharp concentration

doi:10.6342/NTU201800597

inequality for $\frac{1}{n}\sum_{i=1}^{n} Z_i$:

Theorem 2.1 (Bahadur-Ranga Rao's Concentration Inequality [91, Proposition 5], [48]). *Provided that* $\sqrt{m_{2,n}} \geq 1 + (1 + K_n(t^\star))^2$, *then*

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i \geq z\right\} \geq \mathrm{e}^{-n\Lambda_n^*(z)} \frac{\mathrm{e}^{-K_n(t^\star)}}{2\sqrt{2\pi m_{2,n}}}. \tag{2.92}$$

Chaganty and Sethuraman in Ref. [49, Theorem 3.3] considered a more general sequence of random variables $\{Z_n\}_{n\in\mathbb{N}}$, which are not necessarily the sum of random variables.

Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent, real-valued random variables with probability measures $(\mu_i)_{i=1}^n$. Let $Z_n := \sum_{i=1}^n X_i$ and let $\Lambda_n(t) := \log \mathbb{E}\left[\mathrm{e}^{tZ_n}\right]$. Define the Legendre-Fenchel transform of $\frac{1}{n}\Lambda_n(\cdot)$ by:

$$\Lambda_n^*(z) := \sup_{t\in\mathbb{R}} \left\{zt - \frac{1}{n}\Lambda_n(t)\right\}, \quad \forall z \in \mathbb{R}. \tag{2.93}$$

Let $(T_n)_{n\in\mathbb{N}}$ be a bounded sequence of real numbers and $(t_n^\star)_{n\in\mathbb{N}}$ be a sequence satisfying for all $n \in \mathbb{N}$

$$t_n^\star \in (0,1); \quad T_n = \frac{1}{n}\Lambda_n'(t_n^\star); \quad \Lambda_n^*(T_n) = T_n t_n^\star - \frac{1}{n}\Lambda_n(t_n^\star). \tag{2.94}$$

With these definitions, we can now state the following sharp concentration inequality for $\frac{1}{n}Z_n$:

Theorem 2.2 (Chaganty-Sethuraman's Concentration Inequality [49, Theorem 3.3]). *For any* $\eta \in (0,1)$, *there exists an* $N_0 \in \mathbb{N}$ *such that, for all* $n \geq N_0$,

$$\Pr\left\{\frac{1}{n}Z_n \geq T_n, \right\} \geq \frac{1-\eta}{t_n^\star\sqrt{2\pi n m_{2,n}}} \exp\{-n\Lambda_n^*(T_n)\}, \tag{2.95}$$

where $m_{2,n} := \frac{1}{n}\sum_{i=1}^n \mathrm{Var}_{\tilde{\mu}_{n,i}}[X_i]$, *and the measure* $\tilde{\mu}_{n,i}$ *is defined via*

$$\frac{\mathrm{d}\tilde{\mu}_{n,i}}{\mathrm{d}\mu_i}(y) := \frac{\mathrm{e}^{yt_n^\star}}{\mathbb{E}\left[\mathrm{e}^{t_n^\star X_i}\right]}. \tag{2.96}$$

Remark 2.1. Chaganty and Sethuraman proved Theorem 2.2 provided that the following condition is satisfied: there exists $\delta_0 > 0$ such that for any δ and λ with $0 < \delta < \delta_0 < \lambda$, $\sup_{\delta<|t|\leq\lambda t_n^\star} |\exp\{\Lambda_n(t_n^\star + \mathrm{i}t)\}/\exp\{\Lambda_n(t$ $o(1/\sqrt{n})$, where the supremum is defined to be 0 if $\{t : \delta < |t| \leq \lambda t_n^\star\}$ is empty. In the case of Z_n being a sum of random variables, $\exp\{\Lambda_n(t_n^\star + \mathrm{i}t)\}/\exp\{\Lambda_n(t_n^\star)\}$ is the product of the characteristic functions of $\{X_i\}_{i=1}^n$. Since the supremum of a characteristic function on a compact interval not containing 0 is less than 1, this condition is thus satisfied.

We note that the lower bound in Theorem 2.2 for the general sequence of random variables $(X_i)_{i\in\mathbb{N}}$ suffices to establish the converse bound in moderate deviation analysis for c-q channel coding, Theorem 12.2 in Chapter 12 later. We do not particularly consider the case of lattice valued random variables (see e.g. [49, Theorem 3.5]). ◇

doi:10.6342/NTU201800597

Chapter 3

Quantum Entropic Quantities and Notation

In this chapter, we introduce necessary notation in quantum information theory. In Section 3.1, we present various quantum generalizations of the classical Rényi divergence [92], and their mathematical properties. As we will see in quantum hypothesis testing discussed in Chapter 4, some specific definitions of the quantum Rényi divergence naturally arise in the exponent function. In Sections 3.2 and 3.3, we define the conditional Rényi entropies and Rényi mutual information, which play significant roles in th Parts II and III, respectively. We refer the interested readers to books [93, 50, 8] for more comprehensive discussions.

Notation. Throughout this thesis, we consider a finite-dimensional Hilbert space $\mathcal{H}$. The set of density operators (i.e. positive semi-definite operators with unit trace) and the set of full-rank density operators on $\mathcal{H}$ are defined as $\mathcal{S}(\mathcal{H})$ and $\mathcal{S}(\mathcal{H})_{>0}$. For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, we write $\rho \ll \sigma$ if the support of ρ is contained in the support of σ. The identity operator on $\mathcal{H}$ is denoted by $\mathbb{1}_{\mathcal{H}}$. If there is no possibility of confusion, we will skip the subscript $\mathcal{H}$. We use $\mathrm{Tr}\left[\cdot\right]$ as the standard trace function. Let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{R}_{>0}$ denote the set of integers, real numbers, non-negative real numbers, and positive real numbers, respectively. Define $[n] := \{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$.

For a positive semi-definite operator A whose spectral decomposition is $A = \sum_i a_i P_i$, where $(a_i)_i$ and $(P_i)_i$ are the eigenvalues and eigenprojections of A, its power is defined as: $A^p := \sum_{i:a_i \neq 0} a_i^p P_i$. In particular, A^0 denotes the projection onto the support of A. We use $\mathtt{supp}(A)$ to denote the support of the operator A. Further, $A \perp B$ means $\mathtt{supp}(A) \cap \mathtt{supp}(B) = \emptyset$.

Given a pair of positive semi-definite operators $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, we define *quantum relative entropy* [94, 95] as

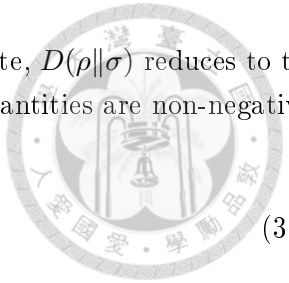
$$D(\rho\|\sigma) := \mathrm{Tr}\left[\rho\left(\log\rho - \log\sigma\right)\right]. \tag{3.1}$$

We define two types of the *quantum relative entropy variances* [14, 15, 16] by

$$V(\rho\|\sigma) := \mathrm{Tr}\left[\rho\left(\log\rho - \log\sigma\right)^2\right] - D(\rho\|\sigma)^2 \tag{3.2}$$

$$\widetilde{V}(\rho\|\sigma) := \int_0^1 \mathrm{d}t\, \mathrm{Tr}\left[\rho^{1-t}(\log\rho - \log\sigma)\rho^t(\log\rho - \log\sigma)\right] - D(\rho\|\sigma)^2. \tag{3.3}$$

doi:10.6342/NTU201800597



They are defined to be $+\infty$ when $\rho \not\ll \sigma$ We note that when ρ and σ commute, $D(\rho\|\sigma)$ reduces to the classical Kullback-Leibler divergence [96]. It is well-known that both the quantities are non-negative, and $D(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$, which in turn shows that

$$V(\rho\|\sigma) > 0 \quad \text{implies} \quad D(\rho\|\sigma) > 0. \tag{3.4}$$

3.1 Quantum Rényi divergence

For density operators $\rho, \sigma \in \mathcal{S}(\mathcal{H})_{>0}$, and every $\alpha \in [0, 1)$, we define the following two families of quantum Rényi divergences [59, 57, 58]:

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log Q_\alpha(\rho\|\sigma), \quad Q_\alpha(\rho\|\sigma) := \mathrm{Tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]; \tag{3.5}$$

$$D^\flat_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log Q^\flat_\alpha(\rho\|\sigma), \quad Q^\flat_\alpha(\rho\|\sigma) := \mathrm{Tr}\left[\mathrm{e}^{\alpha \log \rho + (1-\alpha)\log \sigma}\right]. \tag{3.6}$$

We term the above quantities as the *(Petz) α-Rényi divergence*, and the *log-Euclidean α-Rényi divergence*, respectively. The log-Euclidean Rényi divergence arises from the *log-Euclidean operator mean* (also called the *chaotic mean*): $A \diamondsuit_\alpha B := \exp\left((1-\alpha)\log A + \alpha \log B\right)$ for $0 \le \alpha \le 1$. For general density operators $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, the above definitions can be extended as

$$Q_\alpha(\rho\|\sigma) := \lim_{\delta \searrow 0} Q_\alpha(\rho + \delta \mathbb{1}\|\sigma + \delta \mathbb{1}) \quad \text{and} \quad Q^\flat_\alpha(\rho\|\sigma) := \lim_{\delta \searrow 0} Q^\flat_\alpha(\rho + \delta \mathbb{1}\|\sigma + \delta \mathbb{1}). \tag{3.7}$$

For $\alpha = 1$, we define (see e.g. [58, Lemma III.4]):

$$Q_1(\rho\|\sigma) := \mathrm{Tr}\left[\rho\sigma^0\right] \quad \text{and} \quad Q^\flat_1(\rho\|\sigma) := \mathrm{Tr}\left[\rho\sigma^0\right]; \tag{3.8}$$

$$D_1(\rho\|\sigma) := \lim_{\alpha \to 1} D_\alpha(\rho\|\sigma) = \mathbb{D}(\rho\|\sigma) \quad \text{and} \quad D^\flat_1(\rho\|\sigma) := \lim_{\alpha \to 1} D^\flat_\alpha(\rho\|\sigma) = \mathbb{D}(\rho\|\sigma). \tag{3.9}$$

In addition, these two quantities are related by the Golden-Thompson inequality given in Lemma 2.7:

$$Q^\flat_\alpha(\rho\|\sigma) \le Q_\alpha(\rho\|\sigma), \ \forall \alpha \in [0, 1]. \tag{3.10}$$

The log-Euclidean Rényi divergence is closely related to the quantum version of the *Hellinger arc* in statistics [97, 98], [58, Seciont III]. Lemma 3.1 will useful to prove the variational representations in Sections 5.1 and 9.1 later.

Lemma 3.1 ([58, Theorem III.5]). *Let $\rho, \tau \in \mathcal{S}(\mathcal{H})$ with $\rho \ll \tau$. For all $s > -1$, it follows that*

$$\min_{\sigma \in \mathcal{S}(\mathcal{H})} D(\sigma\|\rho) + sD(\sigma\|\tau) = sD^\flat_{\frac{1}{1+s}}(\rho\|\tau). \tag{3.11}$$

In the following, we provide useful mathematical properties. Most of them can be found in Refs. [99, 58, 100].

doi:10.6342/NTU201800597

Lemma 3.2. *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$. Then,*

$$\alpha \mapsto \log Q_\alpha(\rho\|\sigma) \text{ and } \alpha \mapsto \log Q^\flat_\alpha(\rho\|\sigma) \text{ are convex on } (0,1); \tag{3.12}$$

$$\alpha \mapsto D_\alpha(\rho\|\sigma) \text{ is continuous and monotone increasing on } [0,1]; \tag{3.13}$$

$$\forall \alpha \in (0,1), \quad (\rho,\sigma) \mapsto Q^\flat_\alpha(\rho\|\sigma) \text{ is jointly concave on } \mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H}); \tag{3.14}$$

$$\forall \alpha \in [0,1], \quad \sigma \mapsto D_\alpha(\rho\|\sigma) \text{ is convex and lower semi-continuous on } \mathcal{S}(\mathcal{H}). \tag{3.15}$$

For every $\rho \in \mathcal{S}(\mathcal{H})$ and $\gamma > 0$, the map

$$(\alpha,\sigma) \mapsto Q_\alpha(\rho\|\sigma + \gamma\mathbb{1}) \text{ is continuous on } [0,1] \times \mathcal{S}(\mathcal{H}). \tag{3.16}$$

Moreover, for every $\rho \in \mathcal{S}(\mathcal{H})$, the map

$$(\alpha,\sigma) \mapsto -Q_\alpha(\rho\|\sigma) \text{ is lower-semicontinuous on } [0,1] \times \mathcal{S}(\mathcal{H}), \tag{3.17}$$

and the same argument holds for D_α.

Proof of Lemma 3.2. We note that Eqs. (3.12), (3.13), (3.14), and (3.15) are proved in [99], [58, Lemma III.3, Lemma III.11, Theorem III.14, Corollary III.25], [100, Corollary 2.2][1]. We only prove Eqs. (3.16) and (3.17).

Fix arbitrary $\gamma > 0$, $\alpha_1 \in [0,1]$, $\rho, \sigma_1 \in \mathcal{S}(\mathcal{H})$, $\|\sigma_1 - \sigma_2\|_\infty \leq \varepsilon_1$, and $|\alpha_1 - \alpha_2| \leq \varepsilon_2$. Triangle inequality implies that

$$\begin{aligned} |Q_{\alpha_1}(\rho\|\sigma_1 + \gamma\mathbb{1}) - Q_{\alpha_2}(\rho\|\sigma_2 + \gamma\mathbb{1})| \leq\ & |Q_{\alpha_1}(\rho\|\sigma_2 + \gamma\mathbb{1}) - Q_{\alpha_2}(\rho\|\sigma_2 + \gamma\mathbb{1})| \\ & + |Q_{\alpha_1}(\rho\|\sigma_1 + \gamma\mathbb{1}) - Q_{\alpha_1}(\rho\|\sigma_2 + \gamma\mathbb{1})|. \end{aligned} \tag{3.18}$$

In the following, we upper bound the two terms in the right-hand side of Eq. (3.18), respectively. Without loss of generality, we assume $\alpha_1 \leq \alpha_2$. Direct calculation shows that

$$|Q_{\alpha_1}(\rho\|\sigma_2 + \gamma\mathbb{1}) - Q_{\alpha_2}(\rho\|\sigma_2 + \gamma\mathbb{1})| = \left|\mathrm{Tr}\left[\rho^{\alpha_1}(\sigma_2 + \gamma\mathbb{1})^{1-\alpha_1} - \rho^{\alpha_2}(\sigma_2 + \gamma\mathbb{1})^{1-\alpha_2}\right]\right| \tag{3.19}$$

$$= \left|\mathrm{Tr}\left[\rho^{\alpha_1}\left(\rho^0 - \rho^{\alpha_2-\alpha_1}(\sigma_2 + \gamma\mathbb{1})^{-(\alpha_2-\alpha_1)}\right)(\sigma_2 + \gamma\mathbb{1})^{1-\alpha_1}\right]\right| \tag{3.20}$$

$$\leq d\left\|\rho^{\alpha_1}\left(\rho^0 - \rho^{\alpha_2-\alpha_1}(\sigma_2 + \gamma\mathbb{1})^{-(\alpha_2-\alpha_1)}\right)(\sigma_2 + \gamma\mathbb{1})^{1-\alpha_1}\right\|_\infty \tag{3.21}$$

$$\leq d\|\rho^{\alpha_1}\|_\infty \left\|\rho^0 - \rho^{\alpha_2-\alpha_1}(\sigma_2 + \gamma\mathbb{1})^{-(\alpha_2-\alpha_1)}\right\|_\infty \left\|(\sigma_2 + \gamma\mathbb{1})^{1-\alpha_1}\right\|_\infty \tag{3.22}$$

$$\leq d(1+\gamma)\left\|\rho^0 - \rho^{\alpha_2-\alpha_1}(\sigma_2 + \gamma\mathbb{1})^{-(\alpha_2-\alpha_1)}\right\|_\infty. \tag{3.23}$$

For sufficiently small ε_2, it follows that

$$\left\|\rho^0 - \rho^{\alpha_2-\alpha_1}(\sigma_2 + \gamma\mathbb{1})^{-(\alpha_2-\alpha_1)}\right\|_\infty = 1 - \tilde{\lambda}_{\min}\left(\rho^{\alpha_2-\alpha_1}(\sigma_2 + \gamma\mathbb{1})^{-(\alpha_2-\alpha_1)}\right), \tag{3.24}$$

[1] It was shown in [58, Lemma III.22] that the map $\sigma \mapsto D_\alpha(\rho\|\sigma)$ is lower semi-continuous on $\mathcal{S}(\mathcal{H})$ for all $\alpha \in (0,1)$. The argument can be extended to the range $\alpha \in [0,1]$ by the same method in [58, Lemma III.22].

doi:10.6342/NTU201800597

where we denote by $\tilde{\lambda}_{\min}$ the smallest non-zero eigenvalue. Further, using [67, Problem III.6.14], we have

$$\tilde{\lambda}_{\min}\left(\rho^{\alpha_2-\alpha_1}\left(\sigma_2+\gamma\mathbb{1}\right)^{-(\alpha_2-\alpha_1)}\right) \geq \tilde{\lambda}_{\min}\left(\rho^{\alpha_2-\alpha_1}\right)\tilde{\lambda}_{\min}\left(\left(\sigma_2+\gamma\mathbb{1}\right)^{-(\alpha_2-\alpha_1)}\right) \tag{3.25}$$

$$\geq \left[\frac{\tilde{\lambda}_{\min}(\rho)}{1+\gamma}\right]^{\alpha_2-\alpha_1}. \tag{3.26}$$

Combining Eqs. (3.24) and (3.26) yields

$$\left\|\rho^0-\rho^{\alpha_2-\alpha_1}\left(\sigma_2+\gamma\mathbb{1}\right)^{-(\alpha_2-\alpha_1)}\right\|_\infty \leq 1-\left[\frac{\tilde{\lambda}_{\min}(\rho)}{1+\gamma}\right]^{\alpha_2-\alpha_1} \tag{3.27}$$

$$= (\alpha_2-\alpha_1)\left[\frac{\tilde{\lambda}_{\min}(\rho)}{1+\gamma}\right]+o(\alpha_2-\alpha_1) \tag{3.28}$$

$$\leq \varepsilon_2\left[\frac{\tilde{\lambda}_{\min}(\rho)}{1+\gamma}\right]+o(\varepsilon_2). \tag{3.29}$$

Hence, Eqs. (3.23) and (3.29) give

$$\left|Q_{\alpha_1}(\rho\|\sigma_2+\gamma\mathbb{1})-Q_{\alpha_2}(\rho\|\sigma_2+\gamma\mathbb{1})\right| \leq \varepsilon_2 d(1+\gamma)\left[\frac{\tilde{\lambda}_{\min}(\rho)}{1+\gamma}\right]+o(\varepsilon_2). \tag{3.30}$$

Next, we upper bound the second term in Eq. (3.18). Hölder's inequality given in Lemma 2.5 leads to

$$\left|\mathrm{Tr}\left[\rho^{\alpha_1}\left(\left(\sigma_1+\gamma\mathbb{1}\right)^{1-\alpha_1}-\left(\sigma_2+\gamma\mathbb{1}\right)^{1-\alpha_1}\right)\right]\right| \leq \|\rho^{\alpha_1}\|_1\left\|\left(\sigma_1+\gamma\mathbb{1}\right)^{1-\alpha_1}-\left(\sigma_2+\gamma\mathbb{1}\right)^{1-\alpha_1}\right\|_\infty \tag{3.31}$$

$$\leq d\left\|\left(\sigma_1+\gamma\mathbb{1}\right)^{1-\alpha_1}-\left(\sigma_2+\gamma\mathbb{1}\right)^{1-\alpha_1}\right\|_\infty. \tag{3.32}$$

Then, we apply Lemma 2.12 in Section 2.1 on Eq. (3.32) to obtain

$$\left|Q_{\alpha_1}(\rho\|\sigma_1+\gamma\mathbb{1})-Q_{\alpha_1}(\rho\|\sigma_2+\gamma\mathbb{1})\right| \leq d\left[\left(\varepsilon_1+\gamma\right)^{1-\alpha_1}-\gamma^{1-\alpha_1}\right]. \tag{3.33}$$

Eqs. (3.18), (3.24) and (3.33) thus give

$$\left|Q_{\alpha_1}(\rho\|\sigma_1+\gamma\mathbb{1})-Q_{\alpha_2}(\rho\|\sigma_2+\gamma\mathbb{1})\right| \leq \varepsilon_2\left[\frac{\tilde{\lambda}_{\min}(\rho)}{1+\gamma}\right]+d\left[\left(\varepsilon_1+\gamma\right)^{1-\alpha_1}-\gamma^{1-\alpha_1}\right]+o(\varepsilon_2). \tag{3.34}$$

This implies that, for any $\alpha_1 \in [0,1]$ the left-hand side becomes arbitrary small as $\varepsilon_1, \varepsilon_2 \to 0$, which concludes the continuity of $(\alpha,\sigma) \mapsto Q_\alpha(\rho\|\sigma+\gamma\mathbb{1})$. The assertion for D_α follow immediately. □

Let $\mathcal{X}=\{1,2,\ldots,|\mathcal{X}|\}$ be a finite alphabet, and let $\mathcal{P}(\mathcal{X})$ be the set of probability distributions on $\mathcal{X}$. Let $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ be a c-q channel. We denote a c-q state by:

$$P\circ\mathcal{W} := \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x|\otimes W_x. \tag{3.35}$$

We also express the input distribution $P\in\mathcal{P}(\mathcal{X})$ as a diagonalized matrix with respect to the compu-

doi:10.6342/NTU201800597

tational basis $(|x\rangle)_{x\in\mathcal{X}}$, i.e. $P=\sum_{x\in\mathcal{X}}P(x)|x\rangle\langle x|$.

We define the *conditional quantum relative entropy* of two sets of density operators $\bar{\mathcal{W}},\mathcal{W}$ and $P\in\mathcal{P}(\mathcal{X})$ as

$$D\left(\bar{\mathcal{W}}\|\mathcal{W}|P\right):=\sum_{x\in\mathcal{X}}P(x)D\left(\bar{W}_x\|W_x\right). \tag{3.36}$$

Similarly, we define the following conditional entropic quantities for $\sigma\in\mathcal{S}(\mathcal{H})$ and $P\in\mathcal{P}(\mathcal{X})$:

$$D\left(\mathcal{W}\|\sigma|P\right):=\sum_{x\in\mathcal{X}}P(x)D\left(W_x\|\sigma\right), \tag{3.37}$$

$$D_\alpha\left(\mathcal{W}\|\sigma|P\right):=\sum_{x\in\mathcal{X}}P(x)D_\alpha\left(W_x\|\sigma\right), \tag{3.38}$$

$$V\left(\mathcal{W}\|\sigma|P\right):=\sum_{x\in\mathcal{X}}P(x)V\left(W_x\|\sigma\right), \tag{3.39}$$

$$\widetilde{V}\left(\mathcal{W}\|\sigma|P\right):=\sum_{x\in\mathcal{X}}P(x)\widetilde{V}\left(W_x\|\sigma\right). \tag{3.40}$$

3.2 Conditional Rényi Entropy

For $\rho_{AB}\in\mathcal{S}(AB)$, $\alpha\geq 0$ and $t=\{\,\}$, or $\{\flat\}$, the *quantum conditional Rényi entropies* are given by

$$\begin{aligned}H_\alpha^{t,\uparrow}(A|B)_\rho&:=\sup_{\sigma_B\in\mathcal{S}(B)}-D_\alpha^t\left(\rho_{AB}\|\mathbb{1}_A\otimes\sigma_B\right),\\ H_\alpha^{t,\downarrow}(A|B)_\rho&:=-D_\alpha^t\left(\rho_{AB}\|\mathbb{1}_A\otimes\rho_B\right).\end{aligned} \tag{3.41}$$

In (3.41) When $\alpha=1$ and $t=\{\,\},\{\flat\}$, or $\{*\}$, both quantities coincide with the usual *quantum conditional entropy*:

$$H_1^{t,\uparrow}(A|B)_\rho=H_1^{t,\downarrow}(A|B)_\rho=H(A|B)_\rho:=H(AB)_\rho-H(B)_\rho, \tag{3.42}$$

where $H(A)_\rho:=-\operatorname{Tr}[\rho_A\log\rho_A]$ denotes the *von Neumann entropy* [5].

Proposition 3.1 (Properties of α-Rényi Conditional Entropy). *Given any classical-quantum state $\rho_{XB}\in\mathcal{S}(XB)$, the following holds:*

(a) The map $\alpha\mapsto H_\alpha^\uparrow(X|B)_\rho$ is continuous and monotonically decreasing on $[0,1]$.

(b) The map $\alpha\mapsto\frac{1-\alpha}{\alpha}H_\alpha^\uparrow(X|B)_\rho$ is concave on $(0,1)$.

Proof of Proposition 3.1.

(3.1)-(a) Fix an arbitrary sequence $(\alpha_k)_{k\in\mathbb{N}}$ such that $\alpha_k\in[0,1]$ and $\lim_{k\to+\infty}\alpha_k=\alpha_\infty\in[0,1]$. Let

$$\sigma_k^\star\in\operatorname*{arg\,min}_{\sigma\in\mathcal{S}(\mathcal{H})}D_{\alpha_k}\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma\right),\quad\forall k\in\mathbb{N}\cup\{+\infty\}. \tag{3.43}$$

doi:10.6342/NTU201800597

The definition in Eq. (3.41) implies that

$$\limsup_{k\to+\infty} H^{\uparrow}_{\alpha_k}(X|B)_\rho = -\liminf_{k\to+\infty} D_{\alpha_k}\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_k^\star\right) \tag{3.44}$$

$$\leq -D_{\alpha_\infty}\left(\rho_{XB}\left\|\mathbb{1}_X\otimes\left(\lim_{k\to+\infty}\sigma_k^\star\right)\right.\right) \tag{3.45}$$

$$\leq -\min_{\sigma\in\mathcal{S}(\mathcal{H})} D_{\alpha_\infty}\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma\right) \tag{3.46}$$

$$= H^{\uparrow}_{\alpha_\infty}(X|B)_\rho, \tag{3.47}$$

where, in order to establish (3.45), we used the lower semi-continuity of the map $\sigma\mapsto D_{\alpha_k}(\rho_{XB}\|\mathbb{1}_X\otimes\sigma)$ (see Eq. (3.15) in Lemma 3.2) and the continuity of $\alpha\mapsto D_\alpha\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_k^\star\right)$ (Eq. (3.13) in Lemma 3.2).

Next, we let

$$\sigma_k := (1-\varepsilon_k)\,\sigma_\infty^\star + \varepsilon_k\frac{\mathbb{1}}{d}, \quad \forall k\in\mathbb{N}, \tag{3.48}$$

where $(\varepsilon_k)_{k\in\mathbb{N}}$ is an arbitrary positive sequence that converges to zero. Then, it follows that

$$\liminf_{k\to+\infty} H^{\uparrow}_{\alpha_k}(X|B)_\rho \geq -\limsup_{k\to+\infty}\left\{D_{\alpha_k}\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_k\right)\right\} \tag{3.49}$$

$$= -D_{\alpha_\infty}\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_\infty^\star\right) \tag{3.50}$$

$$= H^{\uparrow}_{\alpha_\infty}(X|B)_\rho. \tag{3.51}$$

Here, equality (3.50) holds because $\mathbb{1}_X\otimes\sigma_k \gg \rho_{XB}$ for all $k\in\mathbb{N}\cup\{+\infty\}$. Thus, the map $(\alpha_k,\sigma_k)\mapsto D_{\alpha_k}(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_k)$ is continuous for $k\in\mathbb{N}\cup\{+\infty\}$. Hence, we prove the continuity.

Now, we show the monotonicity. For all $\sigma_B\in\mathcal{S}(B)$, Eq. (3.15) in Lemma 3.2 implies that $-D_\alpha(\rho_{XB}\|\mathbb{1}\otimes\sigma_B)$ is monotonically decreasing in $\alpha\geq 0$. Since $H^{\uparrow}_\alpha(X|B)_\rho$ is the pointwise supremum of the above function, we conclude that $H^{\uparrow}_\alpha(X|B)_\rho$ is monotonically decreasing in $\alpha\geq 0$. Hence, item (a) is proven.

(3.1)-(b) For convenience, we make a substitution $\alpha = 1/(1+s)$. The concavity for $s\geq 0$ can be proved with the geometric matrix means in [36]. Here, we present another proof by the following matrix inequality. Let $\rho_{XB}=\sum_{x\in\mathcal{X}}P(x)|x\rangle\langle x|\otimes W_x$, $t=\gamma=1$, $i=x$, $k=|\mathcal{X}|$, $A_i=P(x)W_x$, and $Z_i=I_{n,m}$. We obtain the log-convexity of the map by applying Lemma 2.13:

$$p\mapsto \mathrm{Tr}\left(\sum_{x\in\mathcal{X}}(P(x)W_x)^{\frac{1}{p}}\right)^p, \quad \forall p>0, \tag{3.52}$$

which is exactly the concavity of the map $s\mapsto sH^{\uparrow}_{1/(1+s)}(X|B)_\rho$ for all $s>0$.

□

doi:10.6342/NTU201800597

3.3 Rényi Mutual Information

The *mutual information* of a c-q channel $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$ with a prior distribution $P \in \mathcal{P}(\mathcal{X})$ is defined by

$$I(P, \mathcal{W}) := D\left(P \circ \mathcal{W} \| P \otimes P\mathcal{W}\right) = D\left(\mathcal{W} \| P\mathcal{W} | P\right), \tag{3.53}$$

where $P \circ \mathcal{W} := \sum_{x \in \mathcal{X}} P(x)|x\rangle\langle x| \otimes W_x$ and $P\mathcal{W} := \sum_{x \in \mathcal{X}} P(x) W_x$. Hence, the *information radius* or *information capacity*[2] of $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$ is

$$C_{\mathcal{W}} := \sup_{P \in \mathcal{P}(\mathcal{X})} I(P, \mathcal{W}). \tag{3.54}$$

The *conditional information variance* and the *unconditional information variance* of $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$ with a prior distribution $P \in \mathcal{P}(\mathcal{X})$ are defined, respectively, by

$$\begin{aligned} V(P, \mathcal{W}) &:= V\left(\mathcal{W} \| P\mathcal{W} | P\right), \\ U(P, \mathcal{W}) &:= V\left(P \circ \mathcal{W} \| P \otimes P\mathcal{W}\right). \end{aligned} \tag{3.55}$$

Note that $V(P^\star, \mathcal{W}) = U(P^\star, \mathcal{W})$ for every capacity-achieving distribution $P^\star \in \mathcal{P}(\mathcal{X})$, i.e. $I(P^\star, \mathcal{W}) = C_{\mathcal{W}}$, can be easily verified from the similar argument in [12, Lemma 62]. We also define the unconditional information variance in terms of $\widetilde{V}(\rho \| \sigma)$:

$$\widetilde{V}(P, \mathcal{W}) := \widetilde{V}\left(\mathcal{W} \| P\mathcal{W} | P\right). \tag{3.56}$$

The *minimal peripheral information variance* and its variant are defined by

$$V_{\mathcal{W}} := \inf_{P \in \mathcal{P}(\mathcal{X}):\, I(P, \mathcal{W}) = C_{\mathcal{W}}} V(P, \mathcal{W}), \tag{3.57}$$

$$\widetilde{V}_{\mathcal{W}} := \inf_{P \in \mathcal{P}(\mathcal{X}):\, I(P, \mathcal{W}) = C_{\mathcal{W}}} \widetilde{V}(P, \mathcal{W}). \tag{3.58}$$

Furthermore, one can easily verify that

$$V_{\mathcal{W}} > 0 \quad \text{implies} \quad C_{\mathcal{W}} > 0. \tag{3.59}$$

In the following, We define two related information quantities: for every $\alpha \in [0, 1]$,

$$I_\alpha^{(1)}(P, \mathcal{W}) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D_\alpha\left(P \circ \mathcal{W} \| P \otimes \sigma\right); \tag{3.60}$$

$$I_\alpha^{(2)}(P, \mathcal{W}) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D_\alpha\left(\mathcal{W} \| \sigma | P\right). \tag{3.61}$$

The term $I_\alpha^{(1)}(P, \mathcal{W})$ is called the *α-Rényi mutual information* [104, 64, 58, 105] or the *generalized Holevo quantity*. The second term $I_\alpha^{(2)}(P, \mathcal{W})$ can be viewed as a variant of the α-Rényi mutual information, called *α-Augustin mutual information* [106, 107]. It can be verified that these two functions

[2] We note that $C_{\mathcal{W}}$ equals to the capacity of classical communications over quantum channels [101, 102, 103]. It is usually term *classical capacity* [50], though it is a quantity in quantum information processing.

doi:10.6342/NTU201800597

are related by Jensen's inequality:

$$I_\alpha^{(1)}(P,\mathcal{W}) \le I_\alpha^{(2)}(P,\mathcal{W}). \tag{3.62}$$

For the case of $\alpha = 1$, they both equal conventional mutual information, i.e. $I_1^{(1)}(P,\mathcal{W}) = I_1^{(2)}(P,\mathcal{W}) = I(P,\mathcal{W})$. Mosonyi and Ogawa [58, Proposition IV.2] showed that for all $\alpha \in [0,1]$,

$$C_{\alpha,\mathcal{W}} := \sup_{P\in\mathcal{P}(\mathcal{X})} I_\alpha^{(1)}(P,\mathcal{W}) = \sup_{P\in\mathcal{P}(\mathcal{X})} I_\alpha^{(2)}(P,\mathcal{W}), \tag{3.63}$$

and it is termed the *Rényi radius* or the *Rényi capacity* of order α. Moreover, Proposition 3.2 below and the compactness of $\mathcal{P}(\mathcal{X})$ show that the suprema in Eq. (3.63) can be replaced with maxima.

We note that $I_\alpha^{(1)}$ admits a closed form for $\alpha \in (0,1]$ due to the quantum Sibson's identity below. The minimizer in Eq. (3.61) will be studied in Proposition 3.2.

Lemma 3.3 (Quantum Sibson's Identity [108]). *Fix an $\alpha \in (0,1]$. Let $\rho_{AB} \in \mathcal{S}(AB)$ and let $\sigma_B^\star$ be the minimizer of $\min_{\sigma_B\in\mathcal{S}(B)} D_\alpha(\rho_{AB}\|\rho_A\otimes\sigma_B)$. Then, one has*

$$\sigma^\star = \frac{\left(\mathrm{Tr}_A\left[\rho_{AB}^\alpha\right]\right)^{\frac{1}{\alpha}}}{\mathrm{Tr}\left[\left(\mathrm{Tr}_A\left[\rho_{AB}^\alpha\right]\right)^{\frac{1}{\alpha}}\right]}. \tag{3.64}$$

The following proposition presents important properties of α-Rényi mutual information and radius.

Proposition 3.2 (Properties of α-Mutual Information and Radius). *Given any classical-quantum channel $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$, the following holds:*

(a) For every $P \in \mathcal{P}(\mathcal{X})$, $\alpha \mapsto I_\alpha^{(2)}(P,\mathcal{W})$ is monotone increasing on $[0,1]$, and $I_\alpha^{(2)}(P,\mathcal{W}) \le \log\min\{|\mathcal{X}|, d\}$ for all $\alpha \in [0,1]$.

(b) The map $(\alpha,P) \mapsto I_\alpha^{(2)}(P,\mathcal{W})$ is continuous on $[0,1]\times\mathcal{P}(\mathcal{X})$.

(c) For every $(\alpha,P) \in (0,1]\times\mathcal{P}(\mathcal{X})$, there exists a unique $\sigma_{\alpha,P} \in \mathcal{S}(\mathcal{H})$ such that

$$I_\alpha^{(2)}(P,\mathcal{W}) = D_\alpha(\mathcal{W}\|\sigma_{\alpha,P}|P), \tag{3.65}$$

and

$$\mathsf{T}_{\alpha,P}(\sigma) = \sigma \text{ and } \sigma \gg P\mathcal{W} \text{ if and only if } \sigma = \sigma_{\alpha,P}, \tag{3.66}$$

where the map $\mathsf{T}_{\alpha,P}: \mathcal{S}_{P,\mathcal{W}}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ is defined as

$$\mathsf{T}_{\alpha,P}(\sigma) = \sum_{x\in\mathcal{X}} P(x)\frac{\sigma^{\frac{1-\alpha}{2}}W_x^\alpha\sigma^{\frac{1-\alpha}{2}}}{\mathrm{Tr}\left[W_x^\alpha\sigma^{1-\alpha}\right]}. \tag{3.67}$$

(d) The map $(\alpha,P) \mapsto \sigma_{\alpha,P}$ is continuous on $(0,1]\times\mathcal{P}(\mathcal{X})$.

(e) The map $\alpha \mapsto C_{\alpha,\mathcal{W}}$ is continuous and monotone increasing on $[0,1]$.

Proof of Proposition 3.2.

doi:10.6342/NTU201800597

(3.2)-(a) Recalling the definition of $I_\alpha^{(2)}$ given in Eq. (3.61). The statement immediately follows from Eq. (3.13) (see also [58, Lemma IV.5]) because the minimization over $\sigma \in \mathcal{S}(\mathcal{H})$ preserves the monotonicity. Hence, we have $I_\alpha^{(2)}(P,\mathcal{W}) \leq I_1(P,\mathcal{W}) \leq \log\min\{|\mathcal{X}|,d\}$, where the last inequality follows from the well-known upper bound for the Holevo quantity (see e.g. [5, Chapter 12]).

(3.2)-(b) Fix an arbitrary sequence $(\alpha_k,P_k)_{k\in\mathbb{N}}$ such that $\alpha_k \in [0,1]$, $P_k \in \mathcal{P}(\mathcal{X})$, and $\lim_{k\to+\infty}(\alpha_k,P_k) = (\alpha_0,P_0) \in [0,1]\times\mathcal{P}(\mathcal{X})$. Let

$$\sigma_k := \sigma_{\alpha_k,P_k} \in \operatorname*{arg\,min}_{\sigma\in\mathcal{S}(\mathcal{H})} D_{\alpha_k}\left(\mathcal{W}\|\sigma|P_k\right), \quad \forall k\in\mathbb{N}. \tag{3.68}$$

We first choose a subsequence $\{k_l\}_{l\in\mathbb{N}}$ such that

$$\liminf_{k\to+\infty} I_{\alpha_k}^{(2)}(P_k,\mathcal{W}) = \lim_{l\to+\infty} I_{\alpha_{k_l}}^{(2)}(P_{k_l},\mathcal{W}). \tag{3.69}$$

Since $\mathcal{S}(\mathcal{H})$ is compact[3], there exists a convergent subsubsequence $\{k_{l_m}\}_{m\in\mathbb{N}}$ such that $\lim_{m\to+\infty}\sigma_{k_{l_m}} = \sigma_0$ for some $\sigma_0\in\mathcal{S}(\mathcal{H})$. Then, we have

$$\begin{aligned}
\liminf_{k\to+\infty} I_{\alpha_k}^{(2)}(P_k,\mathcal{W}) &= \lim_{m\to+\infty} D_{\alpha_{k_{l_m}}}\left(\mathcal{W}\|\sigma_{k_{l_m}}|P_{k_{l_m}}\right) && (3.70)\\
&= \lim_{m\to+\infty} D_{\alpha_{k_{l_m}}}\left(\mathcal{W}\|\sigma_{k_{l_m}}|P_0\right) && (3.71)\\
&\quad + \lim_{m\to+\infty}\sum_{x\in\mathcal{X}}\left[P_{k_{l_m}}(x)-P_0(x)\right] D_{\alpha_{k_{l_m}}}\left(W_x\|\sigma_{k_{l_m}}\right) && (3.72)\\
&\geq \lim_{m\to+\infty} D_{\alpha_{k_{l_m}}}\left(\mathcal{W}\|\sigma_{k_{l_m}}|P_0\right) && (3.73)\\
&\geq D_{\alpha_0}\left(\mathcal{W}\,\|\,\sigma_0\,|\,P_0\right) && (3.74)\\
&\geq \min_{\sigma\in\mathcal{S}(\mathcal{H})} D_{\alpha_0}\left(\mathcal{W}\|\sigma|P_0\right) && (3.75)\\
&= I_{\alpha_0}^{(2)}(P_0,\mathcal{W}), && (3.76)
\end{aligned}$$

To see why inequality (3.73) holds, we observe that $\mathtt{supp}(P_0)\subseteq\mathtt{supp}(P_k)$ for all sufficiently large $k\in\mathbb{N}$ Further, the upper bound of $I_\alpha^{(2)}(P,\mathcal{W}) \leq \log\min\{|\mathcal{X}|,d\}$ (item (a)) implies that $D_{\alpha_k}(W_x\|\sigma_k) \leq \frac{\log\min\{|\mathcal{X}|,d\}}{P_k(x)}$ for all $x\in\mathtt{supp}(P_k)$. Hence, for $x\in\mathtt{supp}(P_0)$ and for all sufficiently large $m\in\mathbb{N}$, one has $P_{k_{l_m}}(x)\to P_0(x)$ and $D_{\alpha_{k_{l_m}}}(W_x\|\sigma_{k_{l_m}})$ is bounded away from $+\infty$. On the other hand, $P_{k_{l_m}}(x)-P_0(x)\geq 0$ for $x\notin\mathtt{supp}(P_0)$ and all sufficiently large $m\in\mathbb{N}$. In order to establish (3.74), we used the lower semi-continuity of the map $(\alpha,\sigma)\mapsto D_\alpha(W_x\|\sigma)$ for all $x\in\mathtt{supp}(P_0)$ in Eq. (3.17) in Lemma 3.2.

Next, we let

$$\widetilde{\sigma}_k := (1-\varepsilon_k)\,\sigma_{\alpha_0,P_0} + \varepsilon_k\frac{\mathbb{1}}{d}, \quad \forall k\in\mathbb{N}; \tag{3.77}$$

$$\varepsilon_k := \frac{\|P_k-P_0\|_1}{2}. \tag{3.78}$$

[3] Again, the compactness is with respect to the trace norm topology, we transit to the operator norm topology by the finite dimension of the Hilbert space.

doi:10.6342/NTU201800597

The definition of $I_\alpha^{(2)}$ yields

$$\limsup_{k\to+\infty} I_{\alpha_k}^{(2)}(P_k,\mathcal{W}) \leq \limsup_{k\to+\infty}\left\{D_{\alpha_k}\left(\mathcal{W}\|\widetilde{\sigma}_k|P_k\right)\right\} \tag{3.79}$$

$$= \limsup_{k\to+\infty}\left\{D_{\alpha_k}\left(\mathcal{W}\|\widetilde{\sigma}_k|P_\infty\right) + \sum_{x\in\mathcal{X}}\left[P_k(x)-P_0(x)\right]D_{\alpha_k}\left(W_x\|\widetilde{\sigma}_k\right)\right\} \tag{3.80}$$

$$\leq \limsup_{k\to+\infty}\left\{D_{\alpha_k}\left(\mathcal{W}\|\widetilde{\sigma}_k|P_0\right)\right\} + \limsup_{k\to+\infty}\left\{\sum_{x\in\mathcal{X}}\left[P_k(x)-P_0(x)\right]D_{\alpha_k}\left(W_x\|\widetilde{\sigma}_k\right)\right\}, \tag{3.81}$$

where equality (3.80) follows from the definition $I_\alpha^{(2)}$. Inequality (3.81) is due to the subadditivity of superior limits. Then, the convexity of $\sigma\mapsto D_{\alpha_k}(\mathcal{W}\|\sigma|P)$ implies that

$$\limsup_{k\to+\infty}\left\{D_{\alpha_k}\left(\mathcal{W}\|\widetilde{\sigma}_k|P_0\right)\right\} \leq \limsup_{k\to+\infty}\left\{(1-\varepsilon_k)D_{\alpha_k}\left(\mathcal{W}\|\sigma_{\alpha_0,P_0}|P_0\right)+\varepsilon_k\left[D_{\alpha_k}\left(\mathcal{W}\,\|\mathbb{1}/d|\,P_0\right)\right]\right\} \tag{3.82}$$

$$= D_{\alpha_0}\left(\mathcal{W}\|\sigma_{\alpha_0,P_0}|P_0\right) = I_{\alpha_0}^{(2)}(P_0,\mathcal{W}), \tag{3.83}$$

where the last line holds because of the continuity of $\alpha\mapsto D_\alpha(\cdot\|\cdot)$ on $[0,1]$ [58, Corollary III.13]and the finiteness of $D_{\alpha_k}\left(\mathcal{W}\,\|\mathbb{1}/d|\,P_0\right)$ for all $k\in\mathbb{N}$.

It remains to show the second term in Eq. (3.81) is actually zero. Direct calculation shows that

$$\limsup_{k\to+\infty}\left\{\sum_{x\in\mathcal{X}}\left[P_k(x)-P_0(x)\right]D_{\alpha_k}\left(W_x\|\widetilde{\sigma}_k\right)\right\} \tag{3.84}$$

$$\leq \limsup_{k\to+\infty}\left\{\varepsilon_k\cdot\max_{x\in\mathcal{X}}D_{\alpha_k}(W_x\|\widetilde{\sigma}_k)\right\} \tag{3.85}$$

$$\leq \limsup_{k\to+\infty}\left\{\varepsilon_k\cdot\max_{x\in\mathcal{X}}D_{\alpha_k}\left(W_x\middle\|\varepsilon_k\frac{\mathbb{1}}{d}\right)\right\} \tag{3.86}$$

$$= \limsup_{k\to+\infty}\left\{\varepsilon_k\cdot\left[\log\varepsilon_k+\max_{x\in\mathcal{X}}D_{\alpha_k}\left(W_x\middle\|\frac{\mathbb{1}}{d}\right)\right]\right\} \tag{3.87}$$

$$= \limsup_{k\to+\infty}\varepsilon_k\log\varepsilon_k \tag{3.88}$$

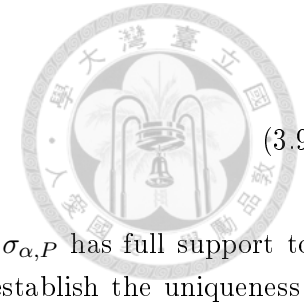
$$= 0, \tag{3.89}$$

where Eq. (3.86) follows from the dominance of α-Rényi divergence [8, Section 4]; equality (3.88) follows the finiteness of $D_\alpha\left(W_x\|\mathbb{1}/d\right)$ for all $x\in\mathcal{X}$ and $\alpha\in[0,1]$. in the last equality (3.89) we use the convention $\lim_{\varepsilon_k\downarrow 0}\varepsilon_k\log\varepsilon_k=0$ as $\varepsilon_k\to 0$ Hence, item (b) is proved.

(3.2)-(c) For $\alpha=1$, it is well-known that (see e.g. [101]) $\sigma_{1,P}=P\mathcal{W}$. Using the fact $P\mathcal{W}\gg W_x$ for all $x\in\mathtt{supp}(P)$, the statements are trivial.

We fix an arbitrary $(\alpha,P)\in(0,1)\times\mathcal{P}(\mathcal{X})$ subsequently. Without loss of generality, we may

doi:10.6342/NTU201800597



further assume

$$\bigcup_{x\in\mathtt{supp}(P)} \mathtt{supp}(W_x) = \mathbb{1}_{\mathcal{H}}, \tag{3.90}$$

and hence $P\mathcal{W}$ has full support. We first show that the minimizer $\sigma_{\alpha,P}$ has full support too. Second, we prove the fixed-point property Eq. (3.66). Finally, we establish the uniqueness of $\sigma_{\alpha,P}$. We remark that the uniqueness has been proven by Dalai and Winter [39, Appendix C]. Here, we provide an alternative proof for the completeness. Our approach follows closely from Hayashi and Tomamichel [104, Appendix C].

Define

$$\mathcal{M}_\alpha(\mathcal{H}) := \arg\min_{\sigma\in\mathcal{S}(\mathcal{H})} D_\alpha\left(\mathcal{W}\|\sigma|P\right) = \arg\max_{\sigma\in\mathcal{S}(\mathcal{H})} g_\alpha(\sigma) = \arg\max_{\sigma\in\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} g_\alpha(\sigma) \tag{3.91}$$

where

$$g_\alpha(\sigma) := \sum_{x\in\mathcal{X}} P(x)\log \mathrm{Tr}\left[W_x^\alpha \sigma^{1-\alpha}\right]. \tag{3.92}$$

To show that the optimizer of $g_\alpha(\cdot)$ has full support, we observe that the directional derivative on the boundary of $\mathcal{S}(\mathcal{H})$ where at least one eigenvalue is zero in a direction that increases its rank diverges to positive infinite. Namely, it suffices to show

$$\lim_{t\to 0} \frac{g_\alpha((1-t)\sigma + t\sigma^\perp) - g_\alpha(\sigma)}{t} = +\infty, \tag{3.93}$$

where $\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ is some singular density operator, and $\sigma^\perp := \frac{(\mathbb{1}_{\mathcal{H}}-\sigma)}{\mathrm{Tr}[\mathbb{1}_{\mathcal{H}}-\sigma]}$. For $x \in \mathtt{supp}(P)$ with $W_x \ll \sigma$, we have $W_x \perp \sigma^\perp$. It is not hard to see that

$$\lim_{t\to 0} P(x) \frac{\log \mathrm{Tr}\left[W_x^\alpha \left((1-t)\sigma + t\sigma^\perp\right)^{1-\alpha}\right] - \log \mathrm{Tr}\left[W_x^\alpha \sigma^{1-\alpha}\right]}{t} \tag{3.94}$$

$$= \lim_{t\to 0} P(x) \frac{\log \mathrm{Tr}\left[W_x^\alpha \left((1-t)^{1-\alpha}\sigma^{1-\alpha} + t^{1-\alpha}(\sigma^\perp)^{1-\alpha}\right)\right] - \log \mathrm{Tr}\left[W_x^\alpha \sigma^{1-\alpha}\right]}{t} \tag{3.95}$$

$$= \lim_{t\to 0} P(x) \frac{(1-\alpha)\log(1-t)}{t} \tag{3.96}$$

$$= \lim_{t\to 0} P(x) \frac{-(1-\alpha)}{1-t} \tag{3.97}$$

$$= -P(x)(1-\alpha) \tag{3.98}$$

$$> -\infty \tag{3.99}$$

where Eq. (3.95) holds because $\sigma \perp \sigma^\perp$; Eq. (3.96) is due to $W_x \perp \sigma^\perp$; and Eq. (3.97) is owing to L'Hôspital's rule.

On the other hand, since σ is singular, there must be some $x \in \mathtt{supp}(P)$ such that $W_x \not\ll \sigma$.

doi:10.6342/NTU201800597

Hence, by denoting $c := \frac{\mathrm{Tr}[W_x^\alpha(\sigma^\perp)^{1-\alpha}]}{\mathrm{Tr}[W_x^\alpha\sigma^{1-\alpha}]} > 0$, Eq. (3.95) leads to

$$\lim_{t\to 0} P(x)\frac{\log\left\{(1-t)^{1-\alpha}+t^{1-\alpha}c\right\}}{t} \tag{3.100}$$

$$= \lim_{t\to 0} P(x)\frac{-(1-\alpha)(1-t)^{-\alpha}+(1-\alpha)t^{-\alpha}c}{(1-t)^{1-\alpha}+t^{1-\alpha}c} \tag{3.101}$$

$$= +\infty, \tag{3.102}$$

where Eq. (3.101) is by L'Hôspital's rule again. Combining Eqs. (3.99) and (3.102) concludes Eq. (3.93).

Next, we show the fixed-point property: $\mathcal{M}_\alpha(\mathcal{H}) = \mathcal{F}_\alpha(\mathcal{H})$, where $\mathcal{F}_\alpha(\mathcal{H}) := \{\sigma \in \mathcal{S}_{>0}(\mathcal{H})\}$ denotes the fixed-points of the map: $\mathsf{T}_{\alpha,P} : \mathcal{S}_{P,\mathcal{W}}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$. A necessary and sufficient condition for σ to be an optimizer is

$$\partial_\omega g_\alpha(\sigma) := \mathsf{D}g_\alpha(\sigma)[\omega-\sigma] = 0, \tag{3.103}$$

for all $\omega \in \mathcal{S}(\mathcal{H})$, where $\mathsf{D}g_\alpha(\sigma)$ denotes the Fréchet derivative of the map g_α (see e.g. [104, Appendix C]). Using the chain rule of Fréchet derivatives, it follows

$$\partial_\omega g_\alpha(\sigma) = \mathrm{Tr}\left[\sum_{x\in\mathcal{X}} P(x)\frac{W_x^\alpha}{\mathrm{Tr}\left[W_x^\alpha\sigma^{1-\alpha}\right]}\partial_\omega\sigma^{1-\alpha}\right] \tag{3.104}$$

$$= \mathrm{Tr}\left[\sum_{x\in\mathcal{X}} P(x)\frac{\sigma^{\frac{-\alpha}{2}}W_x^\alpha\sigma^{\frac{-\alpha}{2}}}{\mathrm{Tr}\left[W_x^\alpha\sigma^{1-\alpha}\right]}\sigma^{\frac{\alpha}{2}}\partial_\omega\sigma^{1-\alpha}\sigma^{\frac{\alpha}{2}}\right]. \tag{3.105}$$

We claim that the operators

$$\left\{\Delta_\omega = \sigma^{\frac{\alpha}{2}}\sigma^{1-\alpha}\partial_\omega\sigma^{\frac{\alpha}{2}} : \omega \in \mathcal{S}(\mathcal{H})\right\} \tag{3.106}$$

span the space of traceless Hermitian operators on $\mathcal{S}(\mathcal{H})$. Let $\sigma = \sum_i \lambda_i|i\rangle\langle i|$ with $\lambda_i > 0$ be the eigenvalue decomposition. One can verify [82, Theorem 3.25] that

$$\langle i|\Delta_\omega|j\rangle = \begin{cases} (\lambda_i\lambda_j)^{\frac{\alpha}{2}}\frac{\lambda_i^{1-\alpha}-\lambda_j^{1-\alpha}}{\lambda_i-\lambda_j}\langle i|\omega-\sigma|j\rangle, & \text{if } \lambda_i \neq \lambda_j \\ (1-\alpha)\langle i|\omega-\sigma|j\rangle, & \text{if } \lambda_i = \lambda_j \end{cases}. \tag{3.107}$$

Therefore, Δ_ω is Hermitian and $\mathrm{Tr}\left[\Delta_\omega\right] = 0$ for all $\omega \in \mathcal{S}(\mathcal{H})$. Moreover, the basis of the traceless Hermitian operators is given by the operators

$$\left\{\Gamma_{ij} = |i\rangle\langle j| + |j\rangle\langle i|,\ \Gamma'_{ij} = \mathrm{i}|i\rangle\langle j| - |j\rangle\langle i|,\ \Gamma''_{ij} = |i\rangle\langle i| - |j\rangle\langle j|\right\}_{i\neq j}. \tag{3.108}$$

For every tuple (i,j) with $i \neq j$ there exists an $\varepsilon > 0$ such that the state $\omega = \sigma + \varepsilon\Gamma_{ij}$ is still in $\mathcal{S}(\mathcal{H})$. For this state, we find that $\Delta_\omega = \eta\Gamma_{ij}$ for some real $\eta > 0$. The similar argument applies to Γ'_{ij} and Γ''_{ij}. Hence, we have verified that the operators $\{\Delta_\omega\}_{\omega\in\mathcal{S}(\mathcal{H})}$ span the space of traceless Hermitian operators.

Armed with the above discussion, the condition that $\partial_\omega g_\alpha(\sigma) = 0$ for all $\omega \in \mathcal{S}(\mathcal{H})$ is equivalent

doi:10.6342/NTU201800597

to the condition that the operators

$$\sum_{x\in\mathcal{X}} P(x)\frac{\sigma^{\frac{-\alpha}{2}}W_x^\alpha\sigma^{\frac{-\alpha}{2}}}{\mathrm{Tr}\left[W_x^\alpha\sigma^{1-\alpha}\right]} \tag{3.109}$$

must be proportional to the identity. Thus, the optimum must be a fixed point of the map $\mathsf{T}_{\alpha,P}(\cdot)$.

Lastly, to prove the uniqueness of the optimizer, it remains to show $\partial_\omega^2 g_\alpha(\sigma) : \mathsf{D}^2 g_\alpha(\sigma)[\omega-\sigma,\omega-\sigma] < 0$ for all $\omega \neq \sigma$ and $\sigma > 0$. Continuing on Eq. (3.104), we have

$$\partial_\omega^2 g_\alpha(\sigma) = -\mathrm{Tr}\left[\sum_{x\in\mathcal{X}} P(x)\frac{W_x^\alpha}{\mathrm{Tr}^2\left[W_x^\alpha\sigma^{1-\alpha}\right]}\partial_\omega\sigma^{1-\alpha}\right] + \mathrm{Tr}\left[\sum_{x\in\mathcal{X}} P(x)\frac{W_x^\alpha}{\mathrm{Tr}\left[W_x^\alpha\sigma^{1-\alpha}\right]}\partial_\omega^2\sigma^{1-\alpha}\right] \tag{3.110}$$

$$< \mathrm{Tr}\left[\sum_{x\in\mathcal{X}} P(x)\frac{W_x^\alpha}{\mathrm{Tr}\left[W_x^\alpha\sigma^{1-\alpha}\right]}\partial_\omega^2\sigma^{1-\alpha}\right], \tag{3.111}$$

where Eq. (3.111) holds by noting that $\partial_\omega\sigma^{1-\alpha} \neq 0$ for all $\omega \neq \sigma$. Further, $\partial_\omega^2\sigma^{1-\alpha} \leq 0$ since $u \mapsto u^{1-\alpha}$ is operator concave. Thus, $\partial_\omega^2 g_\alpha(\sigma) < 0$, item (c) is proved.

(3.2)-(d) We follow the notation in item (d). However, we restrict $(\alpha_k, P_k)_{k\in\mathbb{N}}$ and (α_0, P_0) to be in the set $(0,1]\times\mathcal{P}(\mathcal{X})$. The continuity of $(\alpha, P) \mapsto I_\alpha^{(2)}(P,\mathcal{W})$ in item (b) and Eq. (3.74) thus imply

$$\lim_{k\to+\infty} I_{\alpha_k}^{(2)}(P_k,\mathcal{W}) = D_{\alpha_0}(\mathcal{W}\|\sigma_0|P_0) = I_{\alpha_0}^{(2)}(P_0,\mathcal{W}) = D_{\alpha_0}(\mathcal{W}\|\sigma_{\alpha_0,P_0}|P_0). \tag{3.112}$$

Then, the uniqueness of the minimizer $\sigma_{\alpha,P}$ in item (c) guarantees that $\sigma_0 = \sigma_{\alpha_0,P_0}$. Hence,

$$\lim_{k\to+\infty} \sigma_{\alpha_k,\sigma_k} = \sigma_0 = \sigma_{\alpha_0,\sigma_0}, \tag{3.113}$$

which proves item (d).

(3.2)-(e) Berge's maximum theorem [109, Section IV.3], [110, Lemma 3.1] shows that the continuous map $(\alpha, P) \mapsto I_\alpha^{(2)}(P,\mathcal{W})$ maximized over the compact set $P \in \mathcal{P}(\mathcal{X})$ is still continuous for $\alpha \in [0,1]$.

□

doi:10.6342/NTU201800597

Chapter 4

Quantum Hypothesis Testing

The goal of this chapter is to provide an introduction to quantum hypothesis testing. In Parts II and III later, our finite blocklength bounds heavily rely on the results in this chapter. In Sections 4.1 and 4.2 below, we present the error exponent analysis, while the moderate deviation analysis is given in Section 4.3.

The binary quantum hypothesis testing consists of a null hypothesis and an alternative hypothesis. The null hypothesis and the alternative hypothesis are described by the quantum states $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{S}(\mathcal{H})$, respectively. Given any test $0 \leq Q \leq \mathbb{1}$ that determines the outcome to be null hypothesis ρ, the *type-I error* and *type-II error* of the hypothesis testing are defined as follows:

$$\alpha\left(Q;\rho\right) := \mathrm{Tr}\left[(\mathbb{1}-Q)\rho\right], \tag{4.1}$$

$$\beta\left(Q;\sigma\right) := \mathrm{Tr}\left[Q\sigma\right]. \tag{4.2}$$

Unless $\rho \perp \sigma$, one cannot make both the type-I and type-II errors arbitrary small given the above definitions. Thus, we define the minimum type-I error when the type-II error is below $\mu \in (0,1)$ as

$$\widehat{\alpha}_{\mu}\left(\rho\|\sigma\right) := \min_{0\leq Q\leq \mathbb{1}} \left\{\alpha\left(Q;\rho\right) : \beta\left(Q;\sigma\right) \leq \mu\right\}. \tag{4.3}$$

The following famous quantum Stein's lemma characterizes the trade-off relation between these two errors. That is, the quantum relative entropy $D(\rho\|\sigma)$ serves as a benchmark to determine the asymptotic error behaviors of the optimal type-I error.

Theorem 4.1 (Quantum Stein's Lemma [95], [57], [86]). *Given a binary hypotheses:* $\mathsf{H}_0 : \rho$ *and* $\mathsf{H}_1 : \sigma$*, one has*

$$\lim_{n\to+\infty} \widehat{\alpha}_{\exp\{-nr\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) = \begin{cases} 0, & r < D(\rho\|\sigma) \\ 1, & r > D(\rho\|\sigma) \end{cases}. \tag{4.4}$$

For an n-shot independent extension of the binary hypothesis:

$$\mathsf{H}_0 : \rho^n = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n, \tag{4.5}$$

$$\mathsf{H}_1 : \sigma^n = \sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n, \tag{4.6}$$

doi:10.6342/NTU201800597

we define an error exponent function [86] by

$$\phi_n\left(r|\rho^n\|\sigma^n\right) := \sup_{\alpha\in(0,1]}\left\{\frac{\alpha-1}{\alpha}\left(r-\frac{1}{n}D_\alpha\left(\rho^n\|\sigma^n\right)\right)\right\}, \quad r\geq 0. \tag{4.7}$$

For the case $\rho^n \ll \sigma^n$, it is known that [86, Lemma 4]

$$\phi_n(r|\rho^n\|\sigma^n) = \begin{cases} +\infty, & r\in[0,-\frac{1}{n}\log \mathrm{Tr}\left[(\rho^n)^0\sigma^n\right]), \\ -\log \mathrm{Tr}\left[\rho^n(\sigma^n)^0\right], & r\geq \frac{1}{n}D\left(\rho^n\|\sigma^n\right). \end{cases} \tag{4.8}$$

In the following Sections 4.1 and 4.2, we show that the exponent function ϕ_n will determine how fast the optimal type-I error exponentially decay, i.e.

$$\lim_{n\to+\infty} -\frac{1}{n}\log \widehat{\alpha}_{\exp\{-nr\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) = \phi_1(r|\rho\|\sigma) = \sup_{0\leq\alpha\leq 1}\frac{1-\alpha}{\alpha}\left(D_\alpha(\rho\|\sigma)-r\right). \tag{4.9}$$

4.1 Achievability

Quantum Stein's lemma, given in Theorem 4.1, states that if the exponential decay of the type-II error is not faster than the relative entropy, i.e. $r < D(\rho\|\sigma)$, then the optimal type-I error vanishes asymptotically. The quantum Hoeffding bound makes a step further to investigate the non-asymptotics: how fast does the optimal type-I error decays? The achievability bound is then to give an exponential upper bound for it. This result was first proved by Hayashi [88], and the upper bound can be expressed as Petz's Rényi divergence. Together with the converse bound, discussed in Section 4.2 later, the error exponent for the optimal type-I error in quantum hypothesis testing was solved; see Eq. (4.9).

For the convenience of readers, we provide the proof of the achievability in Theorem 4.2 below.

Theorem 4.2 (Achievability Hoeffding Bound [88], [86, Section 5.5]). *Given a binary hypotheses:* $\mathsf{H}_0 : \rho$ *and* $\mathsf{H}_1 : \sigma$*, and rate* $r < D(\rho\|\sigma)$*, one has*

$$-\frac{1}{n}\log \widehat{\alpha}_{\exp\{-nr\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \geq \phi_1(r|\rho\|\sigma^n), \tag{4.10}$$

where ϕ_n *is defined in Eq.* (4.7).

Proof of Theorem 4.2. Fix an $n\in\mathbb{N}$, $\alpha\in(0,1)$, and let

$$A = \mathrm{e}^{-nx}\,\sigma^{\otimes n} \tag{4.11}$$

$$B = \rho^{\otimes n}, \tag{4.12}$$

where x will be determined later. Consider a sequence of test $\{(\mathbb{1}-Q_n, Q_n)\}$ with $Q_n := \{B-A\geq 0\}$.

doi:10.6342/NTU201800597

Then, Lemma 2.8 gives that

$$\begin{aligned}
\beta(Q_n;\sigma^{\otimes n}) &= \operatorname{Tr}\left[Q_n\sigma^{\otimes n}\right] && (4.13)\\
&= \mathrm{e}^{nx}\operatorname{Tr}\left[Q_nA\right] && (4.14)\\
&\leq \mathrm{e}^{nx\alpha}\,Q_\alpha\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) && (4.15)\\
&= \mathrm{e}^{nx\alpha}\,Q_\alpha\left(\rho\|\sigma\right)^n && (4.16)\\
\alpha(Q_n;\rho^{\otimes n}) &= \operatorname{Tr}\left[(\mathbb{1}-Q_n)\rho^{\otimes n}\right] && (4.17)\\
&= \mathrm{e}^{nx}\operatorname{Tr}\left[(\mathbb{1}-Q_n)B\right] && (4.18)\\
&\leq \mathrm{e}^{-nx(1-\alpha)}\,Q_\alpha\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) && (4.19)\\
&= \mathrm{e}^{-nx(1-\alpha)}\,Q_\alpha\left(\rho\|\sigma\right)^n. && (4.20)
\end{aligned}$$

Now, choose x such that $x\alpha+\log Q_\alpha(\rho\|\sigma)=-r$ to have

$$\beta(Q_n;\sigma^{\otimes n})\leq\exp\{-nr\}. \tag{4.21}$$

Further, it is not hard to see that

$$\alpha(Q_n;\rho^{\otimes n})\leq\exp\{-n\phi_1(r|\rho\|\sigma)\}. \tag{4.22}$$

□

4.2 Optimality

The optimality of the quantum Hoeffding bound means to provide a lower bound to the optimal type-I error. In other words, the performance of the hypothesis testing with any test cannot be improved. This problem was solved by Nagaoka [111]—he showed that asymptotically the error exponent of the optimal type-I error is upper bounded by $\Phi_1(\rho\|\sigma)$; see Theorem 4.3 below. Hence, together with the achievability bound in Theorem 4.2, the error exponent in Eq. (4.9) is fully characterized. The method employed by Nagaoka was introduced by Nussbaum and Szkoła [112], which is a crucial tool to translate a pair of quantum density operators to a pair of classical distributions. This thus plays a significant role in almost all the converse problems in quantum information theory. We provide the knowledge of the Nussbaum-Szkoła mapping in Section 4.2.1 below.

Theorem 4.3 (Asymptotic Converse Hoeffding Bound [111], [86, Section 5.4]). *Given a binary hypotheses:* $\mathsf{H}_0:\rho$ *and* $\mathsf{H}_1:\sigma$*, and rate* $r<D(\rho\|\sigma)$*, one has*

$$\lim_{n\to+\infty}-\frac{1}{n}\log\widehat{\alpha}_{\exp\{-nr\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)\leq\phi_1(r|\rho\|\sigma), \tag{4.23}$$

where ϕ_n *is defined in Eq.* (4.7).

Nagaoka's result in Theorem 4.3 is asymptotic, i.e. it holds when $n\to+\infty$. This motivates us to derive a finite blocklength converse bound. Moreover, we are interested in the tightest converse bound. In the following Theorem 4.4, we establish a sharp converse bound for quantum binary hypothesis testing, which serves as the fundamental tool to prove the sphere-packing bounds both in Slepian-Wolf

doi:10.6342/NTU201800597

coding with QSI (Chapter 7) and classical-quantum channel coding (Chapter 11), and the converse bounds in moderate deviation analysis (see Chapters 8 and 12).

Before stating Theorem 4.4, we introduce some notation. Let

$$\mathsf{H}_0 : \rho^n = \rho_1 \otimes \cdots \otimes \rho_n; \tag{4.24}$$

$$\mathsf{H}_1 : \sigma^n = \sigma_1 \otimes \cdots \otimes \sigma_n, \tag{4.25}$$

where $\rho_x, \sigma_x \in \mathcal{S}(\mathcal{H})$ for $x \in [n]$. Further, denote by (p_i, q_i) be the Nussbaum-Szkoła distribution of (ρ_i, σ_i) [112]. For $\alpha \in [0,1]$, define

$$B_\alpha(\rho^n \| \sigma^n) := \frac{1}{n} \sum_{x \in [n]} \mathbb{E}_{v_{x,\alpha}} \left[\log \frac{p_x}{q_x} \right]; \tag{4.26}$$

$$V_\alpha(\rho^n \| \sigma^n) := \frac{1}{n} \sum_{x \in [n]} \mathbb{E}_{v_{x,\alpha}} \left[\left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{\alpha,x}} \left[\log \frac{p_x}{q_x} \right] \right|^2 \right]; \tag{4.27}$$

$$T_\alpha(\rho^n \| \sigma^n) := \frac{1}{n} \sum_{x \in [n]} \mathbb{E}_{v_{x,\alpha}} \left[\left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{\alpha,x}} \left[\log \frac{p_x}{q_x} \right] \right|^3 \right], \tag{4.28}$$

where (p_x, q_x) is the Nussbaum-Szkoła distribution of (ρ_x, σ_x) for $x \in [n]$, and the *tilted distribution* is

$$v_{x,\alpha}(i,j) := \frac{p_x^\alpha(i,j) q_x^{1-\alpha}(i,j)}{\sum_{\imath,\jmath} p_x^\alpha(\imath,\jmath) q_x^{1-\alpha}(\imath,\jmath)}, \quad \alpha \in [0,1]. \tag{4.29}$$

With the above notation, we have the following converse bound.

Theorem 4.4 (Sharp Converse Bound for Quantum Hypothesis Testing)**.** *Consider a binary hypothesis testing:* $\mathsf{H}_0 : \rho^n = \bigotimes_{i=1}^n \rho_i$ *and* $\mathsf{H}_1 : \sigma^n = \bigotimes_{i=1}^n \sigma_i$ *given in Eq. (4.24) with* $\rho^n \not\perp \sigma^n$*. Let* $r \in \mathbb{R}$ *be such that there exists* $\alpha^\star \in (0,1)$ *such that*

$$\phi_n\left(r | \rho^n \| \sigma^n\right) = \frac{1-\alpha^\star}{\alpha^\star} \left(\frac{1}{n} D_{\alpha^\star}\left(\rho^n \| \sigma^n\right) - r \right). \tag{4.30}$$

Then, for any test Q_n*, either*

$$\alpha\left(Q^n; \rho^n\right) \geq \mathrm{e}^{-n\phi(r|\rho^n\|\sigma^n)} \frac{\mathrm{e}^{-K_n(\alpha)}}{2\sqrt{2\pi n V_{\alpha^\star}(\rho^n \| \sigma^n)}} \left(1 - \frac{1 + (1 + K_n(\alpha^\star)^2)}{2\sqrt{V_{\alpha^\star}(\rho^n \| \sigma^n)}} \right), \tag{4.31}$$

or

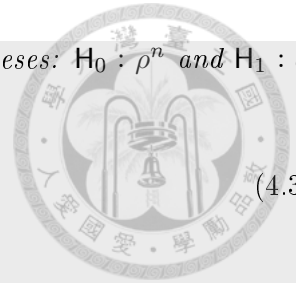
$$\beta\left(Q^n; \sigma^n\right) \geq \mathrm{e}^{-nr} \frac{\mathrm{e}^{-K_n(1-\eta)}}{2\sqrt{2\pi n V_{1-\alpha^\star}(\rho^n \| \sigma^n)}} \left(1 - \frac{1 + (1 + K_n(1-\alpha^\star)^2)}{2\sqrt{V_{1-\alpha^\star}(\rho^n \| \sigma^n)}} \right), \tag{4.32}$$

holds. Here, $K_n(\alpha) := \frac{15\sqrt{2\pi} T_\alpha(\rho^n \| \sigma^n)}{V_\alpha(\rho^n \| \sigma^n)}$.

The proof is delayed to Section 4.2.2.

With the Theorem 4.4 at hand, one can employ the Taylor's expansion of the ϕ_n to obtain the following sharp converse Hoeffding bound, which is the finite blocklength improvement of Nagaoka's result in Theorem 4.3.

doi:10.6342/NTU201800597



Corollary 4.1 (Sharp Converse Hoeffding Bound). *Given a binary hypotheses:* $\mathsf{H}_0 : \rho^n$ *and* $\mathsf{H}_1 : \sigma^n$ *as in Eq. (4.24), and rate:*

$$\frac{1}{n}D_0(\rho^n\|\sigma^n) < r < \frac{1}{n}D(\rho^n\|\sigma^n), \tag{4.33}$$

there exist $K, N_0 \in \mathbb{N}$ *such that for all* $n \geq N_0$*, the following holds*

$$-\log \widehat{\alpha}_{\exp\{-nr\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \leq n\phi_n(r|\rho^n\|\sigma^n) + \frac{1}{2}\left(1 + \left|\phi_n'(r|\rho^n\|\sigma^n)\right|\right)\log n + K. \tag{4.34}$$

where ϕ_n *is defined in Eq. (4.7), and* ϕ_n' *denotes the first-order derivative of* ϕ_n*.*

4.2.1 Nussbaum-Szkoła Distributions

Assume the dimension of the Hilbert space $\mathcal{H}$ is d. Given density operators $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ with spectral decompositions

$$\rho = \sum_{i\in[d]} \lambda_i |x_i\rangle\langle x_i|, \quad \text{and} \quad \sigma = \sum_{j\in[d]} \gamma_j |y_j\rangle\langle y_j|, \tag{4.35}$$

we define the *Nussbaum-Szkoła distributions* [112] $p^{\rho,\sigma}, q^{\rho,\sigma}$ as

$$p^{\rho,\sigma}(i,j) := \lambda_i |\langle x_i|y_j\rangle|^2, \quad q^{\rho,\sigma}(i,j) := \gamma_j |\langle x_i|y_j\rangle|^2. \tag{4.36}$$

The distributions $p^{\rho,\sigma}, q^{\rho,\sigma}$ have the same mathematical properties as the density operators ρ, σ in some cases, and thus are useful in the sequel. First, one can verify that [112, 14],

$$D_\alpha\left(\rho\|\sigma\right) = D_\alpha\left(p^{\rho,\sigma}\|q^{\rho,\sigma}\right), \quad \forall \alpha \in [0,1]. \tag{4.37}$$

Second, for product states $\rho_1 \otimes \rho_2$ and $\sigma_1 \otimes \sigma_2$, we have

$$p^{\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2} = p^{\rho_1,\sigma_1} \otimes p^{\rho_2,\sigma_2}, \quad \text{and} \quad q^{\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2} = q^{\rho_1,\sigma_1} \otimes q^{\rho_2,\sigma_2}. \tag{4.38}$$

Third, $\rho \ll \sigma$ if and only if $p^{\rho,\sigma} \ll q^{\rho,\sigma}$. Moreover, we usually use ω to represent the pair of indices (i,j) in Eq.(4.36), and the distributions $p^{\rho,\sigma}, q^{\rho,\sigma}$ can be thought of as diagonalized matrices, e.g. $\mathrm{Tr}\left[p^{\rho,\sigma}\right] = \sum_{\omega\in[d]\times[d]} p^{\rho,\sigma}(\omega)$.

4.2.2 Proofs of Theorem 4.4 and Corollary 4.1

Let $(\tilde{p}_i, \tilde{q}_i)$ be the Nussbaum-Szkoła distribution [112] of (ρ_i, σ_i) for every $i \in [n]$. Further, we define the (non-normalized) distributions $p_i := \tilde{p}_i q_i^0$, $q_i := \tilde{q}_i p_i^0$, for every $i \in [n]$, and $p^n := \bigotimes_{i=1}^n p_i$ and $q^n := \bigotimes_{i=1}^n q_i$ accordingly. Since $D_\alpha(\rho_i\|\sigma_i) = D_\alpha(\tilde{p}_i\|\tilde{q}_i) = D_\alpha(p_i\|q_i)$, for $\alpha \in (0,1)$, we shorthand

$$\phi_n(r) := \phi_n\left(r|\rho^n\|\sigma^n\right) = \phi_n(r|p^n\|q^n). \tag{4.39}$$

Applying Nagaoka's argument [111] in Eq. (11.57) for any $0 \leq Q_n \leq \mathbb{1}$ and choosing $\delta = \exp\{nr -$

doi:10.6342/NTU201800597

$n\phi_n(r)\}$ yields

$$\alpha\left(Q_n;\rho^n\right)+\delta\beta\left(Q_n;\sigma^n\right)\geq\frac{1}{2}\left(\alpha\left(\mathcal{U};p^n\right)+\mathrm{e}^{nr-n\phi_n(r)}\beta\left(\mathcal{U};q^n\right)\right), \tag{4.40}$$

where $\alpha\left(\mathcal{U};p^n\right):=\sum_{\omega\in\mathcal{U}^c}p^n(\omega)$, $\beta\left(\mathcal{U};q^n\right):=\sum_{\omega\in\mathcal{U}}q^n(\omega)$, and

$$\mathcal{U}:=\left\{\omega:p^n(\omega)\mathrm{e}^{n\phi_n(R_n)}>q^n(\omega)\mathrm{e}^{nR_n}\right\}. \tag{4.41}$$

In the following, we will employ Bahadur-Ranga Rao's concentration inequality, Theorem 2.1 in Section 2.2, to further lower bound $\alpha\left(\mathcal{U};p^n\right)$ and $\beta\left(\mathcal{U};q^n\right)$. Before proceeding, we need to introduce some notation. Let

$$\Lambda_{0,n}(\alpha):=\frac{1}{n}\sum_{x\in[n]}\log\mathbb{E}_{q_x}\left[\mathrm{e}^{\alpha\log\frac{p_x}{q_x}}\right],\quad \Lambda_{1,n}(\alpha):=\frac{1}{n}\sum_{x\in[n]}\log\mathbb{E}_{p_x}\left[\mathrm{e}^{\alpha\log\frac{q_x}{p_x}}\right]. \tag{4.42}$$

Since p^n and q^n share the same support, both $\Lambda_{0,n}(\alpha)$ and $\Lambda_{1,n}(\alpha)$ are smooth functions in $\alpha\in\mathbb{R}$. One can the calculate derivatives as follows

$$\Lambda'_{0,n}(\alpha)=\frac{1}{n}\sum_{x\in[n]}\mathbb{E}_{v_{x,\alpha}}\left[\log\frac{p^n}{q^n}\right];\quad \Lambda'_{1,n}(\alpha)=\frac{1}{n}\sum_{x\in[n]}\mathbb{E}_{v_{x,1-\alpha}}\left[\log\frac{q^n}{p^n}\right] \tag{4.43}$$

$$\Lambda''_{0,n}(\alpha)=\frac{1}{n}\sum_{x\in[n]}\mathrm{Var}_{v_{x,\alpha}}\left[\log\frac{p^n}{q^n}\right];\quad \Lambda''_{1,n}(\alpha)=\frac{1}{n}\sum_{x\in[n]}\mathrm{Var}_{v_{x,1-\alpha}}\left[\log\frac{q^n}{p^n}\right], \tag{4.44}$$

$$T_{0,n}(\alpha):=\frac{1}{n}\sum_{x\in[n]}\mathbb{E}_{v_{x,\alpha}}\left[\left|\log\frac{q_x}{p_x}-\Lambda'_{0,n}(\alpha)\right|^3\right]; \tag{4.45}$$

$$T_{1,n}(\alpha):=\frac{1}{n}\sum_{x\in[n]}\mathbb{E}_{v_{x,1-\alpha}}\left[\left|\log\frac{p_x}{q_x}-\Lambda'_{1,n}(\alpha)\right|^3\right], \tag{4.46}$$

where we denote the tilted distribution by

$$\hat{q}^n_\alpha:=\frac{(p^n)^\alpha(q^n)^{1-\alpha}}{\sum_\omega p^n(\omega)^\alpha q^n(\omega)^{1-\alpha}}. \tag{4.47}$$

Further, we define the *Lengendre-Fenchel transform*:

$$\Lambda^*_{j,n}(z):=\sup_{\alpha\in\mathbb{R}}\left\{(1-\alpha)z-\Lambda_{j,n}(\alpha)\right\},\quad j\in\{0,1\}. \tag{4.48}$$

The quantities $\Lambda^*_{j,n}(z)$ would appear in the lower bounds of $\alpha\left(\mathcal{U};p^n\right)$ and $\beta\left(\mathcal{U};q^n\right)$ obtained by Bahadur-Randga Rao's inequality as shown later.

Now, we are ready to derive the lower bounds for $\alpha\left(\mathcal{U};p^n\right)$ and $\beta\left(\mathcal{U};q^n\right)$. Letting $Z_i=\log p_i-\log q_i$ with probability measure $\mu_i=q_i$, and $z=r-\phi_n(r)$ in Theorem 2.1, the Bahadur-Randga Rao's

doi:10.6342/NTU201800597

inequality gives

$$\alpha\left(\mathcal{U};p^n\right) := \sum_{\omega\in\mathcal{U}^c} p^n(\omega) \tag{4.49}$$

$$= \Pr\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i \geq r-\phi_n(r)\right\} \tag{4.50}$$

$$\geq \exp\left\{-n\Lambda_{0,n}^*\left(\phi_n(r)-r\right)\right\}\frac{\mathrm{e}^{-K_n(\alpha)}}{\sqrt{2\pi\Lambda_n''(\alpha)}}\left(1-\frac{1+(1+K_n(\alpha)^2)}{2\sqrt{\Lambda_n''(\alpha)}}\right) \tag{4.51}$$

$$= \exp\left\{-n\phi_n(r)\right\}\frac{\mathrm{e}^{-K_n(\alpha)}}{\sqrt{2\pi\Lambda_n''(\alpha)}}\left(1-\frac{1+(1+K_n(\alpha)^2)}{2\sqrt{\Lambda_n''(\alpha)}}\right) \tag{4.52}$$

Similarly, applying Theorem 2.1 with $Z_i = \log q_i - \log p_i$, $\mu_i = p_i$, and $z = \phi_n(r) - r$ yields

$$\beta\left(\mathcal{U};q^n\right) := \sum_{\omega\in\mathcal{U}} q^n(\omega) \tag{4.53}$$

$$= \Pr\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i \geq \phi_n(r)-r\right\} \tag{4.54}$$

$$\geq \exp\left\{-n\Lambda_{1,n}^*\left(r-\phi_n(r)\right)\right\}\frac{\mathrm{e}^{-K_n(1-\alpha)}}{\sqrt{2\pi\Lambda_n''(1-\alpha)}}\left(1-\frac{1+(1+K_n(1-\alpha)^2)}{2\sqrt{\Lambda_n''(1-\alpha)}}\right) \tag{4.55}$$

$$= \exp\left\{-nr\right\}\frac{\mathrm{e}^{-K_n(1-\alpha)}}{\sqrt{2\pi\Lambda_n''(1-\alpha)}}\left(1-\frac{1+(1+K_n(1-\alpha)^2)}{2\sqrt{\Lambda_n''(1-\alpha)}}\right). \tag{4.56}$$

Hence, by Eqs. (4.40), (4.52), and (4.56), we conclude our claim. □

4.3 Moderate Deviation Analysis

In this section, we analyze quantum hypothesis testing in the moderate deviation regime. Specifically, we will show that the optimal type-I error asymptotically vanish when the exponential rate of type-II error approaches quantum relative entropy at a speed a_n. Here, $(a_n)_{n\in\mathbb{N}}$ is any sequence satisfying

$$\begin{aligned} &\text{(i)}\ \lim_{n\to+\infty} a_n = 0; \\ &\text{(ii)}\ \lim_{n\to+\infty} a_n\sqrt{n} = +\infty. \end{aligned} \tag{4.57}$$

The achievability part is given in Theorem 4.5. In Section 4.3.1, we provide two proofs. The first one follows from the Theorem 4.2 in Section 4.1, and an asymptotic expansions of the exponent function ϕ_n. The second proof relies on a concentration inequality for noncommutative martingales [113]. The converse part and its proof are provided in Theorem 4.6 and Section 4.3.2.

We remark that the moderate deviation analysis for classical hypothesis testing was studied by Sason [45], and by Watanabe and Hayashi [114]. Moreover, a recent work by Rouzé and Datta [115] formulated the quantum hypothesis problem into a martingale, which is similar to our approach for proving the achievability.

doi:10.6342/NTU201800597

Theorem 4.5 (Achievability). *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be the density operators with non-zero and finite information variance $V := \mathbb{V}(\rho\|\sigma) > 0$. For any sequence of real numbers $\{a_n\}_{n\in\mathbb{N}}$ satisfying Eq. (12.1), there exists a sequence $r_n := \mathbb{D}(\rho\|\sigma) - a_n$ such that*

$$\limsup_{n\to+\infty} \frac{1}{na_n^2} \log \widehat{\alpha}_{\exp\{-nr_n\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \leq -\frac{1}{2V}. \tag{4.58}$$

The proof is provided in Section 4.3.1

Theorem 4.6 (Converse). *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be the density operators with non-zero and finite information variance $V := \mathbb{V}(\rho\|\sigma) > 0$. For any sequence of real numbers $\{a_n\}_{n\in\mathbb{N}}$ satisfying Eq. (12.1), there exists a sequence $r_n := \mathbb{D}(\rho\|\sigma) - a_n$ such that*

$$\liminf_{n\to+\infty} \frac{1}{na_n^2} \log \widehat{\alpha}_{\exp\{-nr_n\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \geq -\frac{1}{2V}. \tag{4.59}$$

The proof is provided in Section 4.3.2

4.3.1 Proof of Theorem 4.5

In this section, we present two proofs of Theorem 4.5. The first one relies on the quantum Hoeffding bound [86] and the Taylor's expansion of the function E_{h}.

The first proof of Theorem 4.5. We start the proof from recalling Audenaet et al.'s achievabilityof the quantum Hoeffding bound in Lemma 2.8:

$$\widehat{\alpha}_{\exp\{-nr\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \leq \exp\left\{-n\left[\sup_{0<\alpha\leq 1}\left\{\frac{\alpha-1}{\alpha}\left(r - D_\alpha\left(\rho\|\sigma\right)\right)\right\}\right]\right\}. \tag{4.60}$$

Since $\mathbb{D}(\rho\|\sigma) > 0$ (due to Eq. (3.4)), we have

$$\mathbb{D}(\rho\|\sigma) - a_n > 0 \tag{4.61}$$

for all sufficiently large n. Choose such n onwards. Then Eq. (4.60) implies that for all sufficiently large n, there exists $r_n = D(\rho\|\sigma) - a_n$ and

$$\begin{aligned}\frac{1}{na_n^2} \log \widehat{\alpha}\left(\rho^{\otimes n}\|\sigma^{\otimes}\right) &\leq \frac{1}{na_n^2} - \frac{1}{a_n^2}\sup_{0<\alpha\leq 1}\left\{\frac{\alpha-1}{\alpha}\left(r - D_\alpha\left(\rho\|\sigma\right)\right)\right\} &(4.62)\\ &= \frac{1}{na_n^2} - \frac{1}{a_n^2}\sup_{s\geq 0}\left\{E_{\mathrm{s}}(s) - sr_n\right\}, &(4.63)\end{aligned}$$

where we substitute $s = \frac{1-\alpha}{\alpha}$ and invoke Eq. (9.7):

$$E_{\mathrm{h}}(s) := E_{\mathrm{h}}(s, P) = sD_{\frac{1}{1+s}}\left(\rho\|\sigma\right). \tag{4.64}$$

with $\mathcal{X} = \{x\}$ and $W_x = \rho$.

Therefore, we apply Taylor's theorem, along with items (c) and (e) in Proposition 9.3, to obtain

$$E_{\mathrm{h}}(s) = s\mathbb{D}(\rho\|\sigma) - \frac{s^2}{2}V + \frac{s^3}{6}\frac{\partial^3 E_{\mathrm{h}}(s)}{\partial s^3}\bigg|_{s=\bar{s}} \tag{4.65}$$

doi:10.6342/NTU201800597

for some $\bar{s} \in [0, s]$ and all $s \geq 0$. Now let $s_n = a_n/V$, for all $n \in \mathbb{N}$. Then for all sufficiently large n and for some $\bar{s}_n \in [0, s_n]$, Eq. (4.65) yields

$$\sup_{s\geq 0} \{E_{\mathrm{h}}(s) - sr_n\} \geq E_{\mathrm{h}}(s_n) - s_n r_n \tag{4.66}$$

$$= \frac{a_n}{V} \left(\mathbb{D}(\rho\|\sigma) - r_n\right) - \frac{a_n^2}{2V} + \frac{a_n^3}{6V^2} \frac{\partial^3 E_{\mathrm{h}}(s)}{\partial s^3}\bigg|_{s=\bar{s}_n} \tag{4.67}$$

$$= \frac{a_n^2}{2V} + \frac{a_n^3}{6V^2} \frac{\partial^3 E_{\mathrm{h}}(s)}{\partial s^3}\bigg|_{s=\bar{s}_n}, \tag{4.68}$$

where we substitute $r_n = D(\rho\|\sigma) - a_n$ in Eq. (4.68).

Note that $s_n = a_n/V \leq 1$ for all sufficiently large n since $\lim_{n\to\infty} a_n = 0$ in Eq. (12.1) and the assumption: $V > 0$. Define

$$\Upsilon := \max_{s\in[0,1]} \left| \frac{\partial^3 E_{\mathrm{h}}(s)}{\partial s^3} \right|, \tag{4.69}$$

From item (a) in Proposition 9.3, $\frac{\partial^3 E_{\mathrm{h}}(s)}{\partial s^3}$ is continuous over $s \geq 0$. Hence the maximum in Eq. (4.69) is well-defined and finite. Therefore, (4.68) leads to

$$\sup \{E_{\mathrm{h}}(s) - sr_n\} \geq \frac{a_n^2}{2V} + \frac{a_n^3}{6V^2} \frac{\partial^3 E_{\mathrm{h}}(s)}{\partial s^3}\bigg|_{s=\bar{s}_n} \tag{4.70}$$

$$\geq \frac{a_n^2}{2V} - \frac{a_n^3}{6V^2} \left| \frac{\partial^3 E_{\mathrm{h}}(s)}{\partial s^3}\bigg|_{s=\bar{s}_n} \right| \tag{4.71}$$

$$\geq \frac{a_n^2}{2V} - \frac{a_n^3}{6V^2} \Upsilon \tag{4.72}$$

for all sufficiently large n.

Substituting Eq. (4.72) into Eq. (4.63) yields

$$\frac{1}{na_n^2} \log \widehat{\alpha}_{\exp\{-nr_n\}} (\rho\|\sigma) \leq \frac{1}{na_n^2} - \frac{1}{2V} \left(1 - \Upsilon \frac{a_n}{3V^2}\right), \tag{4.73}$$

which implies the desired achievability part:

$$\limsup_{n\to+\infty} \frac{1}{na_n^2} \log \widehat{\alpha}_{\exp\{-nr_n\}} (\rho\|\sigma) \leq -\frac{1}{2V}. \tag{4.74}$$

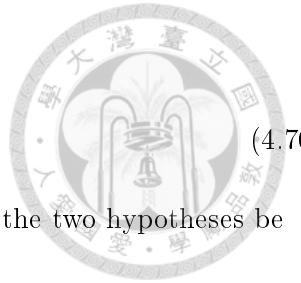
□

In the following, we give an alternative proof of Theorem 4.5 by employing a noncommutative Bennett inequality [113].

The second proof of Theorem 4.5. It is well-known that the Neyman-Pearson (likelihood-ratio) test achieves the optimum type-I error with the constraint of the type-II error. Hence, it suffices to prove that

$$\limsup_{n\to+\infty} \frac{1}{na_n^2} \log \widehat{\alpha}_{\exp\{-nr_n\}} (\rho\|\sigma) = \lim_{n\to\infty} \frac{1}{na_n^2} \log \alpha_n (\eta_n) \geq -\frac{1}{2V}, \tag{4.75}$$

doi:10.6342/NTU201800597



where

$$\eta_n := \mathbb{D}\left(\rho\|\sigma\right) - a_n, \quad n \in \mathbb{N}. \tag{4.76}$$

For notational convenience, we first consider the non-identical case. Let the two hypotheses be

$$\mathsf{H}_0 : \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n \tag{4.77}$$

$$\mathsf{H}_1 : \sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n, \tag{4.78}$$

where $\rho_i, \sigma_i \in \mathcal{S}\left(\mathcal{H}_i\right)$, for every $i \in [n]$. Define the operator

$$L_n := \log \bigotimes_{i=1}^{n} \rho_i - \log \bigotimes_{i=1}^{n} \sigma_i = \sum_{i=1}^{k} \left(\log \rho_i - \log \sigma_i\right), \tag{4.79}$$

which can be seen as the quantum generalization of the Neyman-Pearson log-likelihood ratio.

Next, we formulate the hypothesis testing problem in the noncommutative probability space [116, 117]. Let $\mathfrak{M}_k$ be the von Neumann algebra on the Hilbert space $\bigotimes_{i=1}^{k} \mathcal{H}_k$ with $\mathfrak{M}_0 = \varnothing$, and $(\mathfrak{M}_k)_{k=0}^{n}$ forms a increasing filtration (see e.g. [118]). The normal faithful tracial stae $\tau : \mathfrak{M}_n \to \mathbb{C}$ on $\mathfrak{M}_n$ is defined as $\tau : X \mapsto \mathrm{Tr}\left[\bigotimes_{j=1}^{n} \rho_j X\right]$. Let $\mathbb{E}_{\bigotimes_{j=1}^{n} \rho_j}\left[\,\cdot\,|\mathfrak{M}_k\right] : \mathfrak{M}_n \to \mathfrak{M}_k$ be the conditional expectation of $\mathfrak{M}_n$ with respect to $\mathfrak{M}_k$. For every $k \in \{0, 1, \ldots, n\}$, we let

$$U_k := \mathbb{E}_{\bigotimes_{j=1}^{n} \rho_j}\left[\, L_n \middle|\, \mathfrak{M}_k\right] \tag{4.80}$$

$$= \mathbb{E}_{\bigotimes_{j=1}^{n} \rho_j}\left[\sum_{i=1}^{n} \left(\log \rho_i - \log \sigma_i\right) \middle|\, \mathfrak{M}_k\right] \tag{4.81}$$

$$= \sum_{i=1}^{k} \left(\log \rho_i - \log \sigma_i\right) + \sum_{i=k+1}^{n} \mathbb{E}_{\bigotimes_{j=1}^{n} \rho_j}\left[\log \rho_i - \log \sigma_i\right] \tag{4.82}$$

$$= \sum_{i=1}^{k} \left(\log \rho_i - \log \sigma_i\right) + \sum_{i=k+1}^{n} \mathrm{Tr}\left[\bigotimes_{j=1}^{n} \rho_j \left(\log \rho_i - \log \sigma_i\right)\right] \tag{4.83}$$

$$= \sum_{i=1}^{k} \left(\log \rho_i - \log \sigma_i\right) + \sum_{i=k+1}^{n} \mathbb{D}\left(\rho_i\|\sigma_i\right). \tag{4.84}$$

In particular, we have

$$U_0 = \sum_{i=1}^{n} \mathbb{D}\left(\rho_i\|\sigma_i\right) \tag{4.85}$$

$$U_n = \sum_{i=1}^{n} \left(\log \rho_i - \log \sigma_i\right) = L_n, \tag{4.86}$$

doi:10.6342/NTU201800597

Hence, $\{U_k - U_{k-1}\}_{k=1}^n$ forms a martingale:

$$U_k - U_{k-1} = \log \rho_k - \log \sigma_k - \mathbb{D}(\rho_k \| \sigma_k); \tag{4.87}$$

$$\mathbb{E}_{\bigotimes_{j=1}^n \rho_j} \left[U_k - U_{k-1} \middle| \mathfrak{M}_{k-1} \right] = 0; \tag{4.88}$$

$$\mathbb{E}_{\bigotimes_{j=1}^n \rho_j} \left[(U_k - U_{k-1})^2 \middle| \mathfrak{M}_{k-1} \right] = \mathbb{V}\left(\rho_k \| \sigma_k\right) =: v_k. \tag{4.89}$$

Denote by

$$b_k := \left\| \log \rho_k - \log \sigma_k - \mathbb{D}(\rho_k \| \sigma_k) \right\|_\infty, \tag{4.90}$$

where $\| \cdot \|_\infty$ denotes the operator norm. The martingale is bounded by $\|U_k - U_{k-1}\|_\infty \leq b_k$ for every $k \in [n]$.

Equipped with the notation above, the type-I error can be rephrased as:

$$\alpha_n\left(\eta_n\right) = \operatorname{Tr}\left[\left\{ \bigotimes_{i=1}^n \rho_i - \mathrm{e}^{n\eta_n} \bigotimes_{i=1}^n \sigma_i \leq 0 \right\} \bigotimes_{i=1}^n \rho_i \right] \tag{4.91}$$

$$= \operatorname{Tr}\left[\bigotimes_{i=1}^n \rho_i \left\{ \sum_{i=1}^n \left(\log \rho_i - \log \sigma_i \right) \leq n\eta_n \right\} \right] \tag{4.92}$$

$$= \operatorname{Tr}\left[\bigotimes_{i=1}^n \rho_i \left\{ U_n - U_0 \leq -n a_n \right\} \right] \tag{4.93}$$

$$= \tau\left(\mathbf{1}_{(-\infty, -n a_n)} \left(U_n - U_0 \right) \right) \tag{4.94}$$

$$= \tau\left(\mathbf{1}_{(n a_n, \infty)} \left(U_n - U_0 \right) \right), \tag{4.95}$$

where the third equality (4.93) follows from the definition of η_n in Eq. (4.76) and Eqs. (4.85) and (4.86). The last line (4.95) is due to the symmetry of $U_n - U_0$, i.e. $\mathbb{E}_{\bigotimes_{i=1}^n} \left[U_n - U_0 \right] = 0$.

In the following, we borrow the idea from Sason [45] to employ the noncommutative Bennett inequality to upper bound Eq. (4.95).

Theorem 4.7 (Noncommutative Bennett Inequality [113, Theorem 0.1]). *Let $(X_k)_{k=1}^n$ be a self-adjoint martingale with respect to the filtration $(\mathfrak{M}_k)_{k=0}^n$ such that:* (i) $\mathbb{E}\left[X_k | \mathfrak{M}_{k-1}\right] = 0$; (ii) $\mathbb{E}\left[X_k^2 | \mathfrak{M}_{k-1}\right] = v_k$; (iii) $\|X_k\|_\infty \leq b_k$. *Then for any $x > 0$,*

$$\tau\left(\mathbf{1}_{[x,\infty)} \left(\sum_{k=1}^n X_k \right) \right) \leq \exp\left\{ -\frac{\sum_{k=1}^n v_k}{\sup_{k\in[n]} b_k^2} \varphi\left(\frac{x \sup_{k\in[n]} b_k}{\sum_{k=1}^n v_k} \right) \right\}, \tag{4.96}$$

where $\varphi(u) := (1+u)\log(1+u) - u$.

By applying Theorem 4.7 to Eq. (4.95) with $x = na$ and $X_k = U_k - U_{k-1}$ for ever $k \in [n]$:

$$\alpha_n\left(\eta_n\right) \leq \exp\left\{ -\frac{\sum_{k=1}^n v_k}{\sup_{k\in[n]} b_k^2} \varphi\left(\frac{n a_n \sup_{k\in[n]} b_k}{\sum_{k=1}^n v_k} \right) \right\} \tag{4.97}$$

$$= \exp\left\{ -\frac{n\bar{v}}{B^2} \varphi\left(\frac{a_n b}{\bar{v}} \right) \right\}, \tag{4.98}$$

doi:10.6342/NTU201800597

where

$$b := \sup_{k\in[n]} b_k, \quad B^2 := \sup_{k\in[n]} b_k^2, \quad \bar{v} := \frac{\sum_{k=1}^n v_k}{n}. \tag{4.99}$$

By recalling $\varphi(u) = (1+u)\log(1+u) - u$ and using a scalar inequality [45, Lemma 1]:

$$(1+u)\log(1+u) \geq u + \frac{u^2}{2} - \frac{u^3}{6}, \quad u \geq 0, \tag{4.100}$$

Eq. (4.98) leads to

$$\alpha_n(\eta_n) = \mathrm{Tr}\left[\left\{\bigotimes_{i=1}^n \rho_i - \mathrm{e}^{n\eta_n}\bigotimes_{i=1}^n \sigma_i \leq 0\right\}\bigotimes_{i=1}^n \rho_i\right] \tag{4.101}$$

$$\leq \exp\left\{-\frac{n\bar{v}}{B^2}\left[\frac{(a_n b)^2}{2\bar{v}^2} - \frac{(a_n b)^3}{6\bar{v}^3}\right]\right\} \tag{4.102}$$

$$= \exp\left\{-n\left[\frac{a_n^2 b^2}{2\bar{v}B^2}\left(1 - \frac{a_n b}{3\bar{v}^2}\right)\right]\right\}. \tag{4.103}$$

Now considering the identical case:

$$\bigotimes_{i=1}^n \rho_i = \rho^{\otimes n} \in \mathcal{S}\left(\mathcal{H}^{\otimes n}\right), \quad \text{and} \quad \bigotimes_{i=1}^n \sigma_i = \sigma^{\otimes n} \in \mathcal{S}\left(\mathcal{H}^{\otimes n}\right) \tag{4.104}$$

with $\rho \ll \sigma$ (otherwise $\alpha_n(\eta_n) = 0$ and Eq. (4.75) holds trivially), we have

$$\bar{v} = V, \tag{4.105}$$

$$b = B = \left\|\log\rho - \log\sigma - \mathbb{D}(\rho\|\sigma)\right\|_\infty < \infty, \tag{4.106}$$

where the finiteness of b comes from $\rho \ll \sigma$ and the assumption that the Hilbert space $\mathcal{H}$ is finite-dimensional. From Eq. (4.103), the type-I error is upper bounded by

$$\alpha_n(\eta_n) \leq \exp\left\{-n\left[\frac{a_n^2}{2V}\left(1 - \frac{a_n b}{3V^2}\right)\right]\right\}. \tag{4.107}$$

Finally, recall that $\lim_{n\to\infty} a_n = 0$ in Eq. (12.1). By letting n tend to infinity, we prove the achievability part:

$$\lim_{n\to+\infty} \frac{1}{na_n^2}\log\alpha_n(\eta_n) \leq -\frac{1}{2V}. \tag{4.108}$$

□

4.3.2 Proof of Theorem 4.6

The converse part is a direct consequence of the sharp converse Hoeffding bound, Theorem 4.3, in Section 4.2.

doi:10.6342/NTU201800597

Let $r_n := D(\rho\|\sigma) - a_n$, $\mathcal{X} = \{x\}$ and $W_x = \rho$. We apply Theorem 4.3 with $r = r_n$ to obtain

$$\widehat{\alpha}_{\exp\{-nr_n\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \geq \frac{A}{s_n^\star\sqrt{n}} \exp\left\{-n\left[\sup_{0<\alpha\leq 1}\frac{1-\alpha}{\alpha}\left(D_\alpha(\rho\|\sigma) - (r_n - c_n)\right)\right]\right\}, \tag{4.109}$$

for sufficiently large $n \in \mathbb{N}$ and some constant $A > 0$. Here

$$s_n^\star := \arg\max_{s\geq 0}\left\{sD_{\frac{1}{1+s}}(\rho\|\sigma) - sr_n\right\}. \tag{4.110}$$

Now let

$$\delta_n := a_n + c_n, \quad \forall n \in \mathbb{N}, \tag{4.111}$$

and invoke Proposition 12.2 with $W_x = \rho$, $P(x) = 1$, and substitute $P^\star\mathcal{W}$ with σ to obtain

$$\limsup_{n\to+\infty}\frac{\sup_{s\geq 0}\left\{-s\left(D(\rho\|\sigma) - \delta_n\right) + sD_{\frac{1}{1+s}}(\rho\|\sigma)\right\}}{\delta_n^2} \leq \frac{1}{2V}. \tag{4.112}$$

Moreover, Eq. (12.46) in Proposition 12.2 in Section 12.2 gives that $\lim_{n\to+\infty}\frac{s_n^\star}{\delta_n} \leq 1/V$. Here, we delay the proof of Proposition 12.2 to Section 12.3 for the reason that we unify the proofs for the exponent in quantum hypothesis testing and c-q channel coding there.

Combining Eqs. (4.109) and (4.112) concludes our claim:

$$\liminf_{n\to+\infty}\frac{\log\widehat{\alpha}_{\exp\{-nr_n\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)}{n\delta_n^2} \geq -\frac{1}{2V}. \tag{4.113}$$

□

doi:10.6342/NTU201800597

Part II

Information Storage with a Quantum Helper

doi:10.6342/NTU201800597

Chapter 5

Error Exponent Functions (Source Coding)

In this chapter, we define different versions of the exponent functions and auxiliary functions for Slepian-Wolf coding with QSI. We prove a variational representation in Section 5.1. The properties of the auxiliary function and exponent functions are provided in Sections 5.2 and 5.3, respectively.

For $t = \{\,\}$, $\{*\}$ or $\{\flat\}$, we define

$$E_{\mathrm{r}}^{t}(R) := \max_{0\le s\le 1} \left\{E_0^t(s) + sR\right\}; \tag{5.1}$$

$$E_{\mathrm{sp}}^{t}(R) := \sup_{s\ge 0} \left\{E_0^t(s) + sR\right\}; \tag{5.2}$$

$$E_{\mathrm{sc}}^{t}(R) := \sup_{-1<s<0} \left\{E_0^t(s) + sR\right\}; \tag{5.3}$$

$$E_0^t(s) := -sH_{\frac{1}{1+s}}^{t,\uparrow}(X|B)_\rho, \tag{5.4}$$

where $H_\alpha^{t,\uparrow}$ is the Rényi conditional entropy defined in Section 3.2. For $t = \{\,\}$, i.e. the Petz's Rényi conditional entropy, quantum Sibson's identity given in Lemma 3.3 shows that the auxiliary function $E_0(s)$ admits an closed-form:

$$E_0(s) = -\log \mathrm{Tr}\left[\left(\mathrm{Tr}_X\, \rho_{XB}^{\frac{1}{1+s}}\right)^{1+s}\right] \tag{5.5}$$

We also define another version of the exponent function via $H_\alpha^{\downarrow}$:

$$E_{\mathrm{r}}^{\downarrow}(R) := \max_{0\le s\le 1} \left\{E_0^{\downarrow}(s) + sR\right\}, \tag{5.6}$$

$$E_0^{\downarrow}(s) := -sH_{1-s}^{\downarrow}(X|B)_\rho. \tag{5.7}$$

The Golden-Thompson inequality given in Lemma 2.7 implies that

$$E_{\mathrm{sp}}(R) \le E_{\mathrm{sp}}^{\flat}(R) \tag{5.8}$$

$$E_{\mathrm{r}}(R) \le E_{\mathrm{r}}^{\flat}(R). \tag{5.9}$$

$$E_{\mathrm{sc}}(R) \le E_{\mathrm{sc}}^{\flat}(R). \tag{5.10}$$

doi:10.6342/NTU201800597

Further, since $H_\alpha^\uparrow(X|B)_\rho \leq H_{2-\frac{1}{\alpha}}^\downarrow(X|B)_\rho$ for $\alpha \in [1/2, +\infty]$ [119, Corollary 4], [8, Corollary 5.3]. For $R \in [H_1^\uparrow(X|B)_\rho, H_{1/2}^\uparrow(X|B)_\rho]$, together with Proposition 5.3-(a) below, we have

$$E_{\mathrm{r}}^\downarrow(R) \leq E_{\mathrm{r}}(R) = E_{\mathrm{sp}}(R) \leq E_{\mathrm{sp}}^\flat(R) = E_{\mathrm{r}}^\flat(R). \tag{5.11}$$

In Chapter 6 later, we obtain an achievability bound of the optimal error in terms of $E_{\mathrm{r}}^\downarrow$. We conjecture that it can be further improved by E_{r}.

5.1 Variational Representations

In Theorem 5.1 below, we show that the exponent functions defined in terms of $D^\flat$ admit the variational representations as introduced by Csiszár and J. Körner's [54, 55, 25].

Theorem 5.1 (Variational Representations). *Let ρ_{XB} be a classical-quantum state. Then,*

$$E_{\mathrm{r}}^\flat(R) = \min_{\sigma_{XB}\in\mathcal{S}(XB)} \left\{ D\left(\sigma_{XB}\|\rho_{XB}\right) + \left|R - H(X|B)_\sigma\right|^+ \right\}, \tag{5.12}$$

$$E_{\mathrm{sp}}^\flat(R) = \min_{\sigma_{XB}\in\mathcal{S}(XB)} \left\{ D\left(\sigma_{XB}\|\rho_{XB}\right) : R \leq H(X|B)_\sigma \right\}, \tag{5.13}$$

$$E_{\mathrm{sc}}^\flat(R) = \min_{\sigma_{XB}\in\mathcal{S}(XB)} \left\{ D\left(\sigma_{XB}\|\rho_{XB}\right) + \left|H(X|B)_\sigma - R\right|^+ \right\}. \tag{5.14}$$

Proof of Theorem 5.1. We only provide the proof for Eq (5.13) since Eqs. (5.12) and (5.14) follow similarly. The method of Lagrange multipliers gives that

$$\min_{\sigma_{XB}\in\mathcal{S}(XB)} \left\{ D\left(\sigma_{XB}\|\rho_{XB}\right) : R \leq H(X|B)_\sigma \right\} \tag{5.15}$$

$$= \sup_{s\geq 0} \min_{\sigma_{XB}\in\mathcal{S}(XB)} \left\{ D\left(\sigma_{XB}\|\rho_{XB}\right) + s\left[R - H(X|B)_\sigma\right] \right\} \tag{5.16}$$

$$= \sup_{s\geq 0} \min_{\sigma_{XB}\in\mathcal{S}(XB)} \left\{ D\left(\sigma_{XB}\|\rho_{XB}\right) + \min_{\tau_B\in\mathcal{S}(B)} sD\left(\sigma_{XB}\|\mathbb{1}_B\otimes\tau_B\right) + sR \right\} \tag{5.17}$$

$$= \sup_{s\geq 0} \min_{\tau_B\in\mathcal{S}(B)} \min_{\sigma_{XB}\in\mathcal{S}(XB)} \left\{ D\left(\sigma_{XB}\|\rho_{XB}\right) + sD\left(\sigma_{XB}\|\mathbb{1}_B\otimes\tau_B\right) + sR \right\} \tag{5.18}$$

$$= \sup_{s\geq 0} \min_{\tau_B\in\mathcal{S}(B)} \left\{ sD_{\frac{1}{1+s}}^\flat\left(\rho_{XB}\|\mathbb{1}_B\otimes\tau_B\right) + sR \right\} \tag{5.19}$$

$$= \sup_{s\geq 0} \left\{ E_0^\flat\left(s, \rho_{XB}\right) + sR \right\}, \tag{5.20}$$

where we use the representation $H(X|B)_\sigma = \max_{\tau_B\in\mathcal{S}(B)} -D\left(\sigma_{XB}\|\mathbb{1}_X\otimes\tau_B\right)$ in Eq. (5.17); Eq. (5.19) follows the Lemma 3.1 in Section 3.1, which was proved by Mosonyi and Ogawa [58]; in the last line (5.20) we recall the definition $E_0^\flat\left(s, \rho_{XB}\right) := -sH_{\frac{1}{1+s}}^{\flat,\uparrow}(X|Y)_\rho$.

□

5.2 Properties of Auxiliary Functions

In the following, we collect some useful properties of the auxiliary functions $E_0(s)$ and $E_0^\downarrow(s)$.

doi:10.6342/NTU201800597

Proposition 5.1 (Properties of E_0). *Let ρ_{XB} be a classical-quantum state with $H(X|Y)_\rho > 0$, the auxiliary function $E_0(s)$ defined in Eq. (5.5) admits the following properties.*

(a) (Continuity) The function $s \mapsto E_0(s)$ is smooth for all $s \in (-1, +\infty)$.

(b) (Negativity)

$$E_0(s) \leq 0, \quad s \geq 0 \tag{5.21}$$

with $E_0(0) = 0$.

(c) (Concavity) The function $s \mapsto E_0(s)$ is concave in s for all $s \in (-1, +\infty)$.

(d) (First-order Derivative)

$$\left.\frac{\partial E_0(s)}{\partial s}\right|_{s=0} = -H(X|B)_\rho. \tag{5.22}$$

(e) (Second-order Derivative)

$$\left.\frac{\partial^2 E_0(s)}{\partial s^2}\right|_{s=0} = -V(X|B)_\rho. \tag{5.23}$$

The proof is provided in Section 5.2.1 below.

Proposition 5.2 (Properties of $E_0^\downarrow$). *Let ρ_{XB} be a classical-quantum state with $H(X|Y)_\rho > 0$, the auxiliary function $E_0^\downarrow(s)$ defined in Eq. (5.7) admits the following properties.*

(a) (Continuity) The function $s \mapsto E_0(s)$ is smooth for all $s \in [0, +\infty)$.

(b) (Negativity)

$$E_0^\downarrow(s) \leq 0, \quad s \geq 0 \tag{5.24}$$

with $E_0^\downarrow(0) = 0$.

(c) (Concavity) The function $s \mapsto E_0^\downarrow(s)$ is concave in s for all $s \in (-1, +\infty)$.

(d) (First-order Derivative)

$$\left.\frac{\partial E_0^\downarrow(s)}{\partial s}\right|_{s=0} = -H(X|B)_\rho. \tag{5.25}$$

(e) (Second-order Derivative)

$$\left.\frac{\partial^2 E_0^\downarrow(s)}{\partial s^2}\right|_{s=0} = -V(X|B)_\rho. \tag{5.26}$$

The proof is provided in Section 5.2.2 below.

doi:10.6342/NTU201800597

5.2.1 Proof of Proposition 5.1

Proof of Proposition 5.1.

(5.1)-(a) (Continuity) Since $E_0(s)$ admits a closed-form

$$-\log \mathrm{Tr}\left[\mathrm{Tr}_X \rho_{XB}^{\frac{1}{1+s}}\right]^{1+s}, \quad \forall s > -1. \tag{5.27}$$

It is clearly smooth for all $s > -1$.

(5.1)-(b) (Negativity) The negativity of $E_0(s)$ directly follows from the non-negativity of the conditional Rényi entropy and the definition, Eq. (5.4).

(5.1)-(c) (Concavity) The concavity for $s \geq 0$ can be proved with the geometric matrix means in [36]. Here, we present another proof by the following matrix inequality.

Let $\rho_{XB} = \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x| \otimes W_x$, $t = \gamma = 1$, $i = x$, $k = |\mathcal{X}|$, $A_i = P(x)W_x$, and $Z_i = I_{n,m}$. We obtain the log-convexity of the map by applying Lemma 2.13:

$$p \mapsto \mathrm{Tr}\left(\sum_{x\in\mathcal{X}} (P(x)W_x)^{\frac{1}{p}}\right)^p, \quad \forall p > 0, \tag{5.28}$$

which is exactly the concavity of the map $s \mapsto E_0(s)$ for all $s > -1$.

(5.1)-(d) (First-order derivative) By the definition of $E_0(s)$,

$$\left.\frac{\partial E_0(s)}{\partial s}\right|_{s=0} = -H^{\uparrow}_{\frac{1}{1+s}}(X|B)_\rho - s\left.\frac{\partial H^{\uparrow}_{\frac{1}{1+s}}(X|B)_\rho}{\partial s}\right|_{s=0} = -H_(X|B)_\rho. \tag{5.29}$$

(5.1)-(e) (Second-order derivative) Similar to Item (d), it follows that

$$\left.\frac{\partial^2 E_0(s)}{\partial s^2}\right|_{s=0} = -2\frac{\partial H^{\uparrow}_{\frac{1}{1+s}}(X|B)_\rho}{\partial s} - s\left.\frac{\partial^2 H^{\uparrow}_{\frac{1}{1+s}}(X|B)_\rho}{\partial s^2}\right|_{s=0}. \tag{5.30}$$

The above equation indicates that we need to evaluate the first-order derivative of $H^{\uparrow}_{\frac{1}{1+s}}(X|B)_\rho$ at 0. In the following, we directly deal with the closed-form expression, Eq. (5.5).

To ease the burden of derivations, we denote some notation:

$$f(s) := \mathrm{Tr}_X \rho_{XB}^{1/(1+s)}, \tag{5.31}$$

$$g(s) := f(s)^{(1+s)}, \tag{5.32}$$

$$F(s) := \mathrm{Tr}\left[g(s)\right], \tag{5.33}$$

doi:10.6342/NTU201800597

Then,

$$\frac{\partial E_0(s)}{\partial s} = -\frac{F'(s)}{F(s)} \tag{5.34}$$

$$\frac{\partial^2 E_0(s)}{\partial s^2} = -\frac{F''(s)}{F(s)} - \left(\frac{\partial E_0(s)}{\partial s}\right)^2. \tag{5.35}$$

Direct calculation shows that

$$f'(s) = -\frac{1}{(1+s)^2}\,\mathrm{Tr}_X\,\rho_{XB}^{1/(1+s)}\log\rho_{XB}, \tag{5.36}$$

$$f''(s) = \frac{1}{(1+s)^3}\,\mathrm{Tr}_X\,\rho_{XB}\log\rho_{XB}\cdot\left[2+\frac{\log\rho_{XB}}{(1+s)}\right]. \tag{5.37}$$

Note that[1] $g(s) = \mathrm{e}^{(1+s)\log f(s)}$. By applying the chain rule of the Fréchet derivatives, one can show

$$g'(s) = \mathsf{D}\exp\left[\log g(s)\right]\left((1+s)\mathsf{D}\log\left[f(s)\right]\left(f'(s)\right)+\log f(s)\right). \tag{5.38}$$

Further, we employ Lemma 2.11 and Eqs.(5.33), (5.38), to obtain

$$F'(s) = \mathrm{Tr}\left[g'(s)\left((1+s)\mathsf{D}\log[f(s)](f'(s))+\log f(s)\right)\right], \tag{5.39}$$

$$\begin{aligned} F''(s)|_{s=0} = {} & \mathrm{Tr}\left[g'(s)\left((1+s)\mathsf{D}\log\left[f(s)\right]\left(f'(s)\right)+\log f(s)\right)\right]\big|_{s=0} \\ & + \mathrm{Tr}\left[g(s)\left(2\mathsf{D}\log\left[f(s)\right]\left(f'(s)\right)+(1+s)\left\{\mathsf{D}\log\left[f(s)\right]\left(f''(s)\right)\right.\right.\right. \\ & + \left.\left.\left.\mathsf{D}^2\log\left[f(s)\right]\left(f'(s)\right)\right\}\right)\right]\big|_{s=0}. \end{aligned} \tag{5.40}$$

Before evaluating $F''(s)$ at $s=0$, note that Eqs. (5.31), (5.32), (5.36), (5.37), and (5.38) yield

$$f(0) = g(0) = \rho_B, \tag{5.41}$$

$$f'(0) = -\,\mathrm{Tr}_X\,\rho_{XB}\log\rho_{XB}, \tag{5.42}$$

$$f''(0) = 2\,\mathrm{Tr}_X\,\rho_{XB}\log\rho_{XB} + \mathrm{Tr}_X\,\rho_{XB}\log^2\rho_{XB}, \tag{5.43}$$

$$g'(0) = \mathsf{D}\exp\left[\log g(0)\right]\left((1+0)\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)+\log f(0)\right) \tag{5.44}$$

$$= \mathsf{D}\exp\left[\log f(0)\right]\left(\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)+\log f(0)\right) \tag{5.45}$$

$$= f'(0) + f(0)\log f(0) \tag{5.46}$$

$$= -\,\mathrm{Tr}_X\,\rho_{XB}\log\rho_{XB} + \rho_B\log\rho_B. \tag{5.47}$$

From Eqs. (5.46), (5.39), the first term in Eq. (5.40) leads to

$$\mathrm{Tr}\left[g'(0)\left((1+0)\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)+\log f(0)\right)\right] \tag{5.48}$$

$$= \mathrm{Tr}\left[f'(0)\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)+2f'(0)\log f(0)+f(0)\log^2 f(0)\right] \tag{5.49}$$

$$= \mathrm{Tr}\left[f'(0)\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)-2\,\mathrm{Tr}_X\,\rho_{XB}\log\rho_{XB}\cdot\log\rho_B+\rho_B\log^2\rho_B\right] \tag{5.50}$$

[1] Here, let's assume ρ_{XB} has full support on $\mathcal{S}(XB)$ for brevity. The general case should hold with more technical derivations.

doi:10.6342/NTU201800597

Further, from Eqs. (5.37), (5.42), and (5.43), the second term in Eq. (5.40) leads to

$$\begin{aligned}
&\mathrm{Tr}\left[f(0)\left(2\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)+\left\{\mathsf{D}\log\left[f(0)\right]\left(f''(0)\right)\right.\right.\\
&\quad\left.\left.+\ \mathsf{D}^2\log\left[f(0)\right]\left(f'(0)\right)\right\}\right)\right] &(5.51)\\
&=\mathrm{Tr}\left[2f'(0)+f''(0)-f'(0)\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)\right] &(5.52)\\
&=\mathrm{Tr}\left[\mathrm{Tr}_X\,\rho_{XB}\log^2\rho_{XB}-f'(0)\mathsf{D}\log\left[f(0)\right]\left(f'(0)\right)\right]. &(5.53)
\end{aligned}$$

Combining Eqs. (5.40), (5.50), (5.53) gives

$$F''(0)=\mathrm{Tr}\left[\rho_{XB}\left(\log\rho_{XB}-\log\mathbb{1}_X\otimes\rho_B\right)^2\right]. \tag{5.54}$$

Finally, Eqs. (5.35) and (5.54) conclude our result:

$$\left.\frac{\partial E_0(s)}{\partial s}\right|_{s=0}=-V(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_B)=-V(X|Y)_\rho. \tag{5.55}$$

Moreover, Eq. (5.30) gives

$$\left.\frac{\partial H_\alpha^\uparrow(X|B)_\rho}{\partial\alpha}\right|_{\alpha=0}=\frac{1}{2}V(X|B)_\rho. \tag{5.56}$$

□

5.2.2 Proof of Proposition 5.2

Proof of Proposition 5.2.

(5.2)-(a) (Continuity) Since $E_0^\downarrow(s)=-\log\mathrm{Tr}\left[\rho_{XB}^{1-s}\left(\mathbb{1}_X\otimes\rho_B\right)^s\right]$

$$-\log\mathrm{Tr}\left[\left(\mathrm{Tr}_X\,\rho_{XB}^{\frac{1}{1+s}}\right)^{1+s}\right],\quad\forall s>-1. \tag{5.57}$$

It is smooth for all $s\geq 0$.

(5.2)-(b) (Negativity) The negativity of $E_0^\downarrow(s,\rho_{XB})$ directly follows from the non-negativity of the conditional Rényi entropy and the definition, Eq. (5.4).

(5.2)-(c) (Concavity) The claim follows from the concavity of the map $s\mapsto sD_{1-s}(\cdot\|\cdot)$, Eq. (3.12) in Lemma 3.2.

doi:10.6342/NTU201800597

(5.2)-(d) (First-order derivative) One can verify that

$$\left.\frac{\partial E_0^{\downarrow}(s,\rho_{XB})}{\partial s}\right|_{s=0} = D_{1-s}\left(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B\right) - sD'_{1-s}\left(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B\right)\big|_{s=0} \tag{5.58}$$

$$= D_{1-s}\left(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B\right)\big|_{s=0} \tag{5.59}$$

$$= D(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B) \tag{5.60}$$

$$= -H(X|B)_\rho. \tag{5.61}$$

(5.2)-(e) (Second-order derivative) Continuing from item (d), one obtain

$$\left.\frac{\partial^2 E_0^{\downarrow}(s)}{\partial s^2}\right|_{s=0} = -2D'_{1-s}\left(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B\right) + sD''_{1-s}\left(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B\right)\big|_{s=0} \tag{5.62}$$

$$= -2D'_{1-s}\left(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B\right)\big|_{s=0} \tag{5.63}$$

$$= -V(\rho_{XB}\|\mathbb{1}_X\otimes\rho_B) \tag{5.64}$$

$$= V(X|B)_\rho, \tag{5.65}$$

where in equality (5.64) we use the fact $D'_{1/1+s}(\cdot\|\cdot)|_{s=0} = V(\cdot\|\cdot)/2$ [120, Theorem 2].

□

doi:10.6342/NTU201800597

5.3 Properties of Error Exponent Functions and Saddle-Point

Proposition 5.3 (Properties of the Exponent Function). *Let ρ_{XB} be a classical-quantum state with $H(X|B)_\rho > 0$, the following holds.*

(a) $E_{\mathrm{sp}}(\cdot)$ is convex, differentiable, and monotonically increasing on $[0,+\infty]$. Further,

$$E_{\mathrm{sp}}(R) = \begin{cases} 0, & R \le H_1^\uparrow(X|B)_\rho \\ E_{\mathrm{r}}(R), & H_1^\uparrow(X|B)_\rho \le R \le H_{1/2}^\uparrow(X|B)_\rho \\ +\infty, & R > H_0^\uparrow(X|Y)_\rho \end{cases}. \tag{5.66}$$

(b) Define

$$F_R(\alpha,\sigma_B) := \begin{cases} \dfrac{1-\alpha}{\alpha}\left(R + D_\alpha\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_B\right)\right), & \alpha\in(0,1), \\ 0, & \alpha = 1, \end{cases} \tag{5.67}$$

on $(0,1]\times\mathcal{S}(B)$. For $R\in(H_1^\uparrow(X|B)_\rho, H_0^\uparrow(X|B)_\rho)$, there exists a unique saddle-point $(\alpha^\star,\sigma^\star)\in(0,1)\times\mathcal{S}(B)$ of $F_R(\cdot,\cdot)$ such that

$$F_R(\alpha^\star,\sigma^\star) = \sup_{\alpha\in[0,1]}\inf_{\sigma_B\in\mathcal{S}(B)} F_R(\alpha,\sigma_B) = \inf_{\sigma_B\in\mathcal{S}(B)}\sup_{\alpha\in[0,1]} F_R(\alpha,\sigma_B) = E_{\mathrm{sp}}(R). \tag{5.68}$$

(c) Any saddle-point $(\alpha^\star,\sigma^\star)$ of $F_R(\cdot,\cdot)$ satisfies

$$\mathbb{1}_X\otimes\sigma^\star \gg \rho_{XB}. \tag{5.69}$$

Proof of Proposition 5.3.

(5.3)-(a) Item (a) in Proposition 3.1 shows that the map $\alpha\mapsto H_\alpha^\uparrow(X|B)_\rho$ is monotonically decreasing on $[0,1]$. Hence, from the definition:

$$E_{\mathrm{sp}}(R) := \sup_{\alpha\in(0,1]}\frac{1-\alpha}{\alpha}\left(R - H_\alpha^\uparrow(X|B)_\rho\right), \tag{5.70}$$

it is not hard to verify that $E_{\mathrm{sp}}(R) = +\infty$ for all $R > H_0^\uparrow(H|B)_\rho$; finite for all $R < H_0^\uparrow(H|B)_\rho$; and $E_{\mathrm{sp}}^{\mathsf{SW}}(R) = 0$, for all $R \ge H_1^\uparrow(H|B)_\rho$. Moreover, $E_{\mathrm{sp}}(R) = E_{\mathrm{r}}(R)$ for $R\in[H_1^\uparrow(X|Y)_\rho, H_{1/2}^\uparrow(X|Y)_\rho]$ by the definition in Eq. (5.1).

For every $\alpha\in(0,1]$, the function $\frac{1-\alpha}{\alpha}(R - H_\alpha^\uparrow(X|B)_\rho)$ is an non-decreasing, convex, and continuous function in $R\in\mathbb{R}_{>0}$. Since $E_{\mathrm{sp}}(R)$ is the pointwise supremum of the above function, $E_{\mathrm{sp}}(R)$ is non-decreasing, convex, and lower semi-continuous function for all $R\ge 0$. Furthermore, since a convex function is continuous on the interior of the interval if it is finite [121, Corollary 6.3.3], thus $E_{\mathrm{sp}}(R)$ is continuous for all $R < H_0^\uparrow(X|B)_\rho$, and continuous from the left at $R = H_0^\uparrow(X|B)_\rho$.

doi:10.6342/NTU201800597

(5.3)-(b) Let

$$\mathcal{S}_\rho(B) := \{\sigma_B \in \mathcal{S}(B) : \rho_{XB} \not\perp \mathbb{1}_X \otimes \sigma_B\}. \tag{5.71}$$

Fix an arbitrary $R \in (H_1^\uparrow(X|B)_\rho, H_0^\uparrow(X|B)_\rho)$. In the following, we first prove the existence of a saddle-point of $F_R(\cdot,\cdot)$ on $(0,1] \times \mathcal{S}_\rho(B)$. Ref. [122, Lemma 36.2] states that $(\alpha^\star, \sigma^\star)$ is a saddle point of $F_R(\cdot,\cdot)$ if and only if the supremum in

$$\sup_{\alpha\in(0,1]} \inf_{\sigma\in\mathcal{S}_\rho(B)} F_R(\alpha,\sigma) \tag{5.72}$$

is attained at $\alpha^\star \in (0,1]$, the infimum in

$$\inf_{\sigma\in\mathcal{S}_\rho(B)} \sup_{\alpha\in(0,1]} F_R(\alpha,\sigma) \tag{5.73}$$

is attained at $\sigma^\star \in \mathcal{S}_\rho(B)$, and the two extrema in Eqs. (9.150), (5.73) are equal and finite. We first claim that, $\forall \alpha \in (0,1]$,

$$\inf_{\sigma\in\mathcal{S}_\rho(B)} F_R(\alpha,\sigma) = \inf_{\sigma\in\mathcal{S}(B)} F_R(\alpha,\sigma). \tag{5.74}$$

To see this, observe that for any $\alpha \in (0,1)$, Eqs. (3.5) yield

$$\forall \sigma \in \mathcal{S}(B)\backslash\mathcal{S}_\rho(B), \quad D_\alpha\left(\rho_{XB}\|\mathbb{1}_X \otimes \sigma\right) = +\infty, \tag{5.75}$$

which, in turn, implies

$$\forall \sigma \in \mathcal{S}(B)\backslash\mathcal{S}_\rho(B), \quad F_R(\alpha,\sigma) = +\infty. \tag{5.76}$$

Further, Eq. (5.74) holds trivially when $\alpha = 1$. Hence, Eq. (5.74) yields

$$\sup_{\alpha\in(0,1]} \inf_{\sigma\in\mathcal{S}_\rho(B)} F_R(\alpha,\sigma) = \sup_{\alpha\in(0,1]} \inf_{\sigma\in\mathcal{S}(B)} F_R(\alpha,\sigma) \tag{5.77}$$

Owing to the fact $R < H_0^\uparrow(X|B)_\rho$ and Eq. (5.2), we have

$$E_{\rm sp}(R) = \sup_{\alpha\in(0,1]} \inf_{\sigma\in\mathcal{S}(B)} F_R(\alpha,\sigma) < +\infty, \tag{5.78}$$

which guarantees the supremum in the right-hand side of Eq. (5.78) is attained at some $\alpha \in (0,1]$. Namely, there exists some $\bar{\alpha}_R \in (0,1]$ such that

$$\sup_{\alpha\in(0,1]} \inf_{\sigma\in\mathcal{S}_\rho(B)} F_R(\alpha,\sigma) = \max_{\alpha\in[\bar{\alpha}_R,1]} \inf_{\sigma\in\mathcal{S}(B)} F_R(\alpha,\sigma) < +\infty. \tag{5.79}$$

Thus, we complete our claim in Eq. (5.72). It remains to show that the infimum in Eq.(9.151) is attained at some $\sigma^\star \in \mathcal{S}_\rho(B)$ and the supremum and infimum are exchangeable. To achieve this, we will show that $([\bar{\alpha}_R, 1], \mathcal{S}_\rho(B), F_R)$ is a closed saddle-element (see Definition 5.1 below) and employ the boundness of $[\bar{\alpha}_R, 1] \times \mathcal{S}_\rho(B)$ to conclude our claim.

doi:10.6342/NTU201800597

Definition 5.1 (Closed Saddle-Element [122]). We denote by $\mathtt{ri}$ and $\mathtt{cl}$ the relative interior and the closure of a set, respectively. Let $\mathcal{A}, \mathcal{B}$ be subsets of a real vector space, and $F : \mathcal{A} \times \mathcal{B} \to \mathbb{R} \cup \{\pm\infty\}$. The triple $(\mathcal{A}, \mathcal{B}, F)$ is called a closed saddle-element if for any $x \in \mathtt{ri}\,(\mathcal{A})$ (resp. $y \in \mathtt{ri}\,(\mathcal{B})$),

(i) $\mathcal{B}$ (resp. $\mathcal{A}$) is convex.

(ii) $F(x, \cdot)$ (resp. $F(\cdot, y)$) is convex (resp. concave) and lower (resp. upper) semi-continuous.

(iii) Any accumulation point of $\mathcal{B}$ (resp. $\mathcal{A}$) that does not belong to $\mathcal{B}$ (resp. $\mathcal{A}$), say y_o (resp. x_o) satisfies $\lim_{y \to y_o} F(x, y) = +\infty$ (resp. $\lim_{x \to x_o} F(x, y) = -\infty$).

Fix an arbitrary $\alpha \in \mathtt{ri}\,([\bar{\alpha}_R, 1]) = (\bar{\alpha}_R, 1)$. We check that $(\mathcal{S}_\rho(B), F_R(\alpha, \cdot))$ fulfills the three items in Definition 9.1. (i) The set $\mathcal{S}_\rho(B)$ is clearly convex. (ii) Eq. (3.15) in Lemma 3.2 implies that $\sigma \mapsto D_\alpha(W_x \| \sigma)$ is convex and lower semi-continuous. Since convex combination preservers the convexity and the lower semi-continuity, Eq. (5.67) yields that $\sigma \mapsto F_R(\alpha, \sigma)$ is convex and lower semi-continuous on $\mathcal{S}_\rho(B)$. (iii) Due to the compactness of $\mathcal{S}(B)$, any accumulation point of $\mathcal{S}_\rho(B)$ that does not belong to $\mathcal{S}_\rho(B)$, say σ_o, satisfies $\sigma_o \in \mathcal{S}(B) \backslash \mathcal{S}_\rho(B)$. Eqs. (5.75) and (5.76) then show that $F_R(\alpha, \sigma_o) = +\infty$.

Next, fix an arbitrary $\sigma \in \mathtt{ri}\,(\mathcal{S}_\rho(B))$. Owing to the convexity of $\mathcal{S}_\rho(B)$, it follows that $\mathtt{ri}\,(\mathcal{S}_\rho(B)) = \mathtt{ri}\,(\mathtt{cl}\,(\mathcal{S}_\rho(B)))$ (see e.g. [123, Theorem 6.3]). We first claim $\mathtt{cl}\,(\mathcal{S}_\rho(B)) = \mathcal{S}(B)$. To see this, observe that $\mathcal{S}_{>0}(B) \subseteq \mathcal{S}_\rho(B)$ since a full-rank operator is not orthogonal with ρ_{XB}. Hence,

$$\mathcal{S}(B) = \mathtt{cl}\,(\mathcal{S}_{>0}(B)) \subseteq \mathtt{cl}\,(\mathcal{S}_\rho(B))\,. \tag{5.80}$$

On the other hand, the fact $\mathcal{S}_\rho(B) \subseteq \mathcal{S}(B)$ leads to

$$\mathtt{cl}\,(\mathcal{S}_\rho(B)) \subseteq \mathtt{cl}\,(\mathcal{S}(B)) = \mathcal{S}(B). \tag{5.81}$$

By Eqs. (9.158) and (5.81), we deduce that

$$\mathtt{ri}\,(\mathcal{S}_\rho(B)) = \mathtt{ri}\,(\mathtt{cl}\,(\mathcal{S}_\rho(B))) = \mathtt{ri}\,(\mathcal{S}(B)) = \mathcal{S}_{>0}(B), \tag{5.82}$$

where the last equality in Eq. (5.82) follows from [124, Proposition 2.9]. Hence, we obtain

$$\forall \sigma \in \mathtt{ri}\,(\mathcal{S}_\rho(B)) \quad \text{and} \quad \mathbb{1}_X \otimes \sigma \gg \rho_{XB}. \tag{5.83}$$

Now we verify that $([\bar{\alpha}_R, 1], F_R(\cdot, \sigma))$ satisfies the three items in Definition 9.1. Fix an arbitrary $\sigma \in \mathtt{ri}\,(\mathcal{S}_\rho(B))$. (i) The set $(0, 1]$ is obviously convex. (ii) From Eq. (3.13) in Lemma 3.2, the map $\alpha \mapsto F_R(\alpha, \sigma)$ is continuous on $(0, 1)$. Further, it is not hard to verify that $F_R(1, \sigma) = 0 = \lim_{\alpha \uparrow 1} F_R(\alpha, \sigma)$ from Eqs. (5.83), (9.142), and (3.5). Item (b) in Proposition 3.1 implies that $\alpha \mapsto F_R(\alpha, \sigma)$ on $[\bar{\alpha}_R, 1)$ is concave. Moreover, the continuity of $\alpha \mapsto F_R(\alpha, \sigma)$ on $[\bar{\alpha}_R, 1)$ guarantees the concavity of $\alpha \mapsto F_R(\alpha, \sigma)$ on $[\bar{\alpha}_R, 1]$. (iii) Since $[\bar{\alpha}_R, 1]$ is closed, there is no accumulation point of $[\bar{\alpha}_R, 1]$ that does not belong to $[\bar{\alpha}_R, 1]$.

We are at the position to prove the saddle-point property. The closed saddle-element, along with the boundness of $\mathcal{S}_\rho(B)$ and Rockafellar's saddle-point result [122, Theorem 8], [123, Theorem

doi:10.6342/NTU201800597

37.3] imply that

$$-\infty < \sup_{\alpha\in[\bar{\alpha}_R,1]} \inf_{\sigma\in\mathcal{S}_\rho(B)} F_R(s,\sigma) = \min_{\sigma\in\mathcal{S}_\rho(B)} \sup_{\alpha\in[\bar{\alpha}_R,1]} F_R(s,\sigma). \tag{5.84}$$

Then Eqs. (9.157) and (5.84) lead to the existence of a saddle-point of $F_R(\cdot,\cdot)$ on $(0,1]\times\mathcal{S}_\rho(B)$. Next, we prove the uniqueness. The rate R and item (a) in Proposition 5.3 shows that

$$\sup_{0<\alpha\leq 1} \min_{\sigma\in\mathcal{S}(B)} F_R(\alpha,\sigma) \in \mathbb{R}_{>0}. \tag{5.85}$$

Note that $\alpha^\star = 1$ will not be a saddle point of $F_{R,P}(\cdot,\sigma)$ because $F_R(1,\sigma) = 0$, $\forall\sigma\in\mathcal{S}(B)$, contradicting Eq. (5.85).

Now, fix $\alpha^\star\in(0,1)$ to be a saddle-point of $F_R(\cdot,\cdot)$. Eq. (3.15) in Lemma 3.2 implies that the map $\sigma\mapsto D_{\alpha^\star}(\rho_{XB}\|\mathbb{1}_X\otimes\sigma)$ is strictly convex, and thus the minimizer of Eq. (5.85) is unique. Next, let $\sigma^\star\in\mathcal{S}_\rho(B)$ be a saddle-point of $F_R(\cdot,\cdot)$. Then,

$$F_R(\alpha,\sigma^\star) = \frac{1-\alpha}{\alpha}\left(R - H_\alpha^\uparrow(X|B)_\rho\right). \tag{5.86}$$

Item (b) in Proposition 3.1 then shows that $\frac{1-\alpha}{\alpha}H_\alpha^\uparrow(X|B)_\rho$ is strictly concave on $(0,1)$, which in turn implies that $F_R(\cdot,\sigma^\star)$ is also strictly concave on $(0,1)$. Hence, the maximizer of Eq. (9.163) is unique, which completes item (b) of Proposition 5.3.

(5.3)-(c) As shown in the proof of item (b), $\alpha^\star = 1$ is not a saddle point of $F_R(\cdot,\cdot)$ for any $R < H_0^\uparrow(X|B)_\rho$. We assume $(\alpha^\star,\sigma^\star)$ is a saddle-point of $F_R(\cdot,\cdot)$ with $\alpha^\star\in(0,1)$, it holds that

$$F_R(\alpha^\star,\sigma^\star) = \min_{\sigma\in\mathcal{S}B} F_R(\alpha^\star,\sigma) = \frac{1-\alpha^\star}{\alpha^\star}R + \frac{1-\alpha^\star}{\alpha^\star}\min_{\sigma\in\mathcal{S}(B)} D_{\alpha^\star}(\rho_{XB}\|\mathbb{1}_X\otimes\sigma). \tag{5.87}$$

By quantum Sibson's identity given in Lemma 3.3 (see also [125], [119, Lemma 1], [8, Lemma 5.1]), the minimizer of Eq. (5.87) is

$$\sigma^\star = \frac{\left(\mathrm{Tr}_X\left[\rho_{XB}^{\alpha^\star}\right]\right)^{\frac{1}{\alpha^\star}}}{\mathrm{Tr}\left[\left(\mathrm{Tr}_X\left[\rho_{XB}^{\alpha^\star}\right]\right)^{\frac{1}{\alpha^\star}}\right]}. \tag{5.88}$$

From this expression, it is clear that $\mathbb{1}_X\otimes\sigma^\star \gg \rho_{XB}$, and thus item (c) is proved.

□

doi:10.6342/NTU201800597

Chapter 6

Achievability (Source Coding)

The goal of this chapter is to prove a finite blocklength upper bound for the optimal probability error for Slepian-Wolf coding with QSI.

Theorem 6.1 (n-Shot Achievability Bound)**.** *Consider a Slepian-Wolf coding with a joint classical-quantum state $\rho_{XB} \in \mathcal{S}(XB)$ with $H(X|B)_\rho > 0$. Let $R < H(X|B)_\rho$. The following holds for every $n \in \mathbb{N}$,*

$$-\frac{1}{n}\log \varepsilon^*(n,R) \geq E_{\mathrm{r}}^{\downarrow}(R) - \frac{\log 4}{n}, \tag{6.1}$$

where

$$E_{\mathrm{r}}^{\downarrow}(R) := \sup_{\frac{1}{2}\leq\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(R - H_{2-\frac{1}{\alpha}}^{\downarrow}(X|B)_\rho\right), \tag{6.2}$$

and $H_\alpha^{\downarrow}(X|B)_\rho := -D_\alpha(\rho_{XB}\|\mathbb{1}_X \otimes \rho_B)$ for D_α being Petz's Rényi divergence, see Eq. (3.5).

Proof. Our technique it to use a random coding argument to prove Theorem 6.1. The idea originates from Gallager [56] and later studied by Renes and Renner [41].

We first present an one-shot achievability. It is not hard to extend to the n-tuple cases. Let $f : \mathcal{X} \to \mathcal{I}$ be a random encoder that encodes every source $x \in \mathcal{X}$ into some index $i \in \mathcal{I}$ with equal probability $1/M = 1/|\mathcal{I}|$. Then, the optimal probability of error can be upper bounded by

$$\begin{aligned} \varepsilon^*(1,\log M) &\leq \mathbb{E}_x \mathbb{E}_i \left[\varepsilon(x,i)\right], &(6.3)\\ &= \mathbb{E}_x \mathbb{E}_i \operatorname{Tr}\left[\rho_B^{(x)}\left(\mathbb{1}_B - \Lambda_x^{(i)}\right)\right], &(6.4)\end{aligned}$$

where we denote by $\varepsilon(\mathbf{x}^n, i)$ the error probability conditioned on $\mathbf{x}^n$ being the source and it is encoded into i. Here, the adopted decoder is a pretty good measurement:

$$\Lambda_x^{(i)} := \left(\sum_{\bar{x}:f(\bar{x})=i} \Pi_{\bar{x}}\right)^{-1/2} \Pi_x \left(\sum_{\bar{x}:f(\bar{x})=i} \Pi_{\bar{x}}\right)^{-1/2}, \tag{6.5}$$

where $0 \preceq \Lambda_x^{(i)} \preceq \mathbb{1}_B$ for each $i \in \mathcal{I}$ will be specified later. Applying the Hayashi-Nagaoka inequality

doi:10.6342/NTU201800597

[87, Lemma 2] to obtain

$$\mathbb{1}_X - \Lambda_x^{(i)} \preceq 2\left(\mathbb{1}_B - \Pi_x\right) + 4\sum_{\bar{x}\neq x} \mathbf{1}_{f(\bar{x})=i}\Pi_{\bar{x}}, \tag{6.6}$$

where $\mathbf{1}_{f(\bar{x})=i}$ denotes the indicator function when the event $f(\bar{x}) = i$ is true. Combining Eqs. (6.4) and (6.6) gives

$$\varepsilon(x,i) \leq 2\operatorname{Tr}\left[\rho_B^{(x)}\left(\mathbb{1}_B - \Pi_x\right)\right] + 4\operatorname{Tr}\left[\rho_B^{(x)}\sum_{\bar{x}\neq x}\mathbf{1}_{f(\bar{x})=i}\Pi_{\bar{x}}\right]. \tag{6.7}$$

Taking average over i and using the assumption $\Pr\{f(\bar{x}) = i\} = 1/M$ yield

$$\mathbb{E}_i\left[\varepsilon(x,i)\right] \leq 2\operatorname{Tr}\left[\rho_B^{(x)}\left(\mathbb{1}_B - \Pi_x\right)\right] + 4\Pr\{f(\bar{x}) = i\}\operatorname{Tr}\left[\rho_B^{(x)}\sum_{\bar{x}\neq x}\Pi_{\bar{x}}\right] \tag{6.8}$$

$$= 2\operatorname{Tr}\left[\rho_B^{(x)}\left(\mathbb{1}_B - \Pi_x\right)\right] + \frac{4}{M}\operatorname{Tr}\left[\rho_B^{(x)}\sum_{\bar{x}\neq x}\Pi_{\bar{x}}\right] \tag{6.9}$$

$$\leq 2\operatorname{Tr}\left[\rho_B^{(x)}\left(\mathbb{1}_B - \Pi_x\right)\right] + \frac{4}{M}\operatorname{Tr}\left[\rho_B^{(x)}\sum_{\bar{x}\in\mathcal{X}}\Pi_{\bar{x}}\right]. \tag{6.10}$$

By taking average over x we obtain

$$\varepsilon^*(1,\log M) \leq 2\sum_{x\in\mathcal{X}} P(x)\operatorname{Tr}\left[\rho_B^{(x)}\left(\mathbb{1}_B - \Pi_x\right)\right] + \frac{4}{M}\operatorname{Tr}\left[\rho_B\sum_{\bar{x}\in\mathcal{X}}\Pi_{\bar{x}}\right], \tag{6.11}$$

$$= 2\operatorname{Tr}\left[\rho_{XB}\left(\mathbb{1}_{XB} - \Pi_{XB}\right)\right] + \frac{4}{M}\operatorname{Tr}\left[\mathbb{1}_X\otimes\rho_B\Pi_{XB}\right], \tag{6.12}$$

where $\Pi_{XB} := \sum_{x\in\mathcal{X}}|x\rangle\langle x|\otimes\Pi_x$. Next, we invoke Audenaert *et al.*'s inequality [85, 86]: for every $\boldsymbol{X}, \boldsymbol{Y} \succeq 0$ and $s \in [0,1]$,

$$\operatorname{Tr}\left[\{\boldsymbol{X} - \boldsymbol{Y} \succeq 0\}\boldsymbol{Y} + \{\boldsymbol{Y} - \boldsymbol{X} \prec 0\}\boldsymbol{X}\right] \leq \operatorname{Tr}\left[\boldsymbol{X}^{1-s}\boldsymbol{Y}^s\right]. \tag{6.13}$$

Letting $\boldsymbol{X} = \rho_{XB}$, $\boldsymbol{Y} = \frac{1}{M}\mathbb{1}_X\otimes\rho_B$, $\Pi_{XB} = \left\{\rho_{XB} - \frac{1}{M}\mathbb{1}_X\otimes\rho_B \succeq 0\right\}$, we have one-shot achievability:

$$\varepsilon^*(1,\log M) \leq 4\min_{s\in[0,1]} M^{-s}\operatorname{Tr}\left[\rho_{XB}^{1-s}\left(\mathbb{1}_X\otimes\rho_B\right)^s\right]. \tag{6.14}$$

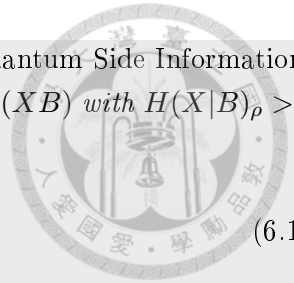
Finally, we consider the n-tuple case. Note that $\rho_{X^nB^n} = \rho_{XB}^{\otimes n}$, and let $M = \exp\{nR\}$. Eqs. (6.14) and (5.6) lead to

$$\varepsilon^*(n,R) \leq 4\exp\left\{-nE_{\mathrm{r}}^{\downarrow}(R)\right\}, \tag{6.15}$$

which completes the proof

□

doi:10.6342/NTU201800597



Conjecture 6.1 (Random Coding Bound for Slepian-Wolf Coding with Quantum Side Information)**.** *Consider a Slepian-Wolf coding with a joint classical-quantum state $\rho_{XB} \in \mathcal{S}(XB)$ with $H(X|B)_\rho > 0$. Let $R < H(X|B)_\rho$. The following holds for every $n \in \mathbb{N}$,*

$$-\frac{1}{n}\log \varepsilon^*(n,R) \geq E_{\mathrm{r}}(R), \tag{6.16}$$

where

$$E_{\mathrm{r}}(R) := \sup_{\frac{1}{2}\leq\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(R - H_\alpha^\uparrow(X|B)_\rho\right), \tag{6.17}$$

and $H_\alpha^\uparrow(X|B)_\rho := \max_{\sigma_B\in\mathcal{S}(B)} -D_\alpha(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_B)$ for D_α being Petz's Rényi divergence, see Eq. (3.5).

doi:10.6342/NTU201800597

Chapter 7

Optimality (Source Coding)

The main result of this section is the finite blocklength converse bound for the optimal error probability—Theorem 7.1. We termed this the sphere-packing bound for Slepian-Wolf coding with QSI, as a counterpart of the sphere-packing bound in classical-quantum channel coding; see Chapter 11. The proof technique relies on an one-shot converse bound in Proposition 7.1 below, and a sharp n-shot converse bound, Theorem 4.3, given in Section 4.2.

Theorem 7.1 (Sphere-Packing Bound for Slepian-Wolf Coding)**.** *theospSW Consider a Slepian-Wolf coding with a joint classical-quantum state $\rho_{XB} \in \mathcal{S}(XB)$ with $H(X|B)_\rho > 0$. Let $R \in (H(X|B)_\rho, H_0^\uparrow(X|B)_\rho)$. Then, there exist $N_0, K \in \mathbb{N}$, such that for all $n \geq N_0$, the following holds:*

$$-\frac{1}{n}\log \varepsilon^*(n,R) \leq E_{\mathrm{sp}}(R) + \frac{1}{2}\left(1 + \left|\frac{\partial E_{\mathrm{sp}}(r)}{\partial r}\Big|_{r=R}\right|\right)\frac{\log n}{n} + \frac{K}{n}, \tag{7.1}$$

where

$$E_{\mathrm{sp}}(R) := \sup_{0\leq\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(R - H_\alpha^\uparrow(X|B)_\rho\right), \tag{7.2}$$

and $H_\alpha^\uparrow(X|B)_\rho := \max_{\sigma_B\in\mathcal{S}(B)} -D_\alpha(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_B)$.

The proof is provided in Section 7.2

7.1 One-Shot Converse Bound (Hypothesis Testing Reduction)

Proposition 7.1 (One-Shot Converse Bound for Error)**.** *Consider a Slepian-Wolf coding with a joint classical-quantum state $\rho_{XB} \in \mathcal{S}(XB)$ and the index size $M < |\mathcal{X}|$. Then,*

$$-\log \varepsilon^*(1, \log M) \leq \min_{\sigma_B\in\mathcal{S}(B)} -\log \widehat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB}\|\tau_X\otimes\sigma_B), \tag{7.3}$$

where τ_X denotes the uniform distribution on the input alphabet $\mathcal{X}$; and $\widehat{\alpha}_\mu(\cdot\|\cdot)$ is defined in Eq. (4.3).

Proof of Proposition 7.1. We first claim that we can reduce to the case of determinstic encoders as follows. Assume for any deterministic encoder $\mathcal{E}: \mathcal{X} \to \mathcal{W}$ with index size $|\mathcal{W}| = M$, any decoder $\mathcal{D}$,

doi:10.6342/NTU201800597

and any state $\sigma_B \in \mathcal{S}(B)$, we have

$$1 - \sum_{x\in\mathcal{X}} p(y) \operatorname{Tr}[\Pi_y^{(\mathcal{E}(y))} \rho_B^y] \geq \hat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB} \| \tau_X \otimes \sigma_B) \tag{7.4}$$

for $\tau_X = \frac{1}{|\mathcal{X}|} \mathbb{1}_X$. Then given a random encoding F, we may average over its constituent deterministic encoders to obtain

$$1 - \sum_{x\in\mathcal{X}} p(y) \sum_{j=1}^{|\mathcal{F}|} P_j \operatorname{Tr}[\Pi_y^{(\mathcal{E}_j(y))} \rho_B^y] \geq \hat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB} \| \tau_X \otimes \sigma_B) \tag{7.5}$$

using that (7.4) holds for each $\mathcal{E}_j$. Then, since the RHS does not depend on the encoding or deocoding, we may minimize over random encodings F and decodings $\mathcal{D}$ to find

$$\varepsilon^* \geq \hat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB} \| \tau_X \otimes \sigma_B). \tag{7.6}$$

Thus

$$-\log \varepsilon^*(1, \log M) \leq -\log \hat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB} \| \tau_X \otimes \sigma_B). \tag{7.7}$$

Since the LHS does not depend on σ, we may minimize over it, yielding

$$-\log \varepsilon^*(1, \log M) \leq \inf_{\sigma_B \in \mathcal{S}(B)} -\log \hat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB} \| \tau_X \otimes \sigma_B) \tag{7.8}$$

which is the conjecture, (7.3).

Fix deterministic encoding $\mathcal{E}$ and a decoding strategy, i.e. a collection of POVMs $\{\mathcal{P}_w\}_{w\in\mathcal{W}}$, given by $\mathcal{P}_w = \{\Pi_{\hat{x}}^{(w)}\}_{\hat{x}\in\mathcal{X}}$. Consider the map $\Lambda : XB \to XB$ such that

$$\Lambda(|x\rangle\langle x| \otimes \sigma_B) = |x\rangle\langle x| \otimes \sum_{\hat{x}} \operatorname{Tr}[\Pi_{\hat{x}}^{(\mathcal{E}(x))} \sigma_B] |\hat{x}\rangle\langle\hat{x}|. \tag{7.9}$$

This is the map that encodes in the second register the probability of each measurement outcome of the POVM $\{\Pi_{\hat{x}}^{(\mathcal{E}(x))}\}_{\hat{x}\in\mathcal{X}}$, when x is in the first register. To see that Λ is completely positive (CP), let us define for each $x \in \mathcal{X}$ the measure-and-prepare map $\Lambda^x : B \to B$ given by

$$\Lambda^x : \sigma_B \mapsto \sum_{\hat{x}} \operatorname{Tr}[\Pi_{\hat{x}}^{(\mathcal{E}(x))} \sigma_B] |\hat{x}\rangle\langle\hat{x}|, \tag{7.10}$$

which is CP (see e.g. [50]). Then writing $L_{|x\rangle\langle x|}$ for left-multiplication by the projector $|x\rangle\langle x|$ and similarly $R_{|x\rangle\langle x|}$ for right-multiplication, we have that

$$\Lambda = \sum_{x\in\mathcal{X}} L_{|x\rangle\langle x|} R_{|x\rangle\langle x|} \otimes \Lambda^x. \tag{7.11}$$

Since $L_A R_A$ is CP for self-adjoint A (since A is its only Kraus operator), and the sum of CP maps is CP, we have that Λ is CP.

We define $\hat{\alpha}_\varepsilon(\rho\|\sigma)$ as the minimum type I error for a binary test discriminating between ρ and σ, with type II error bounded by ε. The type I error of a test T here is $\operatorname{Tr}[(\mathbb{1} - T)\rho]$ and the type II error

doi:10.6342/NTU201800597

is $\mathrm{Tr}[T\sigma]$, and therefore

$$\hat{\alpha}_\varepsilon(\rho\|\sigma) = \inf_{\substack{T:0\le T\le \mathbb{1} \\ \mathrm{Tr}[T\sigma]\le\varepsilon}} \mathrm{Tr}[(\mathbb{1}-T)\rho] \tag{7.12}$$

By writing the optimal type-I error into the hypothesis testing relative entropy [14],

$$-\log\hat{\alpha}_\varepsilon(\rho\|\sigma) = D_{\mathrm{H}}^{\varepsilon}(\sigma\|\rho). \tag{7.13}$$

Since the hypothesis testing relative entropy satisfies the DPI, we have

$$D_{\mathrm{H}}^{\varepsilon}(\rho\|\sigma) \ge D_{\mathrm{H}}^{\varepsilon}(\Phi(\rho)\|\Phi(\sigma)) \tag{7.14}$$

for any CP map Φ. Therefore,

$$\hat{\alpha}_\varepsilon(\rho\|\sigma) = \exp(-D_H^{\varepsilon}(\sigma\|\rho)) \le \exp(-D_H^{\varepsilon}(\Phi(\sigma)\|\Phi(\rho))) = \hat{\alpha}_\varepsilon(\Phi(\rho)\|\Phi(\sigma)). \tag{7.15}$$

We set $\tau_X = \frac{1}{|\mathcal{X}|}\mathbb{1}_X = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}|x\rangle\langle x|$ as the completely mixed state on X and $\sigma_B \in \mathcal{S}(B)$ arbitrary. Then for any $\varepsilon > 0$,

$$\hat{\alpha}_\varepsilon(\rho_{XB}\|\tau_X\otimes\sigma_B) \le \hat{\alpha}_\varepsilon(\Lambda(\rho_{XB})\|\Lambda(\tau_X\otimes\sigma_B)). \tag{7.16}$$

Let us consider these two states:

$$\Lambda(\tau_X\otimes\sigma_B) = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\Lambda(|x\rangle\langle x|\otimes\sigma_B) = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}|x\rangle\langle x|\otimes\sum_{\hat{x}}\mathrm{Tr}[\Pi_{\hat{x}}^{(\mathcal{E}(x))}\sigma_B]|\hat{x}\rangle\langle\hat{x}|, \tag{7.17}$$

and

$$\Lambda(\rho_{XB}) = \sum_{x\in\mathcal{X}}p(x)\Lambda(|x\rangle\langle x|\otimes\rho_B^x) = \sum_{x\in\mathcal{X}}p(x)|x\rangle\langle x|\otimes\sum_{\hat{x}\in\mathcal{X}}\mathrm{Tr}[\Pi_{\hat{x}}^{(\mathcal{E}(x))}\rho_B^x]|\hat{x}\rangle\langle\hat{x}|. \tag{7.18}$$

Now, take a two element POVM (the test) as $T = \sum_y |y\rangle\langle y|\otimes|y\rangle\langle y|$. Then,

$$\mathrm{Tr}[T\Lambda(\rho_{XB})] = \sum_y p(y)\,\mathrm{Tr}[\Pi_y^{(\mathcal{E}(y))}\rho_B^y], \tag{7.19}$$

so this test has type I error of $1-\sum_y p(y)\,\mathrm{Tr}[\Pi_y^{(\mathcal{E}(y))}\rho_B^y]$.

On the other hand,

$$\mathrm{Tr}[T\Lambda(\tau_X\otimes\sigma_B)] = \sum_y\frac{1}{|\mathcal{X}|}\mathrm{Tr}[\Pi_y^{(\mathcal{E}(y))}\sigma_B] = \frac{1}{|\mathcal{X}|}\sum_{w\in\mathcal{W}}\sum_{y\in\mathcal{X}:\mathcal{E}(y)=w}\mathrm{Tr}[\Pi_y^{(\mathcal{E}(y))}\sigma_B]. \tag{7.20}$$

Since

$$\sum_{y\in\mathcal{X}:\mathcal{E}(y)=w}\mathrm{Tr}[\Pi_y^{(\mathcal{E}(y))}\sigma_B] \le \sum_{y\in\mathcal{X}}\mathrm{Tr}[\Pi_y^{(w)}\sigma_B] = \mathrm{Tr}[\sigma_B] = 1, \tag{7.21}$$

we have

$$\mathrm{Tr}[T\Lambda(\tau_X\otimes\sigma_B)] \le \frac{M}{|\mathcal{X}|}. \tag{7.22}$$

doi:10.6342/NTU201800597

That is, this test achieves type II error of $\frac{M}{|\mathcal{X}|}$. As the infimum over all such tests, we have that

$$1-\sum_y p(y)\operatorname{Tr}[\Pi_y^{(\mathcal{E}(y))}\rho_B^y] \geq \hat{\alpha}_{\frac{|\mathcal{W}|}{|\mathcal{X}|}}(\Lambda(\rho_{XB})\|\Lambda(\tau_X\otimes\sigma_B)) \geq \hat{\alpha}_{\frac{|\mathcal{W}|}{|\mathcal{X}|}}(\rho_{XB}\|\tau_X\otimes\sigma_B) \tag{7.23}$$

where the second inequality is by (7.16). Then taking the infimum over $\mathcal{E}$ and $\mathcal{D}$,

$$\hat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB}\|\tau_X\otimes\sigma_B) \leq \varepsilon_{\mathrm{SW}}^*(1,\log|\mathcal{W}|). \tag{7.24}$$

Thus,

$$-\log\varepsilon^*(1,\log M) \leq -\log\hat{\alpha}_{\frac{|\mathcal{W}|}{|\mathcal{X}|}}(\rho_{XB}\|\tau_X\otimes\sigma_B). \tag{7.25}$$

Since this holds independently of $\sigma_B\in\mathcal{S}(B)$, we may minimize over σ_B to find

$$-\log\varepsilon^*(1,\log M) \leq \min_{\sigma_B\in\mathcal{S}(B)} -\log\hat{\alpha}_{\frac{M}{|\mathcal{X}|}}(\rho_{XB}\|\tau_X\otimes\sigma_B),$$

which complete our claim. □

7.2 Proof of Theorem 7.1

Proof of Theorem 7.1. The proof is split into two parts. We first invoke an one-shot converse bound in Proposition 7.1 to relate the optimal error of Slepian-Wolf coding to a binary hypothesis testing problem. Second, we employ a sharp converse Hoeffding bound in Theorem 4.3 to asymptotically expand the optimal type-I error, which yields the desired result in Eq. (7.1).

Applying Proposition 7.1 with $\rho_{X^nB^n}\in\mathcal{S}(X^nB^n)$ and $M=\exp\{nR\}$ gives

$$\begin{aligned}
\log\left(\frac{1}{\varepsilon_{\mathsf{SW}}^*(n,R)}\right) &\leq \min_{\sigma_B^n\in\mathcal{S}(B^n)} -\log\widehat{\alpha}_{\frac{M}{|\mathcal{X}^n|}}(\rho_{X^nB^n}\|U_{X^n}\otimes\sigma_B^n) &(7.26)\\
&\leq -\log\widehat{\alpha}_{\frac{M}{|\mathcal{X}^n|}}\left(\rho_{X^nB^n}\|U_{X^n}\otimes(\sigma_B^\star)^{\otimes n}\right), &(7.27)\\
&= -\log\widehat{\alpha}_{\frac{M}{|\mathcal{X}|^n}}\left(\rho_{XB}^{\otimes n}\|\left(U_X\otimes\sigma_B^\star\right)^{\otimes n}\right), &(7.28)
\end{aligned}$$

where we invoke the saddle-point property in Proposition 5.3-(b) to denote by

$$\sigma_R^\star := \min_{\sigma_B\in\mathcal{S}(B)}\sup_{\alpha\in[0,1]}\frac{1-\alpha}{\alpha}\left(R+D_\alpha\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_B\right)\right). \tag{7.29}$$

Next, we show that the exponent $\phi_n>0$, and thus we can exploit Theorem 4.3 to expand the right-hand side of Eq. (7.28). Let $r=\log|\mathcal{X}|-R$, and note that item (c) in Proposition 5.3 implies

$$\rho_{XB}\ll U_X\otimes\sigma_R^\star. \tag{7.30}$$

doi:10.6342/NTU201800597

One can verify that

$$\phi_n\left(r|\rho_{XB}^{\otimes n}\|\left(U_X\otimes\sigma_R^\star\right)^{\otimes n}\right) = \sup_{\alpha\in(0,1]}\frac{1-\alpha}{\alpha}\left(D_\alpha\left(\rho_{XB}\|U_X\otimes\sigma_R^\star\right)-r\right) \tag{7.31}$$

$$= \sup_{\alpha\in(0,1]}\frac{1-\alpha}{\alpha}\left(D_\alpha\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_R^\star\right)-\log|\mathcal{X}|-r\right) \tag{7.32}$$

$$= E_{\mathrm{sp}}(R) \tag{7.33}$$

$$> 0, \tag{7.34}$$

where ϕ_n is defined in Eq. (2.50); equality (7.33) follows from the saddle-point property, item (b) in Proposition 5.3, and the definition of $E_{\mathrm{sp}}(R)$ in Eq. (5.2); the last inequality (7.34) is due to item (a) in Proposition 5.3 and the given range of R. Further, the positivity of $\phi_n\left(r|\rho_{XB}^{\otimes n}\|\left(\tau_X\otimes\sigma_R^\star\right)^{\otimes n}\right)$ implies that $r > D_0(\rho_{XB}\|\tau_X\otimes\sigma_R^\star)$. By choosing $\varepsilon = r - D_0(\rho_{XB}\|\tau_X\otimes\sigma_R^\star) > 0$, $\rho = \rho_{XB}$ and $\sigma = \tau_X\otimes\sigma_R^\star$, Eq. (7.31) guarantees the positivity of ϕ_n. Hence, we apply Theorem 4.3 on Eq. (7.28) to obtain

$$\log\left(\frac{1}{\varepsilon^*(n,R)}\right)$$

$$\leq n\phi_n\left(r|\rho_{XB}^{\otimes n}\|\left(\tau_X\otimes\sigma_R^\star\right)^{\otimes n}\right)+\frac{1}{2}\left(1+\left|\frac{\partial\phi_n\left(\tilde{r}|\rho_{XB}^{\otimes n}\|\left(\tau_X\otimes\sigma_R^\star\right)^{\otimes n}\right)}{\partial\tilde{r}}\bigg|_{\tilde{r}=r}\right|\right)\log n + K, \tag{7.35}$$

where $K > 0$ is some finite constant independent of n. Finally, combining Eqs. (7.33) and (7.35) completes the proof.

□

doi:10.6342/NTU201800597

Chapter 8

Moderate Deviation Analysis (Source Coding)

In this chapter, we provide the moderate deviation analysis for Slepian-Wolf coding with QSI. As we have shown in Chapters 6 and 7, the optimal probability of error exponentially decay to zero as the compression rate is above the Slepian-Wolf limit $H(X|B)_\rho$. In Theorem 8.1 below, we consider the scenario that the compression rate approaches $H(X|B)_\rho$ from above at a speed a_n, which satisfies

$$\begin{aligned} &\text{(i)} \lim_{n\to+\infty} a_n = 0; \\ &\text{(ii)} \lim_{n\to+\infty} a_n\sqrt{n} = +\infty. \end{aligned} \tag{8.1}$$

Then, the optimal probability of error still goes to zero asymptotically.

Theorem 8.1 (Moderate deviations for the error). *theomodlarge Consider a Slepian-Wolf coding with a joint classical-quantum state $\rho_{XB} \in \mathcal{S}(XB)$ and $V(X|B)_\rho > 0$. For any sequence $(a_n)_{n\in\mathbb{N}}$ satisfying Eq. (1.7),*

$$\lim_{n\to+\infty} \frac{1}{na_n^2} \log \varepsilon^* \left(n, H(X|B)_\rho + a_n\right) = -\frac{1}{2V(X|B)_\rho}, \tag{8.2}$$

where the conditional information variance is defined by $V(X|B)_\rho := V(\rho_{XB}\|\mathbb{1}_X \otimes \rho_B)$ and $V(\rho\|\sigma) := \mathrm{Tr}[\rho\,(\log\rho - \log\sigma)^2] - D(\rho\|\sigma)^2$.

Proof of Theorem 8.1. We shorthand $H = H(X|B)_\rho$, $V = V(X|B)_\rho$ for notational convenience. We first show the achievability, i.e. the "$\geq$" in Eq. (8.2). Let $\{a_n\}_{n\geq 1}$ be any sequence of real numbers satisfying Eq. (8.1). For every $n \in \mathbb{N}$, Theorem 6.1 implies that there exists a sequence of n-block codes with rates $R_n = H + a_n$ such that

$$\varepsilon^*(n, R_n) \leq 4\exp\left\{-n\left[\max_{0\leq s\leq 1}\left\{E_0^\downarrow(s) + sR_n\right\}\right]\right\}. \tag{8.3}$$

Applying Taylor's theorem to $E_0^\downarrow(s)$ at $s = 0$ together with Proposition 5.2 gives

$$E_0^\downarrow(s) = -sH - \frac{s^2}{2}V + \frac{s^3}{6}\frac{\partial^3 E_0^\downarrow(s)}{\partial s^3}\bigg|_{s=\bar{s}}, \tag{8.4}$$

doi:10.6342/NTU201800597

for some $\bar{s} \in [0, s]$. Now, let $s_n = a_n/V$. Then, $s_n \leq 1$ for all sufficiently large n by the assumption in Eq. (8.1) and $V > 0$. For all $s_n \leq 1$, Eq. (8.4) yields

$$\max_{0\leq s\leq 1}\left\{E_0^{\downarrow}(s) + sR_n\right\} \geq E_0^{\downarrow}(s_n) + s_n R_n \tag{8.5}$$

$$= \frac{a_n}{V}(-H + R_n) - \frac{a_n^2}{2V} + \frac{a_n^3}{6V^3}\left.\frac{\partial^3 E_0^{\downarrow}(s)}{\partial s^3}\right|_{s=\bar{s}_n} \tag{8.6}$$

$$= \frac{a_n^2}{2V} + \frac{a_n^3}{6V^3}\left.\frac{\partial^3 E_0^{\downarrow}(s)}{\partial s^3}\right|_{s=\bar{s}_n} \tag{8.7}$$

$$\geq \frac{a_n^2}{2V} - \frac{a_n^3}{6V^3}\left|\left.\frac{\partial^3 E_0^{\downarrow}(s)}{\partial s^3}\right|_{s=\bar{s}_n}\right| \tag{8.8}$$

$$\geq \frac{a_n^2}{2V} - \frac{a_n^3}{6V^3}\Upsilon, \tag{8.9}$$

where $\bar{s}_n \in [0, s_n]$; Eq. (8.7) holds since $R_n = H + a_n$; we denote by

$$\Upsilon = \max_{s\in[0,1]}\left|\frac{\partial^3 E_0^{\downarrow}(s)}{\partial s^3}\right|. \tag{8.10}$$

This quantity is finite due to the compact set $[0, 1]$ and the continuity, item (a) in Proposition 5.2. Therefore, substituting Eq. (8.9) into Eq. (8.3) gives for all sufficiently large $n \in \mathbb{N}$,

$$\frac{1}{na_n^2}\log\left(\frac{1}{\varepsilon^*(n, R_n)}\right) \geq -\frac{\log 4}{na_n^2} + \frac{1}{2V}\left(1 - \Upsilon\frac{a_n}{3V^2}\right). \tag{8.11}$$

Recall Eq. (1.7) and let $n \to +\infty$, which completes the achievability:

$$\liminf_{n\to+\infty}\frac{1}{na_n^2}\log\left(\frac{1}{\varepsilon^*(n, R_n)}\right) \geq \frac{1}{2V}. \tag{8.12}$$

We move on to show the converse, i.e. the "$\leq$" in Eq. (8.2). Let $N_1 \in \mathbb{N}$ be an integer such that $R_n = H + a_n \in (H_1(X|B)_\rho, H_0(X|B)_\rho)$ for all $n \in N_1$ Invoke the one-shot converse bound, Proposition 7.1, with $M = \exp\{nR_n\}$ to obtain for all $n \geq N_1$,

$$\log\left(\frac{1}{\varepsilon^*(n, R_n)}\right) \leq \min_{\sigma_B^n\in\mathcal{S}(B^n)} -\log\widehat{\alpha}_{\frac{M}{|\mathcal{X}^n|}}\left(\rho_{X^nB^n}\|\tau_{X^n}\otimes\sigma_B^n\right) \tag{8.13}$$

$$\leq -\log\widehat{\alpha}_{\frac{M}{|\mathcal{X}^n|}}\left(\rho_{X^nB^n}\|\tau_{X^n}\otimes(\sigma_{R_n}^\star)^{\otimes n}\right) \tag{8.14}$$

$$= -\log\widehat{\alpha}_{\frac{M}{|\mathcal{X}|^n}}\left(\rho_{XB}^{\otimes n}\|(\tau_X\otimes\sigma_{R_n}^\star)^{\otimes n}\right), \tag{8.15}$$

where we denote by $(\alpha_{R_n}^\star, \sigma_{R_n}^\star)$ the unique saddle-point of $\frac{1-\alpha}{\alpha}(R_n - H_\alpha^\uparrow(X|B)_\rho)$.

Next, we verify that we are able to employ Theorem 4.4 to asymptotically expand Eq. (8.15). Equation (8.27) in Proposition 8.1 below shows that $\lim_{n\to+\infty}\alpha_{R_n} = 1$. This together with the closed-

doi:10.6342/NTU201800597

form expression of $\sigma^\star_{R_n}$ [125], [119, Lemma 1], [8, Lemma 5.1] shows that

$$\lim_{n\to+\infty}\sigma^\star_{R_n}=\lim_{n\to+\infty}\frac{\left(\mathrm{Tr}_X\left[\rho_{XB}^{\alpha^\star_{R_n}}\right]\right)^{\frac{1}{\alpha^\star_{R_n}}}}{\mathrm{Tr}\left[\left(\mathrm{Tr}_X\left[\rho_{XB}^{\alpha^\star_{R_n}}\right]\right)^{\frac{1}{\alpha^\star_{R_n}}}\right]}=\rho_B. \tag{8.16}$$

Since $V=V(\rho\|\mathbb{1}_X\otimes\rho_B)>0$, by the continuity of $V(\cdot\|\cdot)$ (c.f. (3.55)), for every $\kappa\in(0,1)$ there exists $N_2\in\mathbb{N}$ such that for all $n\geq N_2$,

$$V\left(\rho_{XB}\|\tau_X\otimes\sigma^\star_{R_n}\right)=V\left(\rho_{XB}\|\mathbb{1}_X\otimes\sigma^\star_{R_n}\right)\geq(1-\kappa)V=:\nu>0. \tag{8.17}$$

Hence, we apply Theorem 4.4 with $r_n=\log|\mathcal{X}|-R_n$, $\rho=\rho_{XB}$ and $\sigma=\tau_X\otimes\sigma^\star_{R_n}$ to obtain for all $n\geq\max\{N_1,N_2\}$,

$$-\log\widehat{\alpha}_{\exp\{-nr_n\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)\leq n\sup_{\alpha\in(0,1]}\frac{1-\alpha}{\alpha}\left(D_\alpha\left(\rho\|\sigma\right)-r_n+\gamma_n\right)+\log\left(s^\star_n\sqrt{n}\right)+K, \tag{8.18}$$

$$=nE_{\mathrm{sp}}(H+a_n+\gamma_n)+\log\left(s^\star_n\sqrt{n}\right)+K, \tag{8.19}$$

for some constant $K>0$, and $s^\star_n:=(1-\alpha^\star_{R_n})/\alpha^\star_{R_n}$. Now, let $\delta_n:=a_n+\gamma_n$, and notice that $\gamma_n=O(\frac{\log n}{n})=o(a_n)$. We invoke Proposition 8.1 below to have

$$\limsup_{n\to+\infty}\frac{E_{\mathrm{sp}}(H(X|B)_\rho+\delta_n)}{a_n^2}=\limsup_{n\to+\infty}\frac{E_{\mathrm{sp}}(H(X|B)_\rho+\delta_n)}{\delta_n^2}\leq\frac{1}{2V}. \tag{8.20}$$

Moreover, Eq. (8.27) in Proposition 8.1 gives that $\lim_{n\to+\infty}\frac{s^\star_n}{\delta_n}=1/V$. Combining Eqs. (1.7), (8.15), (8.19) and (8.20) to conclude our claim

$$\limsup_{n\to+\infty}\frac{1}{na_n^2}\log\left(\frac{1}{\varepsilon^*(n,R_n)}\right)\leq\limsup_{n\to+\infty}-\frac{\log\widehat{\alpha}_{\exp\{-nr_n\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)}{na_n^2} \tag{8.21}$$

$$\leq\frac{1}{2V}+\limsup_{n\to+\infty}\frac{\log\left(s^\star_n\sqrt{n}\right)}{na_n^2} \tag{8.22}$$

$$=\frac{1}{2V}+\limsup_{n\to+\infty}\frac{\log\left(s^\star_n\sqrt{n}\right)}{n\delta_n^2} \tag{8.23}$$

$$=\frac{1}{2V}+\limsup_{n\to+\infty}\frac{\frac{1}{2}\log\left(n\delta_n^2\right)-\log V}{n\delta_n^2} \tag{8.24}$$

$$=\frac{1}{2V}, \tag{8.25}$$

where the last line follows from $\lim_{n\to+\infty}n\delta_n^2=+\infty$. Hence, Eq (8.12) together with Eq. (8.25) completes the proof.

doi:10.6342/NTU201800597

Proposition 8.1 (Error Exponent around Slepian-Wolf Limit). *Let $(\delta_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers with $\lim_{n\to+\infty}\delta_n = 0$. The following hold:*

$$\limsup_{n\to+\infty} \frac{E_{\mathrm{sp}}\left(H(X|B)_\rho + \delta_n\right)}{\delta_n^2} \leq \frac{1}{2V(X|B)_\rho}; \tag{8.26}$$

$$\limsup_{n\to+\infty} \frac{s_n^\star}{\delta_n} = \frac{1}{V(X|B)_\rho}, \tag{8.27}$$

where

$$s_n^\star := \arg\max_{s\geq 0} \left\{ s\left(H(X|B)_\rho + \delta_n\right) - sH^{\uparrow}_{\frac{1}{1+s}}(X|B)_\rho \right\}. \tag{8.28}$$

The proof of Proposition 8.1 is provided in Section 8.1.

□

8.1 Asymptotic Expansion of Error Exponent around Slepian-Wolf Limit

Proof of Proposition 8.1. For notational convenience, we denote by $H := H(X|B)_\rho$, $V := V(X|B)_\rho$.. Thus,

$$E_{\mathrm{sp}}(R) = \sup_{s\geq 0}\left\{sR + E_0(s)\right\}, \tag{8.29}$$

$$\tag{8.30}$$

Let a *critical rate* to be

$$r_{\mathrm{cr}} := \left.\frac{\partial E_0(s)}{\partial s}\right|_{s=1}. \tag{8.31}$$

Let N_0 be the smallest integer such that $H(X|B)_\rho + \delta_n < r_{\mathrm{cr}}$, for all $n \geq N_0$. Since the map $r \mapsto E_{\mathrm{sp}}(r)$ is non-increasing by item (a) in Proposition 5.3, the maximization over s in Eq. (8.29) can be restricted to the set $[0,1]$ for any rate below r_{cr}, i.e.,

$$E_{\mathrm{sp}}\left(H+\delta_n\right) = \max_{0\leq s\leq 1}\left\{s\left(H+\delta_n\right) + E_0(s)\right\}. \tag{8.32}$$

For every $n \in \mathbb{N}$, let $s_n^\star$ attain the maxima in Eq. (8.32) at a rate of $H+\delta_n$. It is not hard to observe that $s_n^\star > 0$ for all $n \geq N_0$ since $s_n^\star = 0$ if and only if $H + \delta_n < H$, which violates the assumption of $\delta_n > 0$ for finite n. Now, we will show Eq. (8.27) and

$$\lim_{n\to+\infty} s_n^\star = 0. \tag{8.33}$$

Let $(s_{n_k}^\star)_{k\in\mathbb{N}}$ be arbitrary subsequences. Since $[0,1]$ are compact, we may assume that

$$\lim_{k\to\infty} s_{n_k}^\star = s_o, \tag{8.34}$$

for some $s_o \in [0,1]$.

doi:10.6342/NTU201800597

Since $s \mapsto E_0(s)$ is strictly concave from item (c) in Proposition 5.1, the maximizer $s_n^\star$ must satisfy

$$\left.\frac{\partial E_0(s)}{\partial s}\right|_{s=s_{n_k}^\star} = -(H + \delta_{n_k}), \tag{8.35}$$

which together with item (a) in Proposition 5.1 implies

$$\lim_{k\to+\infty} \left.\frac{\partial E_0(s)}{\partial s}\right|_{s=s_{n_k}^\star} = \left.\frac{\partial E_0(s)}{\partial s}\right|_{s=s_o} = -H. \tag{8.36}$$

On the other hand, item (d) in Proposition 5.1 gives

$$\left.\frac{\partial E_0(s)}{\partial s}\right|_{s=0} = -H. \tag{8.37}$$

Since item (d) in Proposition 5.1 guarantees

$$\left.\frac{\partial^2 E_0(s)}{\partial s^2}\right|_{s=0} = -V < 0, \tag{8.38}$$

which implies that the first-order derivative $\partial E_0(s)/\partial s$ is strictly decreasing around $s = 0$. Hence, we conclude $s_o = 0$. Because the subsequence is arbitrary, Eq. (8.34) is shown.

Next, from Eqs. (8.35) and Eqs. (8.37), the mean value theorem states that there exists a number $\hat{s}_{n_k} \in \left(0, s_{n_k}^\star\right)$, for each $k \in \mathbb{N}$, such that

$$-\left.\frac{\partial^2 E_0(s)}{\partial s^2}\right|_{s=\hat{s}_{n_k}} = \frac{-H + (H + \delta_{n_k})}{s_{n_k}^\star} = \frac{\delta_{n_k}}{s_{n_k}^\star}. \tag{8.39}$$

When k approaches infinity, items (a) and (e) in Proposition 5.1 give

$$\lim_{k\to+\infty} \left.\frac{\partial^2 E_0(s)}{\partial s^2}\right|_{s=\hat{s}_{n_k}} = \left.\frac{\partial^2 E_0(s)}{\partial s^2}\right|_{s=0} = -V. \tag{8.40}$$

Combining Eqs. (8.39) and (8.40) leads to

$$\lim_{k\to+\infty} \frac{s_{n_k}^\star}{\delta_{n_k}} = \frac{1}{V}. \tag{8.41}$$

Since the subsequence was arbitrary, the above result establishes Eq. (8.27).

Finally, denote by

$$\Upsilon = \max_{s\in[0,1]} \left|\frac{\partial^3 E_0(s)}{\partial s^3}\right| < +\infty. \tag{8.42}$$

For every sufficiently large $n \geq N_0$, we apply Taylor's theorem to the map $s_n^\star \mapsto E_0\left(s_n^\star\right)$ at the original

doi:10.6342/NTU201800597

point to obtain

$$E_{\mathrm{sp}}(H+\delta_n) = s_n^\star (H+\delta_n) + E_0(s_n^\star) \tag{8.43}$$

$$= s_n^\star \delta_n - \frac{(s_n^\star)^2}{2} V + \frac{(s_n^\star)^3}{6} \left. \frac{\partial^3 E_0(s, P_n)}{\partial s^3} \right|_{s=\bar{s}_n} \tag{8.44}$$

$$\leq s_n^\star (H+\delta_n - H) - \frac{(s_n^\star)^2}{2} V + \frac{(s_n^\star)^3 \Upsilon}{6} \tag{8.45}$$

$$\leq \sup_{s \geq 0} \left\{ s\delta_n - \frac{s^2}{2} V \right\} + \frac{(s_n^\star)^3 \Upsilon}{6} \tag{8.46}$$

$$= \frac{\delta_n^2}{2V} + \frac{(s_n^\star)^3 \Upsilon}{6}, \tag{8.47}$$

where $\bar{s}_n$ is some number in $[0, s_n^\star]$. Then, Eqs. (8.27), (8.34), (8.47), and the assumption $\lim_{n\to+\infty} \delta_n = 0$ imply that the desired inequality

$$\limsup_{n\to+\infty} \frac{E_{\mathrm{sp}}(H+\delta_n)}{\delta_n^2} \leq \frac{1}{2V}. \tag{8.48}$$

□

doi:10.6342/NTU201800597

Part III

Information Transmission over a Quantum Channel

doi:10.6342/NTU201800597

Chapter 9

Error Exponent Functions (Channel Coding)

In this chapter, we introduce the auxiliary functions and the exponent functions for classical-quantum channel coding. Thee major properties of those functions are provided in Sections 9.2 and 9.3. Section 9.1 presents the variational representations for the weak sphere-packing exponent.

The *random coding exponent* [35] and *strong sphere-packing exponent* [38] of a c-q channel $\mathcal{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ and a rate $R \geq 0$ are defined by

$$E_{\mathrm{r}}(R) := \max_{P \in \mathcal{P}(\mathcal{X})} E_{\mathrm{r}}(R, P), \tag{9.1}$$

$$E_{\mathrm{sp}}(R) := \max_{P \in \mathcal{P}(\mathcal{X})} E_{\mathrm{sp}}(R, P), \tag{9.2}$$

where

$$E_{\mathrm{r}}(R, P) := \sup_{s \geq 0} \{E_0(s, P) - sR\}, \tag{9.3}$$

$$E_{\mathrm{sp}}(R, P) := \sup_{s \geq 0} \{E_0(s, P) - sR\}, \tag{9.4}$$

and E_0 is the *auxiliary function* of the c-q channel $\mathcal{W}$ (see [34, 35, 36]):

$$E_0(s, P) := -\log \mathrm{Tr} \left[\left(\sum_{x \in \mathcal{X}} P(x) \cdot W_x^{1/(1+s)} \right)^{1+s} \right] \tag{9.5}$$

for all $P \in \mathcal{P}(\mathcal{X})$ and $s \geq 0$.

We will require three variants of the above auxiliary function: $\forall s \geq 0$ and $\sigma \in \mathcal{S}(\mathcal{H})$,

$$E_0^{\downarrow}(s, P, \sigma) := sD_{1-s}\left(P \circ \mathcal{W} \| P \otimes \sigma\right) \tag{9.6}$$

$$E_{\mathrm{h}}(s, P, \sigma) := sD_{\frac{1}{1+s}}\left(\mathcal{W} \| \sigma | P\right), \tag{9.7}$$

$$E_{\mathrm{h}}^{\flat}(s, P, \sigma) := sD^{\flat}_{\frac{1}{1+s}}\left(\mathcal{W} \| \sigma | P\right). \tag{9.8}$$

doi:10.6342/NTU201800597

With this, we define another version fo the random coding exponent:

$$E_{\mathrm{r}}^{\downarrow}(R,P) := \sup_{0\le s\le 1}\left\{E_0^{\downarrow}(s,P,P\mathcal{W}) - sR\right\} \tag{9.9}$$

This quantity will appear in the achievability (see Theorem 12.1 in Chapter 10), and Chapter 12.

The *weak sphere-packing exponent* [37] is defined as

$$\widetilde{E}_{\mathrm{sp}}(R) := \max_{P\in\mathcal{P}(\mathcal{X})} \widetilde{E}_{\mathrm{sp}}(R,P), \tag{9.10}$$

where

$$\widetilde{E}_{\mathrm{sp}}(R,P) := \min_{\mathcal{V}:\mathcal{X}\to\mathcal{S}(\mathcal{H})} \left\{D\left(\mathcal{V}\|\mathcal{W}|P\right) : I(P,\mathcal{V}) \le R\right\}. \tag{9.11}$$

We also need the following definitions: for any $R \ge 0$ and $P \in \mathcal{P}(\mathcal{X})$,

$$E_{\mathrm{sp}}^{(1)}(R,P) := \sup_{0<\alpha\le 1} \frac{1-\alpha}{\alpha}\left(I_\alpha^{(1)}(P,\mathcal{W}) - R\right); \tag{9.12}$$

$$E_{\mathrm{sp}}^{(2)}(R,P) := \sup_{0<\alpha\le 1} \frac{1-\alpha}{\alpha}\left(I_\alpha^{(2)}(P,\mathcal{W}) - R\right), \tag{9.13}$$

Eq. (3.62) implies that (see also Theorem 9.1) $E_{\mathrm{sp}}^{(1)}(R,P) \le E_{\mathrm{sp}}^{(2)}(R,P)$. By quantum Sibson's identity [125], one finds

$$E_{\mathrm{sp}}^{(1)}(R,P) = E_{\mathrm{sp}}(R,P). \tag{9.14}$$

Proposition 3.2 and Eq. (3.63) imply that the two quantities given in Eqs. (9.12) and (9.13) are equal to the strong sphere-packing exponent by maximizing over the input distributions:

$$E_{\mathrm{sp}}(R) = \max_{P\in\mathcal{P}(\mathcal{X})} E_{\mathrm{sp}}^{(1)}(R,P) = \max_{P\in\mathcal{P}(\mathcal{X})} E_{\mathrm{sp}}^{(2)}(R,P). \tag{9.15}$$

Further, we define [25, p. 152], [38, Theorem 6]:

$$R_\infty := C_{0,\mathcal{W}}. \tag{9.16}$$

From the definitions in Eqs. (3.54) and (9.16), it can be verified that $R_\infty \le C_{\mathcal{W}}$ for all c-q channels $\mathcal{W}$. In Proposition 9.6 below, one has $E_{\mathrm{sp}}(R) = +\infty$ for $R < R_\infty$, and $E_{\mathrm{sp}}(R) = 0$ as $R > C_{\mathcal{W}}$. Throughout this paper, we further assume that the considered c-q channel $\mathcal{W}$ satisfies $R_\infty < C_{\mathcal{W}}$.

Lastly, we define

$$\widetilde{E}_{\mathrm{sp}}(R,P,\sigma) := \min_{\bar{\mathcal{W}}:\mathcal{X}\to\mathcal{S}_\circ} \left\{D\left(\bar{\mathcal{W}}\|\mathcal{W}|P\right) : D\left(\bar{\mathcal{W}}\|\sigma|P\right) \le R\right\} \tag{9.17}$$

for all $R > 0$, $P \in \mathcal{P}(\mathcal{X})$, and $\sigma \in \mathcal{S}_{>0}(\mathcal{H})$. From the definitions in Eq. (9.17), it is not hard to see that [86]

$$\widetilde{E}_{\mathrm{sp}}(R,P,\sigma) = 0, \quad \forall R \ge D\left(\mathcal{W}\|\sigma|P\right). \tag{9.18}$$

doi:10.6342/NTU201800597

and

$$E_{\mathrm{sp}}^{(2)}(R,P,\sigma)=\begin{cases}+\infty, & R<D_0\left(\mathcal{W}\|\sigma|P\right),\\ 0, & R\geq D\left(\mathcal{W}\|\sigma|P\right).\end{cases} \tag{9.19}$$

As we will show in Chapter 11, the quantity $E_{\mathrm{sp}}^{(2)}(R,P)$ plays a significant role in the connection between hypothesis testing and channel coding. Moreover, Proposition 9.5 in Section 9.3 below shows that the the optimizer in Eqs. (3.61) and (9.13) forms a saddle-point.

9.1 Variational Representations

This section derives alternative formulations of the strong and weak sphere-packing exponents of Eqs. (9.4)-(9.17), and provides a relation between these two exponents. As we will show later, the derived formulations are essentially optimization problems in the primal domain, while the expressions in Eqs. (9.4) and (9.17) are corresponding dual representations.

We first consider the following convex optimization problem and then exploit it to establish variational formulations of the sphere-packing exponents. Let $\rho,\tau\in\mathcal{S}(\mathcal{H})$ be two density operators. Consider the following convex optimization problem:

$$\begin{aligned}\text{(P)}\quad & e(r):=\inf_{\sigma\in\mathcal{S}(\mathcal{H})} D\left(\sigma\|\rho\right),\\ & \text{subject to}\quad D\left(\sigma\|\tau\right)\leq r.\end{aligned} \tag{9.20}$$

The above primal problem is interpreted as finding the optimal operator $\sigma^\star$ that achieves the minimum relative entropy $e(r)$ to ρ, within r-radius to τ. The following result shows the dual representation of problem (P) via Lagrangian duality.

Lemma 9.1 ([93, Section 3.7], [111], [58, Theorem III.5]). *The dual problem of* (P) *is given by*

$$\text{(D)}\quad \sup_{s\geq 0}\left\{-(1+s)\log Q_{\frac{1}{1+s}}^{\flat}(\rho\|\tau)-sr\right\}. \tag{9.21}$$

Proof. By the method of Lagrange multipliers, the primal problem in Eq. (9.20) can be rewritten as

$$\sup_{s\geq 0}\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{D(\sigma\|\rho)+s\left(D(\sigma\|\tau)-r\right)\right\} \tag{9.22}$$

$$=\sup_{s\geq 0}\left\{(1+s)\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\frac{1}{1+s}D(\sigma\|\rho)+\frac{s}{1+s}D(\sigma\|\tau)\right\}-sr\right\} \tag{9.23}$$

$$=\sup_{s\geq 0}\left\{-(1+s)\log Q_{\frac{1}{1+s}}^{\flat}(\rho\|\tau)-sr\right\}, \tag{9.24}$$

where the last equality follows from Lemma 3.2. □

doi:10.6342/NTU201800597

Theorem 9.1 (Variational Representations of the Sphere-Packing Exponents). *Let $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ be a classical-quantum channel. For any $R>R_\infty$, we have*

$$\widetilde{E}_{\mathrm{sp}}(R,P)=\sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\frac{1-\alpha}{\alpha}\left(D_\alpha^\flat\left(\mathcal{W}\|\sigma|P\right)-R\right)\right\},\quad \textit{and} \tag{9.25}$$

$$E_{\mathrm{sp}}(R,P)\le\sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\frac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{W}\|\sigma|P\right)-R\right)\right\}, \tag{9.26}$$

where $\widetilde{E}_{\mathrm{sp}}(R,P)$ and $E_{\mathrm{sp}}(R,P)$ are defined in Eqs. (9.17) and (9.4), respectively.

Moreover, equality in Eq. (9.26) is attained when maximizing over all prior distributions, i.e.,

$$E_{\mathrm{sp}}(R)=\max_{P\in\mathcal{P}(\mathcal{X})}E_{sp}(R,P)=\max_{P\in\mathcal{P}(\mathcal{X})}\sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\frac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{W}\|\sigma|P\right)-R\right)\right\}. \tag{9.27}$$

Proof. We start with the proof of Eq. (9.25). Observe that

$$\min_{\sigma\in\mathcal{S}(\mathcal{H})}D\left(\mathcal{V}\|\sigma|P\right)=\min_{\sigma\in\mathcal{S}(\mathcal{H})}\sum_{x\in\mathcal{X}}P(x)\operatorname{Tr}\left[V_x\left(\log V_x-\log\sigma\right)\right] \tag{9.28}$$

$$=I(P,\mathcal{V}). \tag{9.29}$$

We find

$$\widetilde{E}_{\mathrm{sp}}(R,P)=\min_{\mathcal{V}:\mathcal{X}\to\mathcal{S}(\mathcal{H})}\left\{D\left(\mathcal{V}\|\mathcal{W}|P\right):I(P,\mathcal{V})\le R\right\} \tag{9.30}$$

$$=\min_{\mathcal{V}:\mathcal{X}\to\mathcal{S}(\mathcal{H})}\left\{D\left(\mathcal{V}\|\mathcal{W}|P\right):\min_{\sigma\in\mathcal{S}(\mathcal{H})}D\left(\mathcal{V}\|\sigma|P\right)\le R\right\} \tag{9.31}$$

$$=\sup_{s\ge 0}\min_{\mathcal{V}:\mathcal{X}\to\mathcal{S}(\mathcal{H})}\left\{D\left(\mathcal{V}\|\mathcal{W}|P\right)+s\left(\min_{\sigma\in\mathcal{S}(\mathcal{H})}D\left(\mathcal{V}\|\sigma|P\right)-R\right)\right\} \tag{9.32}$$

$$=\sup_{s\ge 0}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\min_{\mathcal{V}:\mathcal{X}\to\mathcal{S}(\mathcal{H})}\left\{-sR+\sum_{x\in\mathcal{X}}P(x)D\left(V_x\|W_x\right)+s\cdot D\left(V_x\|\sigma\right)\right\} \tag{9.33}$$

$$=\sup_{s\ge 0}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\sum_{x\in\mathcal{X}}P(x)\min_{V_x\in\mathcal{S}(\mathcal{H})}\left[D\left(V_x\|W_x\right)+s\cdot D\left(V_x\|\sigma\right)-sR\right]\right\} \tag{9.34}$$

$$=\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\sum_{x\in\mathcal{X}}P(x)\min_{V_x\in\mathcal{S}(\mathcal{H})}\left\{D\left(V_x\|W_x\right):D\left(V_x\|\sigma\right)\le R\right\}\right\}. \tag{9.35}$$

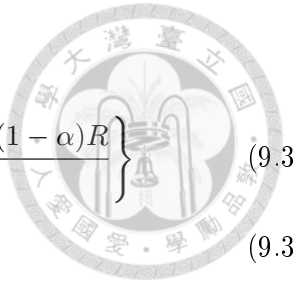
In Eq. (9.32) we introduced the constraint into the objective function via the Lagrange multiplier $s\ge 0$; and Eq. (9.34) follows from the linearity of the convex combination. By Lemma 9.1, the inner minimum over $V_x\in\mathcal{S}(\mathcal{H})$ can be represented as its dual problem:

$$\widetilde{E}_{\mathrm{sp}}(R,P)=\min_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{s\ge 0}\left\{-(1+s)\sum_{x\in\mathcal{X}}P(x)\log\left[Q_{\frac{1}{1+s}}^\flat\left(W_x\|\sigma\right)\right]-sR\right\} \tag{9.36}$$

$$=\min_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{0<\alpha\le 1}\left\{\frac{-\sum_{x\in\mathcal{X}}P(x)\log\left[Q_\alpha^\flat\left(W_x\|\sigma\right)\right]-(1-\alpha)R}{\alpha}\right\}, \tag{9.37}$$

where we substitute $\alpha=1/(1+s)$. From Lemma 3.2, the numerator in the bracket of Eq. (9.37) is a concave-convex saddle function for every $\sigma\in\mathcal{S}(\mathcal{H})$ and every $\alpha\in(0,1]$. Hence, we invoke the minimax

doi:10.6342/NTU201800597



theorem, Lemma 9.2 below, to exchange the order of min-sup in Eq. (9.37):

$$\widetilde{E}_{\mathrm{sp}}(R,P) = \sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\frac{-\sum_{x\in\mathcal{X}}P(x)\log\left[Q_\alpha^\flat\left(W_x\|\sigma\right)\right]-(1-\alpha)R}{\alpha}\right\} \tag{9.38}$$

$$= \sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\frac{1-\alpha}{\alpha}\left(D_\alpha^\flat\left(\mathcal{W}\|\sigma|P\right)-R\right)\right\}, \tag{9.39}$$

where in (9.39) we recall the definition of the log-Euclidean α-Rényi divergence, Eq. (3.6), and hence prove the first claim in Eq. (9.25).

Next, we will prove Eq. (9.26). From Jensen's inequality and the concavity of the logarithm, the right-hand side of Eq. (9.26) implies that

$$\sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{\frac{1-\alpha}{\alpha}\left(\sum_{x\in\mathcal{X}}P(x)D_\alpha\left(W_x\|\sigma\right)-R\right)\right\} \tag{9.40}$$

$$= \sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{-\frac{1}{\alpha}\sum_{x\in\mathcal{X}}P(x)\log\mathrm{Tr}\left[W_x^\alpha\sigma^{1-\alpha}\right]-\frac{1-\alpha}{\alpha}R\right\} \tag{9.41}$$

$$\ge \sup_{0<\alpha\le 1}\min_{\sigma\in\mathcal{S}(\mathcal{H})}\left\{-\frac{1}{\alpha}\log\mathrm{Tr}\left[\sum_{x\in\mathcal{X}}P(x)\left[W_x^\alpha\sigma^{1-\alpha}\right]\right]-\frac{1-\alpha}{\alpha}R\right\} \tag{9.42}$$

$$= E_{\mathrm{sp}}(R,P), \tag{9.43}$$

where the last equality follows from Eq. (9.14).

Finally, Eq. (9.27) follows from the following identity proved by Mosonyi and Ogawa [58, Proposition IV.2]:

$$\max_{P\in\mathcal{P}(\mathcal{X})}\min_{\sigma\in\mathcal{S}(\mathcal{H})}D_\alpha\left(\mathcal{W}\|\sigma|P\right) = \max_{P\in\mathcal{P}(\mathcal{X})}\min_{\sigma\in\mathcal{S}(\mathcal{H})}D_\alpha\left(P\circ\mathcal{W}\|P\otimes\sigma\right), \tag{9.44}$$

Note that the above relation also holds for $D_\alpha^\flat$.

Lemma 9.2 ([104, Proposition 21]). *Let $\mathcal{A}\subset\mathbb{R}_{\ge 0}$ be a convex set and let $\mathcal{B}$ be a compact Hausdorff space. Further, let $f:\mathcal{A}\times\mathcal{B}\to\mathbb{R}$ be concave on $\mathcal{A}$ as well as convex on $\mathcal{B}$. Then*

$$\sup_{x\in\mathcal{A}}\inf_{y\in\mathcal{B}}\frac{f(x,y)}{x} = \inf_{y\in\mathcal{B}}\sup_{x\in\mathcal{A}}\frac{f(x,y)}{x}. \tag{9.45}$$

□

The following corollary is a simple consequence of the variational representations of the sphere-packing exponents in Theorem 9.1 and the Golden-Thompson inequality, Lemma 2.7.

Corollary 9.1. *For any classical-quantum channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, $R>R_\infty$, and $P\in\mathcal{P}(\mathcal{X})$, it holds that*

$$E_{\mathrm{sp}}(R,P)\le\widetilde{E}_{\mathrm{sp}}(R,P). \tag{9.46}$$

doi:10.6342/NTU201800597

9.2 Properties of Auxiliary Functions

In the following, we list the properties of the auxiliary functions E_0, $E_0^{\downarrow}$, E_{h}, and $E_{\mathrm{h}}^{\flat}$ in Propositions 9.1, 9.2, 9.3, and 9.4, respectively. Our ingredients come from properties of Petz's quantum Rényi divergence [59] (see also [126, 120, 8]) and the theory of matrix geometric means.

Proposition 9.1 (Properties of $E_0(s,P)$)**.** *The auxiliary function $E_0(s,P)$, defined in Eq. (9.5), admits the following properties.*

(a) The partial derivatives $\partial E_0(s,P)/\partial s$, $\partial^2 E_0(s,P)/\partial s^2$, $\partial^3 E_0(s,P)/\partial s^3$, and $E_0(s,P)$ are all continuous for $(s,P) \in \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{X})$.

(b) For every $P \in \mathcal{P}(\mathcal{X})$, the function $E_0(s,P)$ is concave in s for all $s \in \mathbb{R}_{\geq 0}$.

(c) For every $P \in \mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial E_0(s,P)}{\partial s}\right|_{s=0} = I(P,\mathcal{W}). \tag{9.47}$$

(d) For every $P \in \mathcal{P}(\mathcal{X})$,

$$\lim_{s\to+\infty} \frac{\partial E_0(s,P)}{\partial s} \leq \frac{\partial E_0(s,P)}{\partial s} \leq I(P,\mathcal{W}),\ \forall s \in \mathbb{R}_{\geq 0}. \tag{9.48}$$

(e) For every $P \in \mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial^2 E_0(s,P)}{\partial s^2}\right|_{s=0} = -V(P,\mathcal{W}). \tag{9.49}$$

The proof is provided in Section 9.2.1.

doi:10.6342/NTU201800597

Proposition 9.2 (Properties of $E_0^{\downarrow}(s,P,\sigma)$). *Consider a classical-quantum channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, a distribution $P\in\mathcal{P}(\mathcal{X})$, and a state $\sigma\in\mathcal{S}(\mathcal{H})$ with $W_x\ll\sigma$ for all $x\in\mathtt{supp}(P)$. Then $E_0^{\downarrow}(s,P,\sigma)$ defined in Eq. (9.6) enjoys the following properties.*

(a) $E_0^{\downarrow}(s,P,\sigma)$ and its partial derivatives $\partial E_0^{\downarrow}(s,P,\sigma)/\partial s$, $\partial^2 E_0^{\downarrow}(s,P,\sigma)/\partial s^2$, $\partial^3 E_0^{\downarrow}(s,P,\sigma)/\partial s^3$ are all continuous in $(s,P)\in\mathbb{R}_{\geq 0}\times\mathcal{P}(\mathcal{X})$.

(b) For every $P\in\mathcal{P}(\mathcal{X})$, the function $E_0^{\downarrow}(s,P,\sigma)$ is concave in $s\in\mathbb{R}_{\geq 0}$.

(c) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s}\right|_{s=0}=D\left(P\circ\mathcal{W}\|P\otimes\sigma\right). \tag{9.50}$$

(d) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\lim_{s\to+\infty}\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s}\leq\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s}\leq D\left(P\circ\mathcal{W}\|P\otimes\sigma\right),\ \forall s\in\mathbb{R}_{\geq 0}. \tag{9.51}$$

(e) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial^2 E_0^{\downarrow}(s,P,\sigma)}{\partial s^2}\right|_{s=0}=-V\left(P\circ\mathcal{W}\|P\otimes\sigma\right). \tag{9.52}$$

The proof is provided in Section 9.2.2.

Properties of E_{h} and $E_{\mathrm{h}}^{\flat}$ will be crucial in the analysis of the converse part of our main result.

Proposition 9.3 (Properties of $E_{\mathrm{h}}(s,P,\sigma)$). *Consider a classical-quantum channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, a distribution $P\in\mathcal{P}(\mathcal{X})$, and a state $\sigma\in\mathcal{S}(\mathcal{H})$ with $W_x\ll\sigma$ for all $x\in\mathtt{supp}(P)$. Then $E_{\mathrm{h}}(s,P,\sigma)$ defined in Eq. (9.7) enjoys the following properties.*

(a) $E_{\mathrm{h}}(s,P,\sigma)$ and its partial derivatives $\partial E_{\mathrm{h}}(s,P,\sigma)/\partial s$, $\partial^2 E_{\mathrm{h}}(s,P,\sigma)/\partial s^2$, $\partial^3 E_{\mathrm{h}}(s,P,\sigma)/\partial s^3$ are continuous for $(s,P)\in\mathbb{R}_{\geq 0}\times\mathcal{P}(\mathcal{X})$.

(b) For every $P\in\mathcal{P}(\mathcal{X})$, the function $E_{\mathrm{h}}(s,P,\sigma)$ is concave in s for all $s\in\mathbb{R}_{\geq 0}$.

(c) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial E_{\mathrm{h}}(s,P,\sigma)}{\partial s}\right|_{s=0}=D\left(\mathcal{W}\|\sigma|P\right). \tag{9.53}$$

(d) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\lim_{s\to+\infty}\frac{\partial E_{\mathrm{h}}(s,P,\sigma)}{\partial s}\leq\frac{\partial E_{\mathrm{h}}(s,P,\sigma)}{\partial s}\leq D\left(\mathcal{W}\|\sigma|P\right),\ \forall s\in\mathbb{R}_{\geq 0}. \tag{9.54}$$

(e) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial^2 E_{\mathrm{h}}(s,P,\sigma)}{\partial s^2}\right|_{s=0}=-V\left(\mathcal{W}\|\sigma|P\right). \tag{9.55}$$

doi:10.6342/NTU201800597

The proof is provided in Section 9.2.3.

Proposition 9.4 (Properties of $E_{\mathrm{h}}^{\flat}(s,P,\sigma)$). *Consider a classical-quantum channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, a distribution $P\in\mathcal{P}(\mathcal{X})$, and a state $\sigma\in\mathcal{S}(\mathcal{H})$ with $W_x\ll\sigma$ for all $x\in\mathtt{supp}(P)$. Then $E_{\mathrm{h}}^{\flat}(s,P,\sigma)$ defined in Eq. (9.8) enjoys the following properties.*

(a) $E_{\mathrm{h}}^{\flat}(s,P,\sigma)$ and its partial derivatives $\partial E_{\mathrm{h}}^{\flat}(s,P,\sigma)/\partial s$, $\partial^2 E_{\mathrm{h}}^{\flat}(s,P,\sigma)/\partial s^2$, $\partial^3 E_{\mathrm{h}}^{\flat}(s,P,\sigma)/\partial s^3$ are all continuous for $(s,P)\in\mathbb{R}_{\geq 0}\times\mathcal{P}(\mathcal{X})$.

(b) For every $P\in\mathcal{P}(\mathcal{X})$, the function $E_{\mathrm{h}}^{\flat}(s,P,\sigma)$ is concave in s for all $s\in\mathbb{R}_{\geq 0}$.

(c) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s}\right|_{s=0}=D\left(\mathcal{W}\|\sigma|P\right). \tag{9.56}$$

(d) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\lim_{s\to+\infty}\frac{\partial E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s}\leq\frac{\partial E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s}\leq D\left(\mathcal{W}\|\sigma|P\right),\ \forall s\in\mathbb{R}_{\geq 0}. \tag{9.57}$$

(e) For every $P\in\mathcal{P}(\mathcal{X})$,

$$\left.\frac{\partial^2 E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s^2}\right|_{s=0}=-\widetilde{V}\left(\mathcal{W}\|\sigma|P\right). \tag{9.58}$$

The proof is provided in Section 9.2.4.

9.2.1 Proof of Proposition 9.1

Fix any c-q channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$. To ease the burden of derivations, we denote some notation:

$$f(s,P):=\sum_{x\in\mathcal{X}}P(x)W_x^{1/(1+s)}\in\mathcal{B}(\mathcal{H})_+, \tag{9.59}$$

$$g(s,P):=f(s,P)^{(1+s)}\in\mathcal{B}(\mathcal{H})_+, \tag{9.60}$$

$$F(s,P):=\mathrm{Tr}\left[g(s,P)\right]\in\mathbb{R}_{\geq 0}, \tag{9.61}$$

for all $(s,P)\in\mathbb{R}_{\geq 0}\times\mathcal{P}(\mathcal{X})$. Clearly, $f(\cdot,\cdot)$ is continuous on $\mathbb{R}_{\geq 0}\times\mathcal{P}(\mathcal{X})$. Direct calculation shows that

$$f'(s,P):=\frac{\partial f(s,P)}{\partial s}=-\frac{1}{(1+s)^2}\sum_{x\in\mathcal{X}}P(x)W_x^{1/(1+s)}\widehat{\log}W_x, \tag{9.62}$$

$$f''(s,P):=\frac{\partial^2 f(s,P)}{\partial s^2}=\frac{1}{(1+s)^3}\sum_{x\in\mathcal{X}}P(x)W_x^{1/(1+s)}\widehat{\log}W_x\left[2+\frac{\widehat{\log}W_x}{(1+s)}\right], \tag{9.63}$$

$$f'''(s,P):=\frac{\partial^3 f(s,P)}{\partial s^3}=-\frac{1}{(1+s)^4}\sum_{x\in\mathcal{X}}P(x)W_x^{1/(1+s)}\widehat{\log}W_x\left[6+\frac{6\widehat{\log}W_x}{(1+s)}+\frac{\widehat{\log}^2W_x}{(1+s)^2}\right], \tag{9.64}$$

doi:10.6342/NTU201800597

where we denote $\widehat{\log}$ by

$$\widehat{\log} x = \begin{cases} \log x, & x > 0, \\ 0, & x = 0. \end{cases} \tag{9.65}$$

From Eqs. (9.62), (9.63), and (9.64), we infer that $f'(s,P)$, $f''(s,P)$, and $f'''(s,P)$ share the same support as $f(s,P)$, and are continuous for all $(s,P) \in \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{X})$ (in the strong topology).

Observe that for all $(s,P) \in \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{X})$,

$$g(s,P)^0 = f(s,P)^0. \tag{9.66}$$

Hence, the operator $g(s,P)$ admit the expression:

$$g(s,P) = g(s,P)^0 \mathrm{e}^{(1+s)\widehat{\log} f(s,P)} g(s,P)^0. \tag{9.67}$$

By applying the chain rule of the Fréchet derivatives, one can calculate that

$$\begin{aligned} g'(s,P) &:= \frac{\partial g(s,P)}{\partial s} \\ &= g(s,P)^0 \mathsf{D} \exp\left[\widehat{\log} g(s,P)\right] \left((1+s)\mathsf{D}\widehat{\log}\left[f(s,P)\right]\left(f'(s,P)\right) + \widehat{\log} f(s,P)\right) g(s,P)^0, \end{aligned} \tag{9.68}$$

$$\begin{aligned} g''(s,P) &:= \frac{\partial^2 f(s,P)}{\partial s^2} \\ &= g(s,P)^0 \mathsf{D}^2 \exp\left[\widehat{\log} g(s,P)\right] \left((1+s)\mathsf{D}\widehat{\log}\left[f(s,P)\right]\left(f'(s,P)\right) + \widehat{\log} f(s,P)\right) g(s,P)^0 \\ &+ g(s,P)^0 \mathsf{D} \exp\left[\widehat{\log} g(s,P)\right] \Big(2\mathsf{D}\widehat{\log}\left[f(s,P)\right]\left(f'(s,P)\right) + (1+s)\Big\{\mathsf{D}\widehat{\log}\left[f(s,P)\right]\left(f''(s,P)\right) \\ &+ \mathsf{D}^2\widehat{\log}\left[f(s,P)\right]\left(f'(s,P)\right)\Big\}\Big) g(s,P)^0, \end{aligned} \tag{9.69}$$

where we use the following integral formulas (see e.g. [82, Example 3.22, Excersize 3.24])

$$\mathsf{D}\widehat{\log}[A](B) = \int_0^{+\infty} (t\mathbb{1} + A)^{-1} B (t\mathbb{1} + A)^{-1} \, \mathrm{d}t, \tag{9.70}$$

$$\mathsf{D}^2\widehat{\log}[A](B) := \mathsf{D}^2\widehat{\log}[A](B,B) = -2\int_0^{+\infty} (t\mathbb{1} + A)^{-1} B (t\mathbb{1} + A)^{-1} B (t\mathbb{1} + A)^{-1} \, \mathrm{d}t \tag{9.71}$$

for all $0 \leq B \ll A$, and (see e.g. [82, Theorem 3.10])

$$\mathsf{D}\exp[A](B) = \int_0^1 \mathrm{e}^{(1-t)A} B \mathrm{e}^{tA} \, \mathrm{d}t, \tag{9.72}$$

$$\mathsf{D}^2\exp[A](B) := \mathsf{D}^2\exp[A](B,B) = 2\int_0^1 \int_0^{t_1} \mathrm{e}^{(1-t_1)A} B \mathrm{e}^{(t_1-t_2)A} B \mathrm{e}^{t_2 A} \, \mathrm{d}t_2 \, \mathrm{d}t_1 \tag{9.73}$$

for all self-adjoint operators A and B. Further, by [77, Theorem 3.5] $\mathsf{D}\exp[\cdot](\cdot)$, $\mathsf{D}^2\exp[\cdot](\cdot)$ are continuous for all self-adjoint operators, and $\mathsf{D}\widehat{\log}[A](B)$, $\mathsf{D}^2\widehat{\log}[A](B)$ are continuous for all $0 \leq B \ll A$.

In the following, we will show that $g'(s,P)$ is continuous for all $(s,P) \in \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{X})$. However, the operation $\mathsf{D}\widehat{\log}[\cdot](\cdot)$ in Eq. (9.68) is only continuous for positive definite operators (see [69, Theorem

doi:10.6342/NTU201800597

3.8]). We need to do little more work to circumvent this problem.

Let $\{s_k, P_k\}_{k\geq 1}$ be an arbitrary sequence with limit $(s_k, P_k) \to (s_o, P_o)$. Observe that if $f(s_k, P_k) \ll f(s_o, P_o)$ for some $k \in \mathbb{N}$, we can only focus on the support of $f(s_o, P_o)$ and treat $f(s_o, P_o)$ as a positive definite operator without loss of generality. Consider any subsequence $\{s_{k_n}, P_{k_n}\}_{n\geq 1}$. Suppose all but a finite number of (s_{k_n}, P_{k_n}) satisfy

$$f(s_{k_n}, P_{k_n}) \ll f(s_o, P_o). \tag{9.74}$$

Then Eq. (9.68), and the continuity of $f(\cdot,\cdot)$, $f'(\cdot,\cdot)$, $\widehat{\mathsf{Dlog}}[f(\cdot,\cdot)](\,\cdot\,)$ (recall that it is continuous for positive definite operators), and $\mathsf{D}\exp[\widehat{\log} f(\cdot,\cdot)](\,\cdot\,)$ (see [69, Theorem 3.8], [77, Theorem 3.5]) imply that $g'(\cdot,\cdot)$ is continuous at (s_o, P_o). If this is not the case, we define

$$\omega_{\min} := \min_{x\in\mathcal{X}} \widetilde{\lambda}_{\min}(W_x), \tag{9.75}$$

$$\omega_{\max} := \max_{x\in\mathcal{X}} \lambda_{\max}(W_x), \tag{9.76}$$

where $\widetilde{\lambda}_{\min}(X)$ denotes the minimum non-zero eigenvalue of an operator X. From Eqs. (9.59), (9.62), (9.75), and (9.76), one can verify that

$$f'(s,P) \leq \frac{f(s,P)}{(1+s)^2} \log \frac{1}{w_{\min}}, \tag{9.77}$$

$$f'(s,P) \geq \frac{f(s,P)}{(1+s)^2} \log \frac{1}{w_{\max}}. \tag{9.78}$$

Then for any subsequence that $f(s_{k_n}, P_{k_n}) \not\ll f(s_o, P_o)$, Eq.s (9.70), (9.77) and (9.78) imply

$$\begin{aligned} \mathbb{R}_{\geq 0} \ni \frac{1}{(1+s_o)^2} \log \frac{1}{\omega_{\max}} &\leq \liminf_{n\to+\infty} \widehat{\mathsf{Dlog}}\left[f(s_{k_n}, P_{k_n})\right]\left(f'(s_{k_n}, P_{k_n})\right) \\ &\leq \limsup_{n\to+\infty} \widehat{\mathsf{Dlog}}\left[f(s_{k_n}, P_{k_n})\right]\left(f'(s_{k_n}, P_{k_n})\right) \leq \frac{1}{(1+s_o)^2} \log \frac{1}{\omega_{\min}} \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{9.79}$$

Invoking the continuity of $f(\cdot,\cdot)$, $g(\cdot,\cdot)^0$, combined with Eqs. (9.68) and (9.79), we infer that[1]

$$\lim_{n\to+\infty} \frac{\partial g(s, P_k)}{\partial s}\bigg|_{s=s_k} = g'(s_o, P_o). \tag{9.80}$$

Hence, we complete the claim of the continuity of $g'(\cdot,\cdot)$. By following the same approach, one can also verify the continuity of $g''(\cdot,\cdot)$ and $g'''(\cdot,\cdot)$.

Recall the definition of $E_0(s,P,W)$ in Eq. (9.5) and Eq. (9.61), we have $E_0(s,P,W) = -\log F(s,P)$.

[1] More precisely, $f(s_{k_n}, P_{k_n})$ and $f(s_o, P_o)$ share some disjoint support since $f(s_{k_n}, P_{k_n}) \not\ll f(s_o, P_o)$. However, owing to the finiteness of Eq. (9.79), the projection $g(s_o, P_o)^0$ "nullifies" those disjoint support, and hence we can only consider the joint support of $f(s_{k_n}, P_{k_n})$ and $f(s_o, P_o)$. The continuity of the operation $\widehat{\mathsf{Dlog}}[f(\cdot,\cdot)](\,\cdot\,)$ on the support of $f(s_o, P_o)$ follows from the previous argument.

doi:10.6342/NTU201800597

By denoting $F'(s,P) := \partial F(s,P)/\partial s$, direct calculation shows that

$$\frac{\partial E_0(s,P)}{\partial s} = -\frac{F'(s,P)}{F(s,P)} \tag{9.81}$$

$$\frac{\partial^2 E_0(s,P)}{\partial s^2} = -\frac{F''(s,P)}{F(s,P)} - \left(\frac{\partial E_0(s,P)}{\partial s}\right)^2, \tag{9.82}$$

$$\frac{\partial^3 E_0(s,P)}{\partial s^3} = -\frac{F'''(s,P)}{F(s,P)} + 3\frac{\partial E_0(s,P)}{\partial s}\frac{\partial^2 E_0(s,P)}{\partial s^2} - \left(\frac{\partial E_0(s,P)}{\partial s}\right)^3. \tag{9.83}$$

Now we are at the position to prove Proposition 9.1:

(9.1-(a)) Recalling from Eq. (9.60), the continuity of $E_0(s,P)$, $\partial E_0(s,P)/\partial s$, $\partial^2 E_0(s,P)/\partial s^2$, and $\partial^3 E_0(s,P)/\partial s^3$ follow from the continuity of $g(\cdot,\cdot)$, $g'(\cdot,\cdot)$, $g''(\cdot,\cdot)$, and $g'''(\cdot,\cdot)$.

(9.1-(b)) To prove the concavity of the map $s \mapsto E_0(s,P)$ for $s \geq 0$, we first provide some useful lemmas and the definition of geometric means. Define the "s-weighted geometric mean" of positive definite matrices A and B by

$$A\#_s B := A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^s A^{1/2}. \tag{9.84}$$

It is known that the geometric mean is jointly concave in the matrix partial order (see e.g. [127]):

$$(\theta A + (1-\theta)B)\,\#_s\,(\theta C + (1-\theta)D) \succeq \theta\,(A\#_s C) + (1-\theta)\,(B\#_s D) \tag{9.85}$$

for all $\theta, s \in [0,1]$.

Now we begin the proof of item (b). Since the geometric means, Eq. (9.84), are defined for positive definite matrices, we first present the proof that only works when all $\{W_x\}_{x\in\mathcal{X}}$ are full rank. The proof can then be extended to include the non-invertible case.

Let X be a random variable with the distribution P, and denote by $\mathbb{E}_X$ the expectation with respect to P. Then it suffices to prove the convexity of the map:

$$f : t \mapsto \log \mathrm{Tr}\left[\left(\mathbb{E}_X\, W_X^{\frac{1}{t}}\right)^t\right] \tag{9.86}$$

for all $t \geq 1$.

Let l, r, and θ be arbitrary numbers $1 \leq l \leq r$, $0 \leq \theta \leq 1$, and define

$$t = \theta l + (1-\theta)r. \tag{9.87}$$

Let $t \equiv 1 + s \geq 1$. Then we prove the convexity of the map f from Eq. (9.86), i.e.

$$f(t) \leq \theta f(l) + (1-\theta) f(r). \tag{9.88}$$

Define the number $\tau \in [0,1]$ by

$$\tau = \frac{l\theta}{t}; \quad 1-\tau = \frac{r(1-\theta)}{t}. \tag{9.89}$$

doi:10.6342/NTU201800597

Then it follows that

$$\frac{1}{t} = \frac{\theta}{t} + \frac{1-\theta}{t} = \frac{\tau}{l} + \frac{1-\tau}{r}. \tag{9.90}$$

The concavity of the geometric means (see Eq. (9.85)) implies that

$$\begin{aligned} \mathbb{E}_X\left[W_X^{1/t}\right] &= \mathbb{E}_X\left[W_X^{\tau/l} W_X^{(1-\tau)/r}\right] && (9.91)\\ &= \mathbb{E}_X\left[W_X^{1/l} \#_{1-\tau} W_X^{1/r}\right] && (9.92)\\ &\preceq \mathbb{E}_X\left[W_X^{1/l}\right] \#_{1-\tau} \mathbb{E}_X\left[W_X^{1/r}\right]. && (9.93) \end{aligned}$$

Now let $A \equiv \mathbb{E}_X\left[W_X^{1/l}\right]$ and $B \equiv \mathbb{E}_X\left[W_X^{1/r}\right]$. Since $x \mapsto x^t$ for $t \geq 1$ is a monotone function, Lemma 2.4 in Section 2.1 leads to

$$\begin{aligned} \mathrm{Tr}\left[\left(\mathbb{E}_X\left[W_X^{1/t}\right]\right)^t\right] &\leq \mathrm{Tr}\left[(A\#_{1-\tau}B)^t\right] && (9.94)\\ &\leq \mathrm{Tr}\left[A^{t\tau}B^{t(1-\tau)}\right] && (9.95)\\ &= \mathrm{Tr}\left[A^{l\theta}B^{r(1-\theta)}\right], && (9.96) \end{aligned}$$

where Eq. (9.95) follows from Lemma 2.6. Finally, applying the matrix Hölder's inequality, Lemma 2.5, in Section 2.1 on the right-hand side of Eq. (9.96), we have

$$\begin{aligned} \mathrm{Tr}\left[\left(\mathbb{E}_X\left[W_X^{1/t}\right]\right)^t\right] &\leq \left(\mathrm{Tr}\left[A^l\right]\right)^\theta \left(\mathrm{Tr}\left[B^r\right]\right)^{1-\theta}\\ &= \left(\mathrm{Tr}\left(\mathbb{E}_X\left[W_X^{1/l}\right]\right)^l\right)^\theta \left(\mathrm{Tr}\left(\mathbb{E}_X\left[W_X^{1/r}\right]\right)^r\right)^{1-\theta}. \end{aligned}$$

Taking the logarithm of the above inequality leads to $f(t) \leq \theta f(l) + (1-\theta) f(r)$. This completes the proof for the special case of invertible channel outputs.

The above proof assumes that every realization of the density operator W_x, $x \in \mathcal{X}$, is positive definite. Hence, each density operator $W_x^{\tau/l} W_x^{(1-\tau)/r}$ can be expressed as a geometric mean $W_x^{1/l} \#_{1-\tau} W_x^{1/r}$. However, if W_x is not invertible for some $x \in \mathcal{X}$, then consider a sequence of positive definite operators $W_{x,\varepsilon} := W_x + \varepsilon I$ that approximate W_x, i.e., $\lim_{\varepsilon \searrow 0} W_{x,\varepsilon} = W_x$. The geometric mean of $W_x^{1/l}$ and $W_x^{1/r}$ is then defined by

$$\left(W_x^{1/l}\right) \#_s \left(W_x^{1/r}\right) := \lim_{\varepsilon \searrow 0} \left(W_{x,\varepsilon}^{1/l}\right) \#_s \left(W_{x,\varepsilon}^{1/r}\right), \tag{9.97}$$

by the continuity of the geometric means. Note that the concavity of the geometric means, and Lemmas 2.1 and 2.6 in Section 2.1 still hold if we use the definition in Eq. (9.97). We can thus obtain a complete the proof of item (b).

(9.1-(c)) This item was discovered by Ogawa and Nagaoka [63, Eq. (12)]. For the sake of completeness, we provide the proof here. Note that

$$g(0, P) = f(0, P). \tag{9.98}$$

doi:10.6342/NTU201800597

The continuity of $g'(\cdot,\cdot)$ and Eq. (9.68) imply that

$$\begin{aligned}
&g'(s,P)\big|_{s=0} \\
&= g(0,P)^0 \mathsf{D}\exp\left[\widehat{\log} g(0,P)\right]\left((1+0)\mathsf{D}\widehat{\log}\left[f(0,P)\right]\left(f'(0,P)\right)+\widehat{\log} f(0,P)\right) g(0,P)^0 \\
&= \mathsf{D}\exp\left[\widehat{\log} f(0,P)\right]\left(\mathsf{D}\widehat{\log}\left[f(0,P)\right]\left(f'(0,P)\right)+f(0,P)\widehat{\log} f(0,P)\right) && (9.99)\\
&= f'(s,P)+\widehat{\log} f(s,P)\Big|_{s=0} && (9.100)\\
&= -\sum_{x\in\mathcal{X}} W_x \log W_x + W_P \widehat{\log} W_P. && (9.101)
\end{aligned}$$

Therefore,

$$\left.\frac{\partial E_0(s,P,W)}{\partial s}\right|_{s=0} = -\frac{F'(0,P)}{F(0,P)} = -\operatorname{Tr}\left[g'(0,P)\right] = I(P,W). \tag{9.102}$$

(9.1-(d)) The concavity of the map $s\mapsto E(s,P)$ in item (b) ensures that $\partial E(s,P)/\partial s$ is decreasing in s. Along with item (c) concludes Eq. (9.48).

(9.1-(e)) By using Lemma 2.11 in Section 2.1, we have

$$\begin{aligned}
F''(s,P)\big|_{s=0} = \operatorname{Tr}&\left[g'(s,P)\left((1+s)\mathsf{D}\widehat{\log}\left[f(s,P)\right]\left(f'(s,P)\right)+\widehat{\log} f(s,P)\right)\right]\Big|_{s=0} \\
&+\operatorname{Tr}\left[g(s,P)\left(2\mathsf{D}\widehat{\log}\left[f(s,P)\right]\left(f'(s,P)\right)+(1+s)\left\{\mathsf{D}\widehat{\log}\left[f(s,P)\right]\left(f''(s,P)\right)\right.\right.\right. \\
&\left.\left.\left.+\ \mathsf{D}^2\widehat{\log}\left[f(s,P)\right]\left(f'(s,P)\right)\right\}\right)\right]\Big|_{s=0}. && (9.103)
\end{aligned}$$

From Eqs. (9.100), (9.101), the first term in Eq. (9.103) yields

$$\begin{aligned}
&\operatorname{Tr}\left[g'(0,P)\left(\mathsf{D}\widehat{\log}\left[f(0,P)\right]\left(f'(0,P)\right)+\widehat{\log} f(0,P)\right)\right] && (9.104)\\
&= \operatorname{Tr}\left[f'(0,P)\mathsf{D}\widehat{\log}\left[f(0,P)\right]\left(f'(0,P)\right)+2f'(0,P)\widehat{\log} f(0,P)+f(0,P)\widehat{\log}^2 f(0,P)\right]. && (9.105)
\end{aligned}$$

Similarly, from Eqs. (9.98),(9.63) the second term in Eq. (9.103) leads to

$$\begin{aligned}
&\operatorname{Tr}\left[f(0,P)\left(2\mathsf{D}\widehat{\log}\left[f(0,P)\right]\left(f'(0,P)\right)+\left\{\mathsf{D}\widehat{\log}\left[f(0,P)\right]\left(f''(0,P)\right)\right.\right.\right. \\
&\quad\left.\left.\left.+\ \mathsf{D}^2\widehat{\log}\left[f(0,P)\right]\left(f'(0,P)\right)\right\}\right)\right] && (9.106)\\
&= \operatorname{Tr}\left[\sum_{x\in\mathcal{X}} P(x)W_x\log^2 W_x - f'(0,P)\mathsf{D}\widehat{\log}\left[f(0,P)\right]\left(f'(0,P)\right)\right]. && (9.107)
\end{aligned}$$

Equation (9.103) combined with Eqs. (9.105), (9.107) gives

$$F''(0,P) = \operatorname{Tr}\left[\sum_{x\in\mathcal{X}} W_x\left(\log W_x - \log W_P\right)^2\right]. \tag{9.108}$$

Recalling Eq. (9.82) completes the proof.

□

doi:10.6342/NTU201800597

9.2.2 Proof of Proposition 9.2

(9.2-(a)) The continuity can be proved by the standard approach of functional calculus (see e.g. [126, Lemma III.1] and [120, Section 4.2]). Let $\widetilde{F}(s) := \sum_{x\in\mathcal{X}} P(x)\,\mathrm{Tr}\left[W_x^{1-s}\sigma^s\right]$. Direct calculation shows that

$$\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s} = -\frac{\widetilde{F}'(s)}{\widetilde{F}(s)}, \tag{9.109}$$

$$\frac{\partial^2 E_0^{\downarrow}(s,P,\sigma)}{\partial s^2} = -\frac{\widetilde{F}''(s)}{\widetilde{F}(s)} + \left(\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s}\right)^2, \tag{9.110}$$

$$\frac{\partial^3 E_0^{\downarrow}(s,P,\sigma)}{\partial s^3} = -\frac{\widetilde{F}'''(s,P)}{\widetilde{F}(s,P)} + 3\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s}\frac{\partial^2 E_0^{\downarrow}(s,P,\sigma)}{\partial s^2} - \left(\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s}\right)^3, \tag{9.111}$$

and

$$\widetilde{F}'(s) = \sum_{x\in\mathcal{X}} P(x)\,\mathrm{Tr}\left[-W_x^{1-s}(\log W_x)\sigma^s + W_x^{1-s}\sigma^s\log\sigma\right], \tag{9.112}$$

$$\begin{aligned}\widetilde{F}''(s) = \sum_{x\in\mathcal{X}} P(x)\,\mathrm{Tr}\big[&W_x^{1-s}(\log^2 W_x)\sigma^s - W_x^{1-s}(\log W_x)\sigma^s\log\sigma \\ &-W_x^{1-s}(\log W_x)\sigma^s\log\sigma + W_x^{1-s}\sigma^s\log^2\sigma\big],\end{aligned} \tag{9.113}$$

$$\begin{aligned}\widetilde{F}'''(s) = \sum_{x\in\mathcal{X}} P(x)\,\mathrm{Tr}\big[&-W_x^{1-s}(\log^3 W_x)\sigma^s + W_x^{1-s}(\log^2 W_x)\sigma^s\log\sigma \\ &+2W_x^{1-s}(\log^2 W_x)\sigma^s\log\sigma - 2W_x^{1-s}(\log W_x)\sigma^s\log^2\sigma \\ &-W_x^{1-s}(\log W_x)\sigma^s\log^2\sigma + W_x^{1-s}\sigma^s\log^3\sigma\big].\end{aligned} \tag{9.114}$$

Since the matrix power function is continuous (with respect to the strong topology; see e.g. [69, Theorem 1.19]), we conclude the continuity of the partial derivatives Eqs. (9.109)-(9.111) in item (a).

(9.2-(b)) The claim follows from the concavity of the map $s \mapsto sD_{1-s}(\cdot\|\cdot)$ (see e.g. [58, Lemma III.11]).

(9.2-(c)) The results can be derived from evaluating Eqs. (9.109) and (9.112) at $s = 0$. We provide an alternative proof here. One can verify

$$\left.\frac{\partial E_0^{\downarrow}(s,P,\sigma)}{\partial s}\right|_{s=0} = D_{1-s}\left(P\circ\mathcal{W}\|P\otimes\sigma\right) - sD'_{1-s}\left(P\circ\mathcal{W}\|P\otimes\sigma\right)\big|_{s=0} \tag{9.115}$$

$$= D_{1-s}\left(P\circ\mathcal{W}\|P\otimes\sigma\right)\big|_{s=0} \tag{9.116}$$

$$= D(P\circ\mathcal{W}\|P\otimes\sigma). \tag{9.117}$$

(9.2-(d)) The concavity of the map $s \mapsto E_0^{\downarrow}(s,P,\sigma)$ in item (b) ensures that $\partial E_0^{\downarrow}(s,P,\sigma)/\partial s$ is non-increasing in s. Along with Eq. (9.117), we conclude Eq. (9.51).

doi:10.6342/NTU201800597

(9.2-(e)) Following from item (c), one obtain

$$\left.\frac{\partial^2 E_0^{\downarrow}(s,P,\sigma)}{\partial s^2}\right|_{s=0} = -2D'_{1-s}\left(P\circ\mathcal{W}\|P\otimes\sigma\right) + sD''_{1-s}\left(P\circ\mathcal{W}\|P\otimes\sigma\right)\big|_{s=0} \tag{9.118}$$

$$= -2D'_{1-s}\left(P\circ\mathcal{W}\|P\otimes\sigma\right)\big|_{s=0} \tag{9.119}$$

$$= -V(P\circ\mathcal{W}\|P\otimes\sigma), \tag{9.120}$$

where the last equality (9.120) follows from the fact $D'_{1/1+s}(\cdot\|\cdot)|_{s=0} = V(\cdot\|\cdot)/2$ [120, Theorem 2].

□

9.2.3 Proof of Proposition 9.3

(9.3-(a)) Direct calculation yields that

$$\frac{\partial E_{\mathrm{h}}(s,P,\sigma)}{\partial s} = D_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) - \frac{s}{(1+s)^2}D'_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) \tag{9.121}$$

$$\frac{\partial^2 E_{\mathrm{h}}(s,P,\sigma)}{\partial s^2} = -\frac{2}{(1+s)^3}D'_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) + \frac{s}{(1+s)^4}D''_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) \tag{9.122}$$

$$\frac{\partial^3 E_{\mathrm{h}}(s,P,\sigma)}{\partial s^3} = \frac{6}{(1+s)^4}D'_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) + \frac{3-3s}{(1+s)^5}D''_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) - \frac{s}{(1+s)^6}D'''_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right). \tag{9.123}$$

From Eqs. (9.121)-(9.123) and the fact that $D_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$, $D'_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$, $D''_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$, and $D'''_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$ are continuous for $(s,P)\in\mathbb{R}_{\geq 0}\times\mathcal{P}(\mathcal{X})$, we deduce the continuity property in item (a).

(9.3-(b)) The proof strategy follows closely with [58, Appendix B]. Let $\psi(\alpha) = \sum_{x\in\mathcal{X}} P(x)\log\mathrm{Tr}\left[W_x^{\alpha}\sigma^{1-\alpha}\right]$. Since $\alpha\mapsto\psi(\alpha)$ is convex for all $\alpha\in(0,1]$ [58, Lemma III.11], it can be written as the supremum of affine functions, i.e.

$$\psi(\alpha) = \sup_{i\in\mathcal{I}}\left\{c_i\alpha + d_i\right\} \tag{9.124}$$

for some index set $\mathcal{I}$. Hence,

$$-E_{\mathrm{h}}(s,P,\sigma) = (1+s)\psi\left(\frac{1}{1+s}\right) = \sup_{i\in\mathcal{I}}\left\{c_i + d_i(1+s)\right\}. \tag{9.125}$$

The right-hand side of Eq. (9.125), in turn, implies that the map $s\mapsto E_{\mathrm{h}}(s,P,\sigma)$ is concave for all $s\in\mathbb{R}_{\geq 0}$.

(9.3-(c)) From Eqs. (9.121), one finds

$$\left.\frac{\partial E_{\mathrm{h}}(s,P,\sigma)}{\partial s}\right|_{s=0} = D\left(\mathcal{W}\|\sigma|P\right). \tag{9.126}$$

doi:10.6342/NTU201800597

(9.3-(d)) The concavity of the map $s \mapsto E_{\mathrm{h}}(s,P,\sigma)$ in item (b) ensures that $\partial E_{\mathrm{h}}(s,P,\sigma)/\partial s$ is non-increasing in s. Along with Eq. (9.126) in item (c), we conclude Eq. (9.54).

(9.3-(e)) Applying $D'_{1/1+s}(\cdot\|\cdot)|_{s=0} = V(\cdot\|\cdot)/2$ [120, Theorem 2], it holds that

$$\left.\frac{\partial^2 E_{\mathrm{h}}(s,P,\sigma)}{\partial s^2}\right|_{s=0} = -V\left(\mathcal{W}\|\sigma|P\right). \tag{9.127}$$

□

9.2.4 Proof of Proposition 9.4

This proof follows similarly from Proposition 9.3.

(9.4-(a)) Direct calculation yields that

$$\frac{\partial E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s} = D^{\flat}_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) - \frac{s}{(1+s)^2} D^{\flat\prime}_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) \tag{9.128}$$

$$\frac{\partial^2 E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s^2} = -\frac{2}{(1+s)^3} D^{\flat\prime}_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) + \frac{s}{(1+s)^4} D^{\flat\prime\prime}_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) \tag{9.129}$$

$$\begin{aligned}\frac{\partial^3 E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s^3} = {} & \frac{6}{(1+s)^4} D^{\flat\prime}_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) + \frac{3-3s}{(1+s)^5} D^{\flat\prime\prime}_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right) \\ & - \frac{s}{(1+s)^6} D^{\flat\prime\prime\prime}_{\frac{1}{1+s}}\left(\mathcal{W}\|\sigma|P\right).\end{aligned} \tag{9.130}$$

From Eqs. (9.128)-(9.130) and the fact that $D^{\flat}_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$, $D^{\flat\prime}_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$, $D^{\flat\prime\prime}_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$, and $D^{\flat\prime\prime\prime}_{1/(1+s)}\left(\mathcal{W}\|\sigma|P\right)$ are continuous for $(s,P) \in \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{X})$, we deduce the continuity property in item (a).

(9.4-(b)) The proof strategy follows closely with [58, Appendix B]. Let

$$\tilde{\psi}(\alpha) = \sum_{x\in\mathcal{X}} P(x) \log \mathrm{Tr}\left[\mathrm{e}^{\alpha \log W_x + (1-\alpha)\log\sigma}\right]. \tag{9.131}$$

Since $\alpha \mapsto \tilde{\psi}(\alpha)$ is convex for all $\alpha \in (0,1]$ [58, Lemma III.11], it can be written as the supremum of affine functions, i.e.

$$\tilde{\psi}(\alpha) = \sup_{i\in\mathcal{I}} \{c_i \alpha + d_i\} \tag{9.132}$$

for some index set $\mathcal{I}$. Hence,

$$-E_{\mathrm{h}}^{\flat}(s,P,\sigma) = (1+s)\tilde{\psi}\left(\frac{1}{1+s}\right) = \sup_{i\in\mathcal{I}} \{c_i + d_i(1+s)\}. \tag{9.133}$$

The right-hand side of Eq. (9.133), in turn, implies that the map $s \mapsto E_{\mathrm{h}}^{\flat}(s,P,\sigma)$ is concave for all $s \in \mathbb{R}_{\geq 0}$.

doi:10.6342/NTU201800597

(9.4-(c)) From Eqs. (9.128), one finds

$$\left.\frac{\partial E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s}\right|_{s=0} = D\left(\mathcal{W}\|\sigma|P\right). \tag{9.134}$$

(9.4-(d)) The concavity of the map $s \mapsto E_{\mathrm{h}}^{\flat}(s,P,\sigma)$ in item (b) ensures that $\partial E_{\mathrm{h}}^{\flat}(s,P,\sigma)/\partial s$ is non-increasing in s. Along with Eq. (9.134) in item (c), we conclude Eq. (9.57).

(9.4-(e)) Following similar steps in [120, Proposition 4], it can be verifies that

$$D_{\alpha}^{\flat\prime}(\rho\|\sigma)\Big|_{\alpha=1} = \lim_{\alpha\uparrow 1}\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\log f(\alpha) = \frac{f(1)f''(1)-(f'(1))^2}{2(f(1))^2}, \tag{9.135}$$

where $f(\alpha) := \mathrm{Tr}\left[\mathrm{e}^{\alpha\log\rho+(1-\alpha)\sigma}\right]$. Further, the Fréchet derivative of the exponential (see e.g. [67, Example X.4.2]) gives

$$f'(\alpha) = \mathrm{Tr}\left[\mathrm{e}^{\alpha\log\rho+(1-\alpha)\log\sigma}\left(\log\rho-\log\sigma\right)\right], \tag{9.136}$$

$$f''(\alpha) = \int_0^1 \mathrm{d}t\,\mathrm{Tr}\left[\mathrm{e}^{t(\alpha\log\rho+(1-\alpha)\log\sigma)}\left(\log\rho-\log\sigma\right)\mathrm{e}^{(1-t)(\alpha\log\rho+(1-\alpha)\log\sigma)}\left(\log\rho-\log\sigma\right)\right], \tag{9.137}$$

Therefore, Eq. (9.135) equals

$$D_{\alpha}^{\flat\prime}(\rho\|\sigma)\Big|_{\alpha=1} = \frac{1}{2}\left(\int_0^1 \mathrm{d}t\,\mathrm{Tr}\left[\rho^{1-t}(\log\rho-\log\sigma)\rho^{t}(\log\rho-\log\sigma)\right] - D(\rho\|\sigma)^2\right) \tag{9.138}$$

$$= \frac{1}{2}\widetilde{V}(\rho\|\sigma). \tag{9.139}$$

Finally, combining with Eq. (9.129) yields

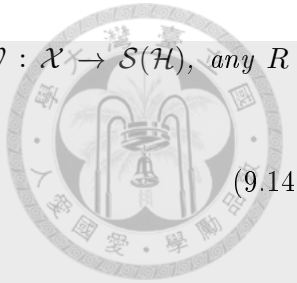
$$\left.\frac{\partial^2 E_{\mathrm{h}}^{\flat}(s,P,\sigma)}{\partial s^2}\right|_{s=0} = -\widetilde{V}\left(\mathcal{W}\|\sigma|P\right). \tag{9.140}$$

□

9.3 Properties of Error Exponent Functions and Saddle-Point

As we will show in Chapter 11, the quantity $E_{\mathrm{sp}}^{(2)}(R,P)$ plays a important role in the connection between hypothesis testing and channel coding. Moreover, in the last Section 9.1, we observe that the error-exponent functions can be represented as a sup-min formulation. In the following Proposition 9.5 we show that the pair of the optimizers in the error-exponent functions form a saddle-point.

doi:10.6342/NTU201800597



Proposition 9.5 (Saddle-Point). *Consider a classical-quantum channel $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$, any $R \in (R_\infty, C_\mathcal{W})$, and $P \in \mathcal{P}(\mathcal{X})$. Let*

$$\mathcal{S}_{P,\mathcal{W}}(\mathcal{H}) := \{\sigma \in \mathcal{S}(\mathcal{H}) : \forall x \in \mathtt{supp}(P),\, W_x \not\perp \sigma\}. \tag{9.141}$$

Define

$$F_{R,P}(\alpha,\sigma) := \begin{cases} \dfrac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{W}\|\sigma|P\right) - R\right), & \alpha \in (0,1) \\ 0, & \alpha = 1 \end{cases}, \tag{9.142}$$

on $(0,1] \times \mathcal{S}(\mathcal{H})$, and denote by

$$\mathcal{P}_R(\mathcal{X}) := \left\{ P \in \mathcal{P}(X) : \sup_{0<\alpha\le 1} \inf_{\sigma\in\mathcal{S}(\mathcal{H})} F_{R,P}(\alpha,\sigma) \in \mathbb{R}_{>0} \right\}. \tag{9.143}$$

The following holds

(a) For any $P \in \mathcal{P}(\mathcal{X})$, $F_{R,P}(\cdot,\cdot)$ has a saddle-point on $(0,1] \times \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ with the saddle-value:

$$\min_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{0<\alpha\le 1} F_{R,P}(\alpha,\sigma) = \sup_{0<\alpha\le 1} \min_{\sigma\in\mathcal{S}(\mathcal{H})} F_{R,P}(\alpha,\sigma) = E_{\mathrm{sp}}^{(2)}(R,P). \tag{9.144}$$

(b) If $P \in \mathcal{P}_R(\mathcal{X})$, the saddle-point is unique.

(c) Fix $P \in \mathcal{P}_R(\mathcal{X})$. Any saddle-point $(\alpha^\star_{R,P}, \sigma^\star_{R,P})$ of $F_{R,P}(\cdot,\cdot)$ satisfies $\alpha^\star_{R,P} \in (0,1)$ and

$$\sigma^\star_{R,P} \gg W_x, \quad \forall x \in \mathtt{supp}(P). \tag{9.145}$$

The proof is provided in Section 9.3.1.

The following Proposition 9.6 discusses the continuity and differentiability of the error-exponent functions. The proof is shown in Section 9.3.2.

doi:10.6342/NTU201800597

Proposition 9.6 (Properties of Error-Exponent Functions). *Consider a classical-quantum channel $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$ with $R_\infty < C_\mathcal{W}$. We have*

(a) Given every $P \in \mathcal{P}(\mathcal{X})$, $E_{\mathrm{sp}}^{(2)}(\cdot, P)$ is convex and non-increasing on $[0, +\infty]$, and continuous on $\left[I_0^{(2)}(P, \mathcal{W}), +\infty\right]$. For every $R > R_\infty$, $E_{\mathrm{sp}}^{(2)}(R, \cdot)$ is continuous on $\mathcal{P}(\mathcal{X})$. Further,

$$E_{\mathrm{sp}}^{(2)}(R, P) = \begin{cases} +\infty, & R < I_0^{(2)}(P, \mathcal{W}) \\ 0, & R \geq I_1^{(2)}(P, \mathcal{W}) \end{cases}. \tag{9.146}$$

(b) $E_{\mathrm{sp}}(\cdot)$ is convex and non-increasing on $[0, +\infty]$, and continuous on $[R_\infty, +\infty]$. Further,

$$E_{\mathrm{sp}}(R) = \begin{cases} +\infty, & R < R_\infty \\ 0, & R \geq C_\mathcal{W} \end{cases}. \tag{9.147}$$

(c) Consider any $R \in (R_\infty, C_\mathcal{W})$ and $P \in \mathcal{P}_R(\mathcal{X})$ (see Eq. (9.143)). The function $E_{\mathrm{sp}}^{(2)}(\cdot, P)$ is differentiable with

$$s_{R,P}^\star = -\left.\frac{\partial E_{\mathrm{sp}}^{(2)}(r, P)}{\partial r}\right|_{r=R} \in \mathbb{R}_{>0}, \tag{9.148}$$

where $s_{R,P}^\star := (1 - \alpha_{R,P}^\star)/\alpha_{R,P}^\star$, and $\alpha_{R,P}^\star$ is the optimizer in Eq. (9.13).

(d) $s_{R,(\cdot)}^\star$ in Eq. (9.148) is continuous on $\mathcal{P}_R(\mathcal{X})$.

Given any $R \in (R_\infty, C_\mathcal{W})$ and $P \in \mathcal{P}_R(\mathcal{X})$, we denote a *maximum absolute value subgradient* of the sphere-packing exponent at R by

$$\left|E_{\mathrm{sp}}'(R)\right| := \max_{P: E_{\mathrm{sp}}^{(2)}(R,P) = E_{\mathrm{sp}}(R)} s_{R,P}^\star. \tag{9.149}$$

Note that the term $\left|E_{\mathrm{sp}}'(R)\right|$ in Eq. (9.149) is well-defined and finite by item (d) in Proposition 9.6.

Figure 9.1 below depicts different cases of the $E_{\mathrm{sp}}(R)$ over rate R.

9.3.1 Proof of Proposition 9.5

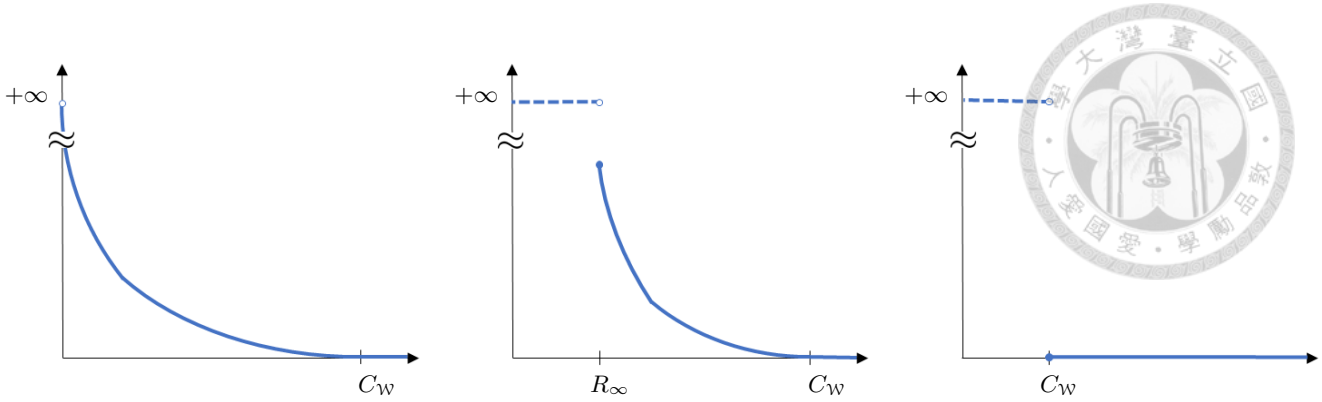
(9.5-(a)) Fix arbitrary $R > R_\infty$ and $P \in \mathcal{P}(\mathcal{X})$. In the following, we prove the existence of a saddle-point of $F_{R,P}(\cdot, \cdot)$ on $(0, 1] \times \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$. Ref. [122, Lemma 36.2] states that $(\alpha^\star, \sigma^\star)$ is a saddle point of $F_{R,P}(\cdot, \cdot)$ if and only if the supremum in

$$\sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(\alpha, \sigma) \tag{9.150}$$

is attained at $\alpha^\star \in (0, 1]$, the infimum in

$$\inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} \sup_{\alpha \in (0,1]} F_{R,P}(\alpha, \sigma) \tag{9.151}$$

doi:10.6342/NTU201800597



(a) $0 = R_\infty < C_{\mathcal{W}}$. (b) $0 < R_\infty < C_{\mathcal{W}}$. (c) $0 < R_\infty = C_{\mathcal{W}}$.

Figure 9.1: This figure illustrates three cases of the strong sphere-packing exponent $E_{\mathrm{sp}}(R)$ over $R \geq 0$. In the first case $0 = R_\infty < C_{\mathcal{W}}$ (the left figure), $E_{\mathrm{sp}}(R)$ is only infinite at $R = 0$ and finite otherwise. In the second case $0 < R_\infty < C_{\mathcal{W}}$ (the central figure), $E_{\mathrm{sp}}(R) = +\infty$ for $R < R_\infty$, and $E_{\mathrm{sp}}(R) < +\infty$ for $R \geq R_\infty$. In the third case $0 < R_\infty = C_{\mathcal{W}}$ (the right figure), $E_{\mathrm{sp}}(R) = +\infty$ for $R < C_{\mathcal{W}}$, and $E_{\mathrm{sp}}(R) = 0$ for $R \geq C_{\mathcal{W}}$. Without loss of generality, we assume $R_\infty < C_{\mathcal{W}}$ to exclude the last case throughout this paper.

is attained at $\sigma^\star \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$, and the two extrema in Eqs. (9.150), (9.151) are equal and finite. We first claim that, $\forall \alpha \in (0, 1]$,

$$\inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma). \tag{9.152}$$

To see this, observe that for any $\alpha \in (0, 1)$, Eqs. (3.5) and (3.38) yield

$$\forall \sigma \in \mathcal{S}(\mathcal{H}) \backslash \mathcal{S}_{P,\mathcal{W}}(\mathcal{H}), \quad D_\alpha \left(\mathcal{W} \| \sigma | P\right) = +\infty, \tag{9.153}$$

which, in turn, implies

$$\forall \sigma \in \mathcal{S}(\mathcal{H}) \backslash \mathcal{S}_{P,\mathcal{W}}(\mathcal{H}), \quad F_{R,P}(\alpha, \sigma) = +\infty. \tag{9.154}$$

Further, Eq. (9.152) holds trivially when $\alpha = 1$. Hence, Eq. (9.152) yields

$$\sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) \tag{9.155}$$

Owing to the fact $R > R_\infty$ and Eq. (9.13), we have

$$E_{\mathrm{sp}}^{(2)}(R, P) = \sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) < +\infty, \tag{9.156}$$

which guarantees the supremum in the right-hand side of Eq. (9.156) is attained at some $\alpha \in (0, 1]$. Namely, there exists some $\bar{\alpha}_{R,P} \in (0, 1]$ such that

$$\sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \max_{\alpha \in [\bar{\alpha}_{R,P}, 1]} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) < +\infty. \tag{9.157}$$

doi:10.6342/NTU201800597

Thus, we complete our claim in Eq. (9.150). It remains to show that the infimum in Eq.(9.151) is attained at some $\sigma^\star \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ and the supremum and infimum are exchangeable. To achieve this, we will show that $\left([\bar{\alpha}_{R,P}, 1], \mathcal{S}_{P,\mathcal{W}}(\mathcal{H}), F_{R,P}\right)$ is a closed saddle-element (see Definition 9.1 below) and employ the boundness of $[\bar{\alpha}_{R,P}, 1] \times \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ to conclude our claim.

Definition 9.1 (Closed Saddle-Element [122]). We denote by $\mathtt{ri}$ and $\mathtt{cl}$ the relative interior and the closure of a set, respectively. Let $\mathcal{A}, \mathcal{B}$ be subsets of a real vector space, and $F : \mathcal{A} \times \mathcal{B} \to \mathbb{R} \cup \{\pm\infty\}$. The triple $(\mathcal{A}, \mathcal{B}, F)$ is called a closed saddle-element if for any $x \in \mathtt{ri}\,(\mathcal{A})$ (resp. $y \in \mathtt{ri}\,(\mathcal{B})$),

(i) $\mathcal{B}$ (resp. $\mathcal{A}$) is convex.

(ii) $F(x, \cdot)$ (resp. $F(\cdot, y)$) is convex (resp. concave) and lower (resp. upper) semi-continuous.

(iii) Any accumulation point of $\mathcal{B}$ (resp. $\mathcal{A}$) that does not belong to $\mathcal{B}$ (resp. $\mathcal{A}$), say y_o (resp. x_o) satisfies $\lim_{y \to y_o} F(x, y) = +\infty$ (resp. $\lim_{x \to x_o} F(x, y) = -\infty$).

Fix an arbitrary $\alpha \in \mathtt{ri}\,([\bar{\alpha}_{R,P}, 1]) = (\bar{\alpha}_{R,P}, 1)$. We check that $\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H}), F_{R,P}(\alpha, \cdot)\right)$ fulfills the three items in Definition 9.1. (i) The set $\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ is clearly convex. (ii) Eq. (3.15) in Lemma 3.2 implies that $\sigma \mapsto D_\alpha(W_x \| \sigma)$ is convex and lower semi-continuous. Since convex combination preservers the convexity and the lower semi-continuity, Eq. (9.142) yields that $\sigma \mapsto F_{R,P}(\alpha, \sigma)$ is convex and lower semi-continuous on $\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$. (iii) Due to the compactness of $\mathcal{S}(\mathcal{H})$, any accumulation point of $\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ that does not belong to $\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$, say σ_o, satisfies $\sigma_o \in \mathcal{S}(\mathcal{H}) \backslash \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$. Eqs. (9.153) and (9.154) then show that $F_{R,P}(\alpha, \sigma_o) = +\infty$.

Next, fix an arbitrary $\sigma \in \mathtt{ri}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right)$. Owing to the convexity of $\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$, it follows that $\mathtt{ri}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right) = \mathtt{ri}\left(\mathtt{cl}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right)\right)$ (see e.g. [123, Theorem 6.3]). We first claim $\mathtt{cl}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right) = \mathcal{S}(\mathcal{H})$. To see this, observe that $\mathcal{S}_{>0}(\mathcal{H}) \subseteq \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ since a full-rank density operator is not orthogonal with every W_x, $x \in \mathcal{X}$. Hence,

$$\mathcal{S}(\mathcal{H}) = \mathtt{cl}\left(\mathcal{S}_{>0}(\mathcal{H})\right) \subseteq \mathtt{cl}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right). \tag{9.158}$$

On the other hand, the fact $\mathcal{S}_{P,\mathcal{W}}(\mathcal{H}) \subseteq \mathcal{S}(\mathcal{H})$ leads to

$$\mathtt{cl}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right) \subseteq \mathtt{cl}\left(\mathcal{S}(\mathcal{H})\right) = \mathcal{S}(\mathcal{H}). \tag{9.159}$$

By Eqs. (9.158) and (9.159), we deduce that

$$\mathtt{ri}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right) = \mathtt{ri}\left(\mathtt{cl}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right)\right) = \mathtt{ri}\left(\mathcal{S}(\mathcal{H})\right) = \mathcal{S}_{>0}(\mathcal{H}), \tag{9.160}$$

where the last equality in Eq. (9.160) follows from [124, Proposition 2.9]. Hence, we obtain

$$\forall \sigma \in \mathtt{ri}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right) \quad \text{and} \quad \forall x \in \mathcal{X}, \quad \sigma \gg W_x. \tag{9.161}$$

Now we verify that $([\bar{\alpha}_{R,P}, 1], F_{R,P}(\cdot, \sigma))$ satisfies the three items in Definition 9.1. Fix an arbitrary $\sigma \in \mathtt{ri}\left(\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})\right)$. (i) The set $(0, 1]$ is obviously convex. (ii) From Eq. (3.13) in Lemma 3.2, the map $\alpha \mapsto F_{R,P}(\alpha, \sigma)$ is continuous on $(0, 1)$. Further, it is not hard to verify that $F_{R,P}(1, \sigma) = 0 = \lim_{\alpha \uparrow 1} F_{R,P}(\alpha, \sigma)$ from Eqs. (9.161), (9.142), and (3.5). Item (c) in

doi:10.6342/NTU201800597

Proposition 3.2 implies that $\alpha \mapsto F_{R,P}(\alpha,\sigma)$ on $[\bar{\alpha}_R, 1)$ is concave. Moreover, the continuity of $\alpha \mapsto F_{R,P}(\alpha,\sigma)$ on $[\bar{\alpha}_{R,P}, 1)$ guarantees the concavity of $\alpha \mapsto F_{R,P}(\alpha,\sigma)$ on $[\bar{\alpha}_{R,P}, 1]$. (iii) Since $[\bar{\alpha}_{R,P}, 1]$ is closed, there is no accumulation point of $[\bar{\alpha}_{R,P}, 1]$ that does not belong to $[\bar{\alpha}_{R,P}, 1]$.

We are at the position to prove item (a) of Proposition 9.5. The closed saddle-element, along with the boundness of $\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ and Rockafellar's saddle-point result [122, Theorem 8], [123, Theorem 37.3] imply that

$$-\infty < \sup_{\alpha\in[\bar{\alpha}_{R,P},1]} \inf_{\sigma\in\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(s,\sigma) = \min_{\sigma\in\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} \sup_{\alpha\in[\bar{\alpha}_{R,P},1]} F_{R,P}(s,\sigma). \tag{9.162}$$

Then Eqs. (9.157) and (9.162) lead to the existence of a saddle-point of $F_{R,P}(\cdot,\cdot)$ on $(0,1]\times\mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$. Hence, item (a) is proved.

(9.5-(b)) Fix arbitrary $R \in (R_\infty, C_\mathcal{W})$ and $P \in \mathcal{P}_R(\mathcal{X})$. We have

$$\sup_{0<\alpha\le 1} \min_{\sigma\in\mathcal{S}(\mathcal{H})} F_{R,P}(\alpha,\sigma) \in \mathbb{R}_{>0}. \tag{9.163}$$

First note that $\alpha^\star = 1$ will not be a saddle point of $F_{R,P}(\cdot,\sigma)$ because $F_{R,P}(1,\sigma) = 0, \forall\sigma\in\mathcal{S}(\mathcal{H})$, contradicting Eq. (9.163).

Now, fix $\alpha^\star \in (0,1)$ to be a saddle-point of $F_{R,P}(\cdot,\cdot)$. Eq. (3.15) in Lemma 3.2 implies that the map $\sigma \mapsto D_{\alpha^\star}(\mathcal{W}\|\sigma|P)$ is strictly convex, and thus the minimizer of Eq. (9.163) is unique. Next, let $\sigma^\star \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ be a saddle-point of $F_{R,P}(\cdot,\cdot)$. Then,

$$F_{R,P}(\alpha,\sigma^\star) = \frac{1-\alpha}{\alpha}\left(I_\alpha^{(2)}(P,\mathcal{W}) - R\right). \tag{9.164}$$

Item (c) in Proposition 3.2 then shows that $\frac{1-\alpha}{\alpha} I_\alpha^{(2)}(P,\mathcal{W})$ is strictly concave on $(0,1)$, which in turn implies that $F_{R,P}(\cdot,\sigma^\star)$ is also strictly concave on $(0,1)$. Hence, the maximizer of Eq. (9.163) is unique.

(9.5-(c)) As shown in the proof of item (b), $\alpha^\star = 1$ is not a saddle point of $F_{R,P}(\cdot,\cdot)$ for any $R > R_\infty$ and $P \in \mathcal{P}_R(\mathcal{X})$. We assume $(\alpha^\star,\sigma^\star)$ is a saddle-point of $F_{R,P}(\cdot,\cdot)$ with $\alpha^\star \in (0,1)$, it holds that

$$F_{R,P}(\alpha^\star,\sigma^\star) = \min_{\sigma\in\mathcal{S}(\mathcal{H})} F_{R,P}(\alpha^\star,\sigma) = \frac{\alpha^\star-1}{\alpha^\star} R + \frac{1-\alpha^\star}{\alpha^\star} \min_{\sigma\in\mathcal{S}(\mathcal{H})} D_{\alpha^\star}(\mathcal{W}\|\sigma|P). \tag{9.165}$$

Then, it is clear from Proposition 3.2-(c) in Section 3.3 that

$$\sigma^\star \gg W_x, \quad \forall x \in \mathtt{supp}(P), \tag{9.166}$$

and thus item (c) is proved.

□

9.3.2 Proof of Proposition 9.6

(9.6-(a)) Fix any arbitrary $P \in \mathcal{P}(\mathcal{X})$. Item (b) in Proposition 3.2 shows that the map $\alpha \mapsto I_\alpha^{(2)}(P,\mathcal{W})$ is monotone increasing on $[0,1]$. Hence, from the definition in Eq. (9.13), it is not hard to verify that

doi:10.6342/NTU201800597

$E_{\rm sp}^{(2)}(R,P)=+\infty$ for all $R\in(0,I_0^{(2)}(P,\mathcal{W}))$; finite for all $R>I_0^{(2)}(P,\mathcal{W})$; and $E_{\rm sp}^{(2)}(R,P)=0$, for all $R\geq I_1^{(2)}(P,\mathcal{W})$.

For every $\alpha\in(0,1]$, the function $\frac{1-\alpha}{\alpha}(I_\alpha^{(2)}(P,\mathcal{W})-R)$ in Eq. (9.13) is an non-increasing, convex, and continuous function in $R\in\mathbb{R}_{>0}$. Since $E_{\rm sp}^{(2)}(R,P)$ is the pointwise supremum of the above function, $E_{\rm sp}^{(2)}(R,P)$ is non-increasing, convex, and lower semi-continuous function for all $R\geq 0$. Furthermore, since a convex function is continuous on the interior of the interval if it is finite [121, Corollary 6.3.3], thus $E_{\rm sp}^{(2)}(R,P)$ is continuous for all $R>I_0^{(2)}(P,\mathcal{W})$, and continuous from the right at $R=I_0^{(2)}(P,\mathcal{W})$.

To establish the continuity of $E_{\rm sp}^{(2)}(R,P)$ in $P\in\mathcal{P}(\mathcal{X})$, we first claim that there exists some $\bar{\alpha}_R\in(0,1]$ such that for every $P\in\mathcal{P}(\mathcal{X})$,

$$\sup_{\alpha\in(0,1]}\frac{1-\alpha}{\alpha}\left(I_\alpha^{(2)}(P,\mathcal{W})-R\right)=\sup_{\alpha\in[\bar{\alpha}_R,1]}\frac{1-\alpha}{\alpha}\left(I_\alpha^{(2)}(P,\mathcal{W})-R\right). \tag{9.167}$$

Recall that $R>R_\infty=\max_{P\in\mathcal{P}(\mathcal{X})}I_0^{(2)}(P,\mathcal{W})$. The continuity, item (a) in Proposition 3.2, implies that there is an $\bar{\alpha}_R>0$ such that

$$R\geq I_{\bar{\alpha}_R}^{(2)}(P,\mathcal{W}),\quad\forall P\in\mathcal{P}(\mathcal{X}). \tag{9.168}$$

Then, Eq. (9.168) and the monotone increases of the map $\alpha\mapsto I_\alpha^{(2)}(P,\mathcal{W})$ yield that,

$$\frac{1-\alpha}{\alpha}\left(I_\alpha^{(2)}(P,\mathcal{W})-R\right)<0,\quad\forall P\in\mathcal{P}(\mathcal{X}),\text{ and }\alpha\in(0,\bar{\alpha}_R). \tag{9.169}$$

The non-negativity of $E_{\rm sp}^{(2)}(R,P)\geq 0$ ensures that the maximizer $\alpha^\star$ will not happen in the region $(0,\bar{\alpha}_R)$, and thus Eq. (9.167) is evident. Finally, Berge's maximum theorem [109, Section IV.3], [110, Lemma 3.1] coupled with the compactness of $[\bar{\alpha}_R,1]$ and item (a) in Proposition 3.2 complete our claim:

$$P\mapsto E_{\rm sp}^{(2)}(R,P)=\sup_{\alpha\in[\bar{\alpha}_R,1]}\frac{1-\alpha}{\alpha}\left(I_\alpha^{(2)}(P,\mathcal{W})-R\right)\text{ is continuous on }\mathcal{P}(\mathcal{X}). \tag{9.170}$$

(9.6-(b)) The statement follows since item (a) holds for any $P\in\mathcal{P}(\mathcal{X})$.

(9.6-(c)) For any $R\in(R_\infty,C_{\mathcal{W}})$ and $P\in\mathcal{P}_R(\mathcal{X})$, item (b) in Proposition 9.5 shows that the optimizer $\alpha_{R,P}^\star$ is unique. Moreover, Eq. (9.148) follows from in Lemma 2.14-(d) in Section 2.2.

(9.6-(d)) The proof of this item is similar to [91, Proposition 3.4]. Fix any $P_o\in\mathcal{P}_R(\mathcal{X})$ and consider arbitrary $\{P_k\}_{k\in\mathbb{N}}$ such that $P_k\in\mathcal{P}_R(\mathcal{X})$, $\forall k\in\mathbb{N}$, and $\lim_{n\to+\infty}P_k=P_o$. Following from Eq. (9.148), we have

$$s_{R,P_k}^\star=-\left.\frac{\partial E_{\rm sp}^{(2)}(r,P_k)}{\partial r}\right|_{r=R}. \tag{9.171}$$

doi:10.6342/NTU201800597

Given any $R \in (R_\infty, C_{\mathrm{W}})$, the continuity of $E_{\mathrm{sp}}^{(2)}(R,\cdot)$ (see item (a)) implies that

$$\lim_{k\to+\infty} E_{\mathrm{sp}}^{(2)}(R,P_k) = E_{\mathrm{sp}}^{(2)}(R,P_o). \tag{9.172}$$

Then, continuity of the first-order derivative in [128, Corollary VI.6.2.8], we have

$$\lim_{k\to+\infty} s_{R,P_k}^\star = \lim_{k\to+\infty} -\left.\frac{\partial E_{\mathrm{sp}}^{(2)}(r,P_k)}{\partial r}\right|_{r=R} = -\left.\frac{\partial E_{\mathrm{sp}}^{(2)}(r,P_o)}{\partial r}\right|_{r=R} = s_{R,P_o}^\star, \tag{9.173}$$

which completes the proof.

□

doi:10.6342/NTU201800597

Chapter 10

Achievability (Channel Coding)

In the error exponent regime (i.e. large deviation regime), the achievability for information transmissions means that one has to construct a coding strategy and show the probability of error achieves the desired upper bound given a fixed transmission rate. The finite blocklength achievability bound for classical-quantum channel exponent was first studied by Burnashev, Holevo [34, 35], and Winter [37]. Specifically, Burnashev and Holevo [34] introduced the following random coding exponent $E_{\mathrm{r}}(R)$ and the auxiliary function $E_0(s,P)$ (see also Eqs. (9.1) and (9.5)):

$$E_{\mathrm{r}}(R) = \sup_{0\leq s\leq 1}\sup_{P\in\mathcal{P}(\mathcal{X})}\left\{E_0(s,P)-sR\right\}; \tag{10.1}$$

$$E_0(s,P) = -\log \mathrm{Tr}\left[\left(\sum_{x\in\mathcal{X}} P(x)W_x^{\frac{1}{1+s}}\right)^{1+s}\right]. \tag{10.2}$$

By quantum Sibson's identity given in Lemma 3.3, it is easy to show that the random coding exponent can be expressed the Rényi capacity with Petz's version (see Eqs. (3.63) and (3.5)):

$$E_{\mathrm{r}}(R) = \sup_{\frac{1}{2}\leq\alpha\leq 1}\frac{1-\alpha}{\alpha}\left(C_{\alpha,\mathcal{W}}-R\right). \tag{10.3}$$

Further, they showed that [34, 35] for pure-state c-q channels (i.e. the channel outputs are all rank-one density operators), there exists a random coding strategy and some decoder (POVM) such that the average error probability over the ensemble, denoted by $\mathsf{P}_{\mathrm{e}}(n,R)$, can be upper bounded as

$$\mathsf{P}_{\mathrm{e}}(n,R)\leq 4\exp\{-nE_{\mathrm{r}}(R)\},\quad \forall R<C_{\mathcal{W}},\, n\in\mathbb{N}. \tag{10.4}$$

However, for general c-q channels (i.e. the channel outputs are possibly non-rank-one density operators), the random coding bound by the exponent function in Eq. (10.3) is still open.

The slightly weaker and the best to date achievability bound was later proven by Hayashi [87, 88, 129]:

$$\mathsf{P}_{\mathrm{e}}(n,R)\leq 4\exp\{-nE_{\mathrm{r}}^{\downarrow}(R)\},\quad \forall R<C_{\mathcal{W}},\, n\in\mathbb{N}. \tag{10.5}$$

doi:10.6342/NTU201800597

The above bound holds for all c-q channels. However, it can be shown that

$$E_{\mathrm{r}}^{\downarrow}(R) \leq E_{\mathrm{r}}(R), \quad \forall R < C_{\mathcal{W}}. \tag{10.6}$$

Recently, Dalai [130] proposed a method to prove Eqs. (10.3) and (10.5). For the sake of completeness, we provide the proof below.

Theorem 10.1 (Dalai [130]). *Given any classical-quantum channels $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$, and any random codes with size M and distribution $P \in \mathcal{P}(\mathcal{X})$, we have the one-shot bound:*

$$\mathsf{P}_{\mathrm{e}}(1, \log M) \leq 6(M-1)^s \exp\left\{-E_0^{\downarrow}(s, P)\right\}, \quad \forall s \in [0,1]. \tag{10.7}$$

Let the transmission rate be $R := \frac{1}{n}\log M < C_{\mathcal{W}}$. The n-shot bound is then:

$$\mathsf{P}_{\mathrm{e}}(n, R) \leq 6 \exp\left\{-nE_{\mathrm{r}}^{\downarrow}(R, P)\right\}, \quad \forall n \in \mathbb{N}. \tag{10.8}$$

For pure-state classical-quantum channels,

$$\mathsf{P}_{\mathrm{e}}(1, \log M) \leq 6(M-1)^s \exp\left\{-E_0(s, P)\right\}, \quad \forall s \in [0,1]. \tag{10.9}$$

Proof of Theorem 10.1. Assume the channel output of a random code is $\{W_{x_1}, \cdots, W_{x_M}\}$ where x_i has an i.i.d. $P(x_i)$. Construct a POVM $\{\Pi_{x_i}\}_{i\in[M]}$ by

$$\Pi_{x_i} := \left(\sum_j \pi_{x_j}\right)^{-\frac{1}{2}} \pi_{x_i} \left(\sum_j \pi_{x_j}\right)^{-\frac{1}{2}}, \tag{10.10}$$

where

$$\pi_{x_i} := \left\{ W_{x_i}^{\alpha s} - \left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^s > 0 \right\}, \quad \alpha \in (0,1],\ s \in (0,1]. \tag{10.11}$$

Using Hayashi-Nagaoka inequality, Lemma 2.9, we have

$$\mathbb{1} - \Pi_{x_i} \leq 2(\mathbb{1} - \pi_{x_i}) + 4\sum_{j\neq i} \pi_{x_j}. \tag{10.12}$$

Hence, the average probability of error given realizations $(x_1, \ldots, x_M)$ can be upper bounded as

$$\Pr\{\text{error}|(x_1, \ldots, x_M)\} = \frac{1}{M} \operatorname{Tr}\left[W_{x_i}(\mathbb{1} - \Pi_{x_i})\right] \tag{10.13}$$

$$\begin{aligned} &\leq 2\frac{1}{M}\sum_i \operatorname{Tr}\left[W_{x_i}\left\{W_{x_i}^{\alpha s} - \left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^s \leq 0\right\}\right] \\ &\quad + 2\frac{1}{M}\sum_i \operatorname{Tr}\left[W_{x_i}\left\{W_{x_i}^{\alpha s} - \left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^s > 0\right\}\right]. \end{aligned} \tag{10.14}$$

doi:10.6342/NTU201800597

For $0 < \alpha \leq 1$ and $0 < s \leq 1$, using 2.10 to bound the first term in Eq. (10.14) as

$$\operatorname{Tr}\left[W_{x_i}\left\{W_{x_i}^{\alpha s} - \left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{s} \leq 0\right\}\right] \leq \operatorname{Tr}\left[W_{x_i}^{1-\alpha s}\left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{s}\right]. \tag{10.15}$$

Recalling the operator concavity of $u \mapsto u^s$, we take expectation of the random code to obtain

$$\begin{aligned} 2\mathbb{E}\operatorname{Tr}\left[W_{x_i}^{1-\alpha s}\left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{s}\right] &= 2\operatorname{Tr}\left[\mathbb{E}_x\left[W_x^{1-\alpha s}\right]\mathbb{E}\left[\left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{s}\right]\right] && (10.16)\\ &\leq 2\operatorname{Tr}\left[\mathbb{E}_x\left[W_x^{1-\alpha s}\right]\left(\mathbb{E}\left[\sum_{j\neq i} W_{x_j}^{\alpha}\right]\right)^{s}\right] && (10.17)\\ &= 2(M-1)^s \operatorname{Tr}\left[\mathbb{E}_x\left[W_x^{1-\alpha s}\right]\mathbb{E}_x\left[W_x^{\alpha}\right]^{s}\right]. && (10.18) \end{aligned}$$

For the second term in Eq. (10.14), we re-index it to have

$$4\frac{1}{M}\sum_i \operatorname{Tr}\left[\left(\sum_{j\neq i} W_{x_j}\right)\left\{W_{x_i}^{\alpha s}\left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{s}\right\}\right]. \tag{10.19}$$

Again, using Lemma 2.10 yields

$$\operatorname{Tr}\left[\left(\sum_{j\neq i} W_{x_j}\right)\left\{W_{x_i}^{\alpha s}\left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{s}\right\}\right] \leq \operatorname{Tr}\left[\left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{s} W_{x_i}^{1-\alpha s}\right]. \tag{10.20}$$

Taking expectation and combining with Eq. (10.18), we have

$$\mathsf{P}_{\mathrm{e}}(n,R) \leq 6(M-1)^s \operatorname{Tr}\left[\mathbb{E}_x\left[W_x^{1-\alpha s}\right]\mathbb{E}_x\left[W_x^{\alpha}\right]^{s}\right]. \tag{10.21}$$

Invoking the definition of $E_{\mathrm{r}}^{\downarrow}(R)$ and choosing $\alpha = 1/(1+s)$, we obtain Eq. (10.7).

For pure-state c-q channels, Eq. (10.21) can be rewritten as

$$\mathsf{P}_{\mathrm{e}}(n,R) \leq 6(M-1)^s \operatorname{Tr}\left[\mathbb{E}_x\left[W_x\right]\right]^{1+s}, \tag{10.22}$$

because $W_x^p = W_x$ for $p \geq 0$ for pure-state c-q channels. The above expression equals to Eq. (10.9), which completes the proof. □

Remark 10.1. To obtain the Eq. (10.9) for general c-q channels, one possible way of the above method is to employ the inequality

$$\left(\sum_{j\neq i} W_{x_j}\right)^{\alpha} \leq \sum_{j\neq i} W_{x_j}^{\alpha}, \quad \forall \alpha \in [0,1], \tag{10.23}$$

doi:10.6342/NTU201800597

which in turn implies

$$\sum_{j\neq i} W_{x_j} \leq \left(\sum_{j\neq i} W_{x_j}^{\alpha}\right)^{\frac{1}{\alpha}}. \tag{10.24}$$

Unfortunately, the operator inequality in Eq. (10.23) does not hold for general density operators W_{x_i}. The inequality only holds under the weak majorization. ◇

Lastly, the following Conjecture 10.1 was posed by Holevo [35]. Note that to achieve Eq. (10.26), the right-hand side of Eq. (10.25) allows to have any sub-exponential prefactors $\exp\{o(n)\}$.

Conjecture 10.1 (Random Coding Bound for Classical-Quantum Channels)**.** *Given any classical-quantum channels $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$, transmission rate $R < C_{\mathcal{W}}$, and random codes with distribution $P \in \mathcal{P}(\mathcal{X})$, one has*

$$\mathsf{P}_{\mathrm{e}}(n, R) \leq \exp\{-nE_{\mathrm{r}}(R, P)\}, \quad \forall n \in \mathbb{N}. \tag{10.25}$$

In particular,

$$\varepsilon^*(n, R) \leq \exp\{-nE_{\mathrm{r}}(R)\}, \quad \forall n \in \mathbb{N}. \tag{10.26}$$

doi:10.6342/NTU201800597

Chapter 11

Optimality (Channel Coding)

In this chapter, we present the weak and strong sphere-packing bounds for c-q channels. In Section 11.1, we first review existing approaches of proving classical sphere-packing bound. In Section 11.2, we provide the proof of a weak sphere-packing bound by using Wolfowitz strong converse. This bound is new in the quantum scenario and will be used in the moderate deviation analysis in Section 12. In Section 11.3, we prove our main result of a finite blocklength strong sphere-packing bound for c-q channels, see Theorem 11.1 below, which improve Dalai's prefactor [38, 39] from the order of subexponential $\mathrm{e}^{O(\sqrt{n})}$ to polynomial. Lastly, in Section 11.4, we obtain exact asymptotics (i.e. exact prefactors) of the strong sphere-packing bound for a symmetric c-q channels, which can be seen as a generalization of classical symmetric channels [21].

Theorem 11.1 (Finite Blocklength Strong Sphere-Packing Bound of Constant Composition Codes)**.** *Consider a classical-quantum channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ and $R\in(R_\infty,C_{\mathcal{W}})$. For every $\gamma>0$, there exist an $N_0\in\mathbb{N}$ and a constant $A>0$ such that for all constant composition codes $\mathcal{C}_n$ of length $n\geq N_0$ with message size $|\mathcal{C}_n|\geq\exp\{nR\}$, we have*

$$\bar{\varepsilon}\left(\mathcal{C}_n\right)\geq\frac{A}{n^{\frac{1}{2}\left(1+\left|E'_{\mathrm{sp}}(R)\right|+\gamma\right)}}\exp\left\{-nE_{\mathrm{sp}}(R)\right\}. \tag{11.1}$$

The following corollary generalizes the refined sphere-packing bound for constant composition codes to arbitrary codes by using the standard argument [30, p. 95]. We delay the proof to the end of Section 11.3.5.

Corollary 11.1 (Finite Blocklength Strong Sphere-Packing Bound of General Codes)**.** *Consider a classical-quantum channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ and $R\in(R_\infty,C_{\mathcal{W}})$. There exist some $t>1/2$ and $N_0\in\mathbb{N}$ such that for all codes of length $n\geq N_0$, we have*

$$\varepsilon^*\left(n,R\right)\geq n^{-t}\exp\left\{-nE_{\mathrm{sp}}(R)\right\}. \tag{11.2}$$

Theorem 11.1 yields

$$\log\frac{1}{\bar{\varepsilon}(\mathcal{C}_n)}\leq nE_{\mathrm{sp}}(R)+\frac{1}{2}\left(1+\left|E'_{\mathrm{sp}}(R)\right|\right)\log n+o(\log n), \tag{11.3}$$

where the term $\frac{1}{2}\left(1+\left|E'_{\mathrm{sp}}(R)\right|\right)$ can be viewed as a second-order term (see the discussions in [18,

doi:10.6342/NTU201800597

Section 4.4]). On the other hand, for the case of classical *non-singular* channels[1], it was shown that [131, Theorem 3.6], for all constant composition codes $\mathcal{C}_n$ and rate $R \in (C_{1/2,\mathcal{W}}, C_{\mathcal{W}})$,

$$\log \frac{1}{\bar{\varepsilon}(\mathcal{C}_n)} \geq nE_{\mathrm{r}}(R) + \frac{1}{2}\left(1 + \left|E'_{\mathrm{r}}(R)\right|\right)\log n + \Omega(1), \tag{11.4}$$

where $E_{\mathrm{r}}(R)$ is the random coding exponent defined in Eq. (9.1), and note that $E_{\mathrm{r}}(R) = E_{\mathrm{sp}}(R)$ for all $R \geq C_{1/2,\mathcal{W}}$ [21, p. 160], [36]. Hence our result, Theorem 11.1, matches the achievability up to the logarithmic order. We note that whether the third order $o(\log n)$ in Eq. (11.3) can be improved to $O(1)$ is still unknown even for the classical case.

11.1 Literature Review of Classical Sphere-Packing Bound

This section reviews existing proof approaches of classical sphere-packing bounds:

$$\varepsilon^*(n,R) \geq f(n)\exp\left\{-n\left[E_{\mathrm{sp}}(R-g(n))\right]\right\}, \tag{11.5}$$

$$\varepsilon^*(n,R) \geq f(n)\exp\left\{-n\left[\widetilde{E}_{\mathrm{sp}}(R-g(n))\right]\right\}, \tag{11.6}$$

where $f(n)$ is the pre-factor of the bound, and $g(n)$ is the back-off from the rate. We remark that E_{sp} coincides with $\widetilde{E}_{\mathrm{sp}}$ in the classical case. The reason why we distinguish the notation E_{sp} and $\widetilde{E}_{\mathrm{sp}}$ here is because of their possible quantum generalizations (recalling that they are not equal in the quantum case, i.e. Theorem 9.1 in Section 9.1). Table 11.1 below summarizes the comparisons of existing results.

Bounds\Settings	Finite blocklength	Composition dependent	Pre-factor $f(n)$	Rate back-off $g(n)$	Classical-quantum channels	Tightness
(a) Shannon-Gallager-Berlekamp [30]	No	Yes	$\mathrm{e}^{-O(\sqrt{n})}$	$O\left(\frac{\log n}{n}\right)$	Dalai [38]	Strong
(b) Haroutunian [31] Omura [133] Csisár-Korner [25]	No	Yes	$\mathrm{e}^{-o(n)}$	$o(1)$	Winter [37]	Weak
(c) Blahut [32]	No	No	$\mathrm{e}^{-O(\sqrt{n})}$	$O\left(n^{-\frac{1}{2}}\right)$	Eqs. (11.148) & (11.153)	Strong
(d) Altuğ-Wagner [91]	Yes	Yes	$n^{-\frac{1}{2}\left(1+\left\|E'_{\mathrm{sp}}(R)\right\|+o(1)\right)}$	0	Theorem 11.1	Strong
(e) Elkayam-Feder [134]	Yes	Yes	$O\left(n^{-t}\right)$	$O\left(\frac{\log n}{n}\right)$	Unknown	Unknown
(f) Agustin-Nakiboğlu [135, 106, 105, 136]	Yes	No	$O\left(n^{-t}\right)$	0	Unknown	Unknown

Table 11.1: Different sphere-packing bounds are compared by (i) the bound is finite blocklength or asymptotical; (ii) whether or not they are dependent on the constant composition codes; (iii) & (iv) the asymptotics of $f(n)$ and $g(n)$; (v) the corresponding c-q generalizations. The parameter t in rows (e) and (f) is some value in the range $t > 1/2$; and (vi) whether their error exponent expressions for c-q channels are in the strong form (Eq. (1.4)) or weak form (Eq. (12.51)).

(a) Shannon, Gallager and Berlekamp obtained the first classical sphere-packing bound Eq. (11.5),

[1] For classical *singular* channels, one has $\log \frac{1}{\bar{\varepsilon}(\mathcal{C}_n)} \geq nE_{\mathrm{r}}(R) + \frac{1}{2}\log n + \Omega(1)$ [131]. Further, it was conjectured that [132] that $\log \frac{1}{\bar{\varepsilon}(\mathcal{C}_n)} \leq nE_{\mathrm{sp}}(R) + \frac{1}{2}\log n + o(\log n)$, for all asymmetric classical singular channels and constant composition codes. However, such a result remains open.

doi:10.6342/NTU201800597

where [30, Theorem 5]

$$f(n) = \mathrm{e}^{-O(\sqrt{n})}; \quad g(n) = O\left(\frac{\log n}{n}\right). \tag{11.7}$$

Their method is based on distinguishing two codewords, followed by Chebyshev's inequality. The works [137] and [138] further improved the coefficients in $f(n)$ and $g(n)$ for short to moderate blocklengths.

Remarkably, Shannon-Gallager-Berlekamp's result can be extended to c-q channels with almost the same asymptotics in Eq. (11.7) [38]. See also the result by Dalai and Winter for constant composition codes [39].

(b) Haroutunian [31], Omura [133], Csiszár and Körner [25], Ahlswede [139] subsequently proposed a sphere-packing exponent using discrimination functions (i.e. the relative entropy function in Eq. (1.5)), and obtained the following classical sphere-packing bound for constant composition codes $\mathcal{C}_n$:

$$\bar{\varepsilon}(\mathcal{W}, \mathcal{C}_n) \geq \frac{1}{2} \exp\left\{-n\widetilde{E}_{\mathrm{sp}}(R-\delta)(1+\delta)\right\}, \tag{11.8}$$

for all $\delta > 0$ and all sufficiently large $n \in \mathbb{N}$, and $\bar{\varepsilon}(W, \mathcal{C}_n)$ denotes the average error of the code $\mathcal{C}_n$. The idea is to apply strong converse bounds [140, 141, 142, 133, 25] to a dummy channel, and then use a data-processing inequality for the discrimination function between the dummy and true channels. Recently, Altuğ and Wagner employed a particular strong converse result, Wolfowitz's strong converse result [143], and obtained a form of Eq. (11.6) with [43, Lemma 3]:

$$g(n) = O\left(\frac{1}{\sqrt{n}}\right). \tag{11.9}$$

Following the arguments in [139, Theorem 49], Winter proved a weak sphere-packing bound Eq. (11.8) for constant composition codes in c-q channels [37, Theorem II.20]. We remark that Altuğ and Wagner's result [43] can also be extended to a weak sphere-packing bound for c-q channels when combining Winter's approach [37] with Sharma and Warsi's strong converse result [125, Theorem 3].

(c) Blahut related the channel coding problem to hypothesis testing [32, Theorem 20] (see also [23, Theorem 10.2.1]) and independently obtained a weak sphere-packing bound Eq. (11.6) with

$$f(n) = \mathrm{e}^{-O(\sqrt{n})}; \quad g(n) = O\left(\frac{1}{\sqrt{n}}\right). \tag{11.10}$$

In Section 11.3, we generalize Blahut's result to a strong sphere-packing bound for c-q channels.

(d) In Ref. [48], Altuğ and Wagner applied a sharp concentration inequality to refine the sphere-packing bound Eq. (11.7) with

$$f(n) = \mathrm{e}^{-O(\sqrt{n})}; \quad g(n) = O\left(\frac{\log n}{n}\right), \tag{11.11}$$

doi:10.6342/NTU201800597

for some $t > 1/2$ and all sufficiently large $n \in \mathbb{N}$.

(e) Elkayam and Feder [134] established a general expression for the error probability in terms of the cumulative distribution function [144, Theorem 6]. Combined with the method of types and Polyanskiy's minimax meta-converse [145, Theorem 3], they proved a classical sphere-packing bound for constant composition codes with

$$f(n) = \Theta\left(n^{-t}\right); \quad g(n) = O\left(\frac{\log n}{n}\right), \tag{11.12}$$

for some $t > 1/2$. This sphere-packing bound also had a polynomial pre-factor; however, it is unknown whether this method can be extended to c-q channels.

11.2 A Weak Sphere-Packing Bound via Wolfowitz Strong Converse

Theorem 11.2 (Weak Converse Bound with Polynomial Prefactors). *Consider a classical-quantum channel $\mathcal{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ with $\mathcal{S}_\circ := \overline{\mathsf{im}(\mathcal{W})}$, an arbitrary rate $R \geq 0$, and $\sigma \in \mathcal{S}_{>0}(\mathcal{H})$. For any $\eta \in (0, \frac{1}{2})$ and $c > 0$, let $N_0 \in \mathbb{N}$ such that for all $n \geq N_0$,*

$$c \cdot \mathrm{e}^{-\xi\sqrt{n}} \leq \frac{\eta}{2}, \tag{11.13}$$

where $\xi = \sqrt{2A/\eta}$ and $A := \max_{\rho \in \mathcal{S}_\circ} V(\rho\|\sigma)$. Then, it holds that for all $n \geq N_0$,

$$\widehat{\alpha}_{\exp\{-nR\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| \sigma^{\otimes n}\right) \geq f(\eta) \exp\left\{-n\left[\frac{\widetilde{E}_{\mathrm{sp}}\left(R - \frac{2\xi}{\sqrt{n}}, P_{\mathbf{x}^n}, \sigma\right)}{1-\eta}\right]\right\}, \tag{11.14}$$

where $f(\eta) = \exp\left\{-\frac{h(1-\eta)}{1-\eta}\right\}$ and $h(p) := -p\log p - (1-p)\log(1-p)$ is the binary entropy function.

Remark 11.1. Consider a constant composition code with common type $P_{\mathbf{x}^n}$ on a finite input alphabet $\mathcal{X}$. Recall the definition of the weak sphere-packing exponent [37, 26]:

$$\widetilde{E}_{\mathrm{sp}}(R, P_{\mathbf{x}^n}) := \min_{\bar{\mathcal{W}}: \mathcal{X} \to \mathcal{S}(\mathcal{H})} \left\{D\left(\bar{\mathcal{W}} \| \mathcal{W} | P_{\mathbf{x}^n}\right) : I(P_{\mathbf{x}^n}, \bar{\mathcal{W}}) \leq R\right\}. \tag{11.15}$$

Theorem 11.2, along with the one-shot converse bound (see Proposition 11.2 in Section 11.3.1 later), establishes a weak sphere-packing bound with polynomial prefactors, which generalizes Altuğ and Wagner's result [43, Lemma 3] to c-q channels: for any $\eta \in (0, \frac{1}{2})$ and for all sufficiently large n such that Eq. (11.13) holds, we have

$$\begin{aligned}
\varepsilon_{\max}(\mathcal{W}, P_{\mathbf{x}^n}) &\geq \max_{\sigma \in \mathcal{S}(\mathcal{H})} \widehat{\alpha}_{\exp\{-nR\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| \sigma^{\otimes n}\right) && (11.16)\\
&\geq \widehat{\alpha}_{\exp\{-nR\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^\star)^{\otimes n}\right) && (11.17)\\
&\geq f(\eta) \exp\left\{-n\left[\frac{\widetilde{E}_{\mathrm{sp}}\left(R - \frac{2\xi}{\sqrt{n}}, P_{\mathbf{x}^n}\right)}{1-\eta}\right]\right\}, && (11.18)
\end{aligned}$$

where $\sigma^\star := P_{\mathbf{x}^n}\bar{\mathcal{W}}^\star$ and $\bar{\mathcal{W}}^\star$ is an arbitrary minimizer in Eq. (11.15). Moreover, Eq. (11.18) im-

doi:10.6342/NTU201800597

proves the prefactor of Winter's weak sphere-packing bound [37] from the order of subexponential to polynomial. ◇

Proof of Theorem 11.2. Consider an arbitrary sequence $\mathbf{x}^n \in \mathcal{X}^n$ and a test Q_n on $\mathcal{H}^{\otimes n}$. For two c-q channels $\bar{\mathcal{W}}, \mathcal{W} : \mathcal{X} \to \mathcal{S}_\circ$, the data-processing inequality implies that

$$D\left(\bar{W}_{\mathbf{x}^n}^{\otimes n} \| W_{\mathbf{x}^n}^{\otimes n}\right) \geq \left[1 - \alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})\right] \log \frac{1 - \alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})}{1 - \alpha(Q_n; W_{\mathbf{x}^n}^{\otimes n})} + \alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n}) \log \frac{\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})}{\alpha(Q_n; W_{\mathbf{x}^n}^{\otimes n})} \tag{11.19}$$

$$= -h\left(\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})\right) - \alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n}) \log \alpha(Q_n; W_{\mathbf{x}^n}^{\otimes n}) - \left[1 - \alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})\right] \log\left(1 - \alpha(Q_n; W_{\mathbf{x}^n}^{\otimes n})\right) \tag{11.20}$$

$$\geq -\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n}) \log \alpha(Q_n; W_{\mathbf{x}^n}^{\otimes n}) - h\left(\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})\right), \tag{11.21}$$

where the last inequality (11.21) follows since the third term in (11.20) is non-negative. Continuing from Eq. (11.21), we have

$$\alpha(Q_n; W_{\mathbf{x}^n}^{\otimes n}) \geq \exp\left\{-\frac{D\left(\bar{W}_{\mathbf{x}^n}^{\otimes n} \middle\| W_{\mathbf{x}^n}^{\otimes n}\right) + h\left(\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})\right)}{\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})}\right\} \tag{11.22}$$

$$= \exp\left\{-\frac{n\, D\left(\bar{\mathcal{W}} \middle\| \mathcal{W} \middle| P_{\mathbf{x}^n}\right) + h\left(\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})\right)}{\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})}\right\}, \tag{11.23}$$

where Eq. (11.23) follows from the additivity of the relative entropy and the empirical distribution $P_{\mathbf{x}^n}$.

The next step is to replace $\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n})$ with a lower bound that does not depend on the dummy channel $\bar{W}$, provided that $\bar{W}$ satisfies certain conditions. This can be done using Proposition 11.1, Wolfowitz's strong converse bound. We delay its proof in Section 11.2.1 below.

Proposition 11.1 (Wolfowitz's Strong Converse)**.** *Let $\mathcal{S}_\circ \subseteq \mathcal{S}(\mathcal{H})$ be closed and let $\bar{\mathcal{W}} : \mathcal{X} \to \mathcal{S}_\circ$ be an arbitrary classical-quantum channel. Consider the binary hypothesis testing:*

$$\mathsf{H}_0 : \bar{W}_{\mathbf{x}^n}^{\otimes n}, \tag{11.24}$$

$$\mathsf{H}_1 : \sigma^{\otimes n}, \tag{11.25}$$

where $\mathbf{x}^n \in \mathcal{X}^n$ and $\sigma \in \mathcal{S}_{>0}(\mathcal{H})$. For any test Q_n such that $\beta(Q_n; \sigma^{\otimes n}) \leq \mathrm{e}^{-nR}$ and $D\left(\bar{W}_{\mathbf{x}^n} \| \sigma | P_{\mathbf{x}^n}\right) \leq R - 2\kappa$, it holds that

$$\alpha\left(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n}\right) > 1 - \frac{A}{n\kappa^2} - \mathrm{e}^{-n\kappa}, \tag{11.26}$$

where $A := \max_{\rho \in \mathcal{S}_\circ} V(\rho \| \sigma)$.

Fix $0 < \eta < \frac{1}{2}$, and let $\xi^2 := \frac{2A}{\eta}$. Note that ξ^2 is finite because $A < +\infty$. For all $n \geq N_0$, we have

$$c \cdot \mathrm{e}^{-\xi\sqrt{n}} \leq \frac{\eta}{2} \tag{11.27}$$

by assumption in Theorem 11.2. Choose $\kappa = \xi/\sqrt{n}$. For any $\bar{\mathcal{W}} : \mathcal{X} \to \mathcal{S}_\circ$ with $D\left(\bar{\mathcal{W}} \| \sigma | P_{\mathbf{x}^n}\right) \leq R - \frac{2\xi}{\sqrt{n}}$ and any test Q_n such that $\beta(Q_n; \sigma^{\otimes n}) \leq \mathrm{e}^{-nR}$, Proposition 11.1 gives a lower bound to the type-I

doi:10.6342/NTU201800597

error:

$$\alpha(Q_n; \bar{W}_{\mathbf{x}^n}^{\otimes n}) \geq 1 - \frac{A}{n\kappa^2} - \mathrm{e}^{-n\kappa} \geq 1 - \eta. \tag{11.28}$$

Hence, combining Eqs. (11.23) and (11.28) yields that, for any $\beta(Q_n; \sigma^{\otimes n}) \leq c\mathrm{e}^{-nR}$,

$$\alpha(Q_n; W_{\mathbf{x}^n}^{\otimes n}) \geq \max_{\bar{\mathcal{W}}: D(\bar{\mathcal{W}}\|\sigma|P_{\mathbf{x}^n}) \leq R - \frac{2\xi}{\sqrt{n}}} \exp\left\{ -\frac{n\, D\left(\bar{\mathcal{W}} \middle\| \mathcal{W} \middle| P_{\mathbf{x}^n}\right) + h\left(1-\eta\right)}{1-\eta} \right\}, \tag{11.29}$$

$$= \exp\left\{ -\frac{h\left(1-\eta\right)}{1-\eta} \right\} \exp\left\{ -\frac{n\, \widetilde{E}_{\mathrm{sp}}\left(R - \frac{2\xi}{\sqrt{n}}, P_{\mathbf{x}^n}, \sigma\right)}{1-\eta} \right\}, \tag{11.30}$$

which concludes Theorem 11.2. □

11.2.1 Proof of Wolfowitz's Strong Converse, Proposition 11.1

This proof follows similar steps by Sharma and Warsi [125, Theorem 3], which uses generalized divergences to prove Wolfowitz's strong converse.

To prove our claim, we first introduce notation for generalized divergences. For any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\gamma > 0$, define the *hockey-stick divergence* by

$$\mathcal{D}_\gamma(\rho\|\sigma) := \mathrm{Tr}\left[(\rho - \gamma\sigma)_+\right], \tag{11.31}$$

where $A_+ := A\{A \geq 0\}$ denotes the positive part of A. This divergence satisfies the data-processing inequality (DPI):

$$\mathrm{Tr}\left[(\rho - \gamma\varrho)_+\right] \geq \mathrm{Tr}\left[(\mathcal{N}(\rho) - \gamma\mathcal{N}(\varrho))_+\right], \tag{11.32}$$

for any completely positive and trace-preserving map $\mathcal{N} : \mathcal{S}(\mathcal{H}_{\mathrm{in}}) \to \mathcal{S}(\mathcal{H}_{\mathrm{out}})$ [125, Lemma 4]. Let

$$\rho_p := p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|, \quad \text{and} \quad \sigma_q := q|0\rangle\langle 0| + (1-q)|1\rangle\langle 1|, \tag{11.33}$$

for $0 \leq p, q \leq 1$ and some orthonormal basis $\{|0\rangle, |1\rangle\}$, and define

$$\mathrm{d}_\gamma\left(p\|q\right) := \mathcal{D}_\gamma\left(\rho_p\|\sigma_q\right). \tag{11.34}$$

Note that the quantity $\mathrm{d}_\gamma\left(p\|q\right)$ is independent of the choice of the basis $\{|0\rangle, |1\rangle\}$. Now we are ready to prove Proposition 11.1.

Proof of Proposition 11.1. Fix an arbitrary test Q_n on $\mathcal{H}^{\otimes n}$. For notational convenience, we shorthand $\rho^n = \bar{W}_{\mathbf{x}^n}^{\otimes n}$, $\tau^n = \sigma^{\otimes n}$, $\alpha = \alpha(Q_n; \rho^n)$ and $\beta = (Q_n; \tau^n)$. Further, we assume $\beta(Q_n; \tau^n) \leq \mathrm{e}^{-nR}$. From the definition of the classical divergence, Eqs. (11.31) and (11.34), and any $\gamma > 0$, we find

$$\mathrm{d}_\gamma(1-\alpha\|\beta) = (1 - \alpha - \gamma\beta)_+ + (\alpha - \gamma\left[1-\beta\right])_+ \tag{11.35}$$

$$\geq 1 - \alpha - \gamma\beta \tag{11.36}$$

$$\geq 1 - \alpha - \gamma\mathrm{e}^{-nR}. \tag{11.37}$$

doi:10.6342/NTU201800597

On the other hand, DPI for the measurement map $\mathrm{Tr}[Q_n(\cdot)]|0\rangle\langle 0| + (1 - \mathrm{Tr}[Q_n(\cdot)])|1\rangle\langle 1|$ implies that

$$\mathcal{D}_\gamma\left(\rho^n \| \tau^n\right) \geq \mathrm{d}_\gamma\left(\mathrm{Tr}[Q_n\rho^n] \| \,\mathrm{Tr}[Q_n\tau^n]\right) = \mathrm{d}_\gamma(1-\alpha\|\beta). \tag{11.38}$$

Hence, Eqs. (11.37) and (11.38) lead to

$$\alpha \geq 1 - \mathcal{D}_\gamma\left(\rho^n \| \tau^n\right) - \gamma \mathrm{e}^{-nR}. \tag{11.39}$$

Since

$$\begin{aligned} \mathcal{D}_\gamma\left(\rho^n \| \tau^n\right) &= \mathrm{Tr}\left[\{\rho^n - \gamma\tau^n \geq 0\}\left(\rho^n - \gamma\tau^n\right)\right] && (11.40)\\ &\leq \mathrm{Tr}\left[\{\rho^n - \gamma\tau^n \geq 0\}\,\rho^n\right], && (11.41) \end{aligned}$$

Eq. (11.39) gives

$$\alpha \geq 1 - \mathrm{Tr}\left[\{\rho^n - \gamma\tau^n \geq 0\}\,\rho_n\right] - \gamma \mathrm{e}^{-nR}. \tag{11.42}$$

Next, invoking Lemma 11.1 below, for all $\log\gamma > D\left(\rho^n \| \tau^n\right)$, we have

$$\begin{aligned} \alpha &\geq 1 - \frac{V\left(\rho^n \| \tau^n\right)}{\left[\log\gamma - D\left(\rho^n \| \tau^n\right)\right]^2} - \gamma \mathrm{e}^{-nR} && (11.43)\\ &= 1 - \frac{V\left(\bar{\mathcal{W}} \| \sigma | P_{\mathbf{x}^n}\right)}{n\left[\frac{\log\gamma}{n} - D\left(\bar{\mathcal{W}} \| \sigma | P_{\mathbf{x}^n}\right)\right]^2} - \gamma \mathrm{e}^{-nR} && (11.44) \end{aligned}$$

Finally, recall $D\left(\bar{\mathcal{W}} \| \sigma | P_{\mathbf{x}^n}\right) \leq R - 2\kappa$ and $A := \max_{\rho \in \mathcal{S}_\circ} V\left(\rho \| \sigma\right)$ and choose $\log\gamma = nD\left(\bar{\mathcal{W}} \| \sigma | P_{\mathbf{x}^n}\right) + n\kappa$. Then, Eq. (11.44) yields, for any test Q_n and $\beta(Q_n; \sigma^{\otimes n}) \leq \mathrm{e}^{-nR}$,

$$\begin{aligned} \alpha\left(Q_n; \bar{\mathcal{W}}_{\mathbf{x}^n}^{\otimes n}\right) &\geq 1 - \frac{V\left(\bar{\mathcal{W}} \| \sigma | P_{\mathbf{x}^n}\right)}{n\kappa^2} - \mathrm{e}^{-n\kappa} && (11.45)\\ &\geq 1 - \frac{A}{n\kappa^2} - \mathrm{e}^{-n\kappa}, && (11.46) \end{aligned}$$

which concludes the proof.

Lemma 11.1 (Quantum Chebyshev's Inequality [108, Lemma 6]). *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ and assume $\log\gamma > D(\rho\|\sigma)$. Then*

$$\mathrm{Tr}\left[\rho\left\{\rho - \gamma\sigma \geq 0\right\}\right] \leq \frac{V(\rho\|\sigma)}{\left[\log\gamma - D(\rho\|\sigma)\right]^2}. \tag{11.47}$$

□

11.3 A Strong Sphere-Packing Bound

The goal the section is to prove Theorem 11.1, the strong sphere-packing bound for c-q channels with a polynomial pre-factor. To establish this result, we combine Blahut's insight of relating a channel coding problem to binary hypothesis testing [32, 23] with a sharp concentration inequality introduced in Section 2.2. Our proof consists of three major steps: (i) reduce the channel coding problem to

doi:10.6342/NTU201800597

binary hypothesis testing (Proposition 11.2 in Section 11.3.1); (ii) bound its type-I error from below (Propositions 11.3 and 11.5 in Sections 11.3.2 and 11.3.3); (iii) employ Theorem 9.1 in Section 9.1 to relate the derived bound to the strong sphere-packing exponent. The proof of Theorem 11.1 and Corollary 11.1 will be give in Section 11.3.5.

11.3.1 One-Shot Converse Bound (Hypothesis Testing Reduction)

We first present a proof that relates the decoding error of a code to binary hypothesis testing. Proposition 11.2 below is similar to the meta-converse in Ref. [12]. However, the idea dates back to Blahut [32].

Proposition 11.2. *For any classical-quantum channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ and any code $\mathcal{C}_n$ with message size M, it follows that*

$$\varepsilon_{\max}\left(\mathcal{C}_n\right) \geq \max_{\sigma \in \mathcal{S}(H)} \min_{\mathbf{x}^n \in \mathcal{C}_n} \widehat{\alpha}_{\frac{1}{M}}\left(W_{\mathbf{x}^n}^{\otimes n} \| \sigma^{\otimes n}\right). \tag{11.48}$$

Proof of Proposition 11.2. Let $\mathbf{x}^n(m)$ be the codeword encoding the message $m \in \{1, \ldots, M\}$. Define a binary hypothesis testing problem:

$$\mathsf{H}_0 : W_{\mathbf{x}^n(m)}^{\otimes n}, \tag{11.49}$$

$$\mathsf{H}_1 : \sigma^n := \bigotimes_{i=1}^{n} \sigma_i, \tag{11.50}$$

where $\sigma^n \in \mathcal{S}\left(\mathcal{H}^{\otimes n}\right)$ can be viewed as a dummy channel output. Since $\sum_{m=1}^{M} \beta\left(\Pi_{n,m}; \sigma^n\right) = 1$ for any POVM $\Pi_n = \{\Pi_{n,1}, \ldots, \Pi_{n,M}\}$, and $\beta\left(\Pi_{n,m}; \sigma^n\right) \geq 0$ for every $m \in \mathcal{M}$, there must exist a message $m \in \mathcal{M}$ for any code $\mathcal{C}_n$ such that $\beta\left(\Pi_{n,m}; \sigma^n\right) \leq \frac{1}{M}$. Fix $\mathbf{x}^n := \mathbf{x}^n\left(m\right)$. Then

$$\varepsilon_{\max}\left(\mathcal{C}_n\right) \geq \varepsilon_m\left(\mathcal{C}_n\right) = \alpha\left(\Pi_{n,m}; W_{\mathbf{x}^n}^{\otimes n}\right) \geq \widehat{\alpha}_{\frac{1}{M}}\left(W_{\mathbf{x}^n}^{\otimes n} \| \sigma^n\right). \tag{11.51}$$

Since the above inequality (11.51) holds for every $\sigma^n \in \mathcal{S}\left(\mathcal{H}^{\otimes n}\right)$, it follows that

$$\varepsilon_{\max}\left(\mathcal{C}_n\right) \geq \max_{\sigma \in \mathcal{S}(H)} \min_{\mathbf{x}^n \in \mathcal{C}_n} \widehat{\alpha}_{\frac{1}{M}}\left(W_{\mathbf{x}^n}^{\otimes n} \| \sigma^{\otimes n}\right). \tag{11.52}$$

□

11.3.2 Chebyshev's Type Converse Bound

In the following Proposition, we generalize Blahut's one-shot converse Hoeffding bound [32, Theorem 10] to the quantum setting. This result is essentially a Chebyshev-type bound. We will employ it to lower bound the error of "bad sequences" that yield smaller error exponent in Section 11.3.5.

doi:10.6342/NTU201800597

Proposition 11.3 (Chebyshev's Type Converse Hoeffding Bound). *Consider the following binary hypothesis testing problem:* $\mathsf{H}_0 : \rho$ *versus* $\mathsf{H}_1 : \sigma$, *where* $\rho, \sigma \in \mathcal{S}(\mathcal{H})$. *For every* $r \geq 0$ *and* $\nu > 0$, *we have*

$$\widehat{\alpha}_{\frac{1}{4}\exp\{-(r+\nu)\}}(\rho\|\sigma) \geq \frac{1}{2}\left(\frac{1}{2} - \frac{K(\rho,\sigma)}{\nu^2}\right)\exp\{-\nu - \phi(r|\rho\|\sigma)\} \tag{11.53}$$

where

$$\phi(r|\rho\|\sigma) := \sup_{\alpha\in(0,1]}\left\{\frac{1-\alpha}{\alpha}\left(D_\alpha(\rho\|\sigma) - r\right)\right\}, \tag{11.54}$$

and

$$K(\rho,\sigma) := V(\hat{q}_t\|q) + V(\hat{q}_t\|p) \in \mathbb{R}_{\geq 0}, \tag{11.55}$$

where (p, q) *are the Nussbaum-Szkoła distributions of* (ρ, σ), *and*

$$\hat{q}_t(\omega) = \frac{p^{1-t}(\omega)q^t(\omega)}{\sum_{\omega\in\mathtt{supp}(p)\cap\mathtt{supp}(q)} p^{1-t}(\omega)q^t(\omega)}, \quad \omega \in \mathtt{supp}(p)\cap\mathtt{supp}(q) \tag{11.56}$$

for some $t \in [0, 1]$.

Proof of Prosition 11.3. If ρ and σ have disjoint supports, then Eq. (11.53) trivially holds since $D_\alpha(\rho\|\sigma) = +\infty$ for all $\alpha \in [0, 1]$. Hence, we assume ρ and σ have non-disjoint support in the following. Let $\mathcal{B} := \mathtt{supp}(p)\cap\mathtt{supp}(q)$ be the intersection of the joint support of p and q. Fix $\phi(r) := \phi(r|\rho\|\sigma) = \phi(r|p\|q)$ since $D_\alpha(\rho\|\sigma) = D_\alpha(p\|q)$.

For any test $0 \leq Q \leq \mathbb{1}$, Nagaoka showed that [111, Lemma 1] (see also [86, Proposition 2], [112]):

$$\alpha(Q;\rho) + \delta\beta(Q;\sigma) \geq \frac{1}{2}\left(\sum_{\omega:p(\omega)\leq\delta q(\omega)} p(\omega) + \sum_{\omega:p(\omega)>\delta q(\omega)} \delta q(\omega)\right), \quad \forall\delta \geq 0. \tag{11.57}$$

Let $r > 0$, $\delta = e^{r-\phi(r)}$, and $\mu \geq 0$ that will be specified later. Eq. (11.57) implies that

$$\widehat{\alpha}_\mu(\rho\|\sigma) \geq \frac{1}{2}\left(\sum_{\omega:p(\omega)\mathrm{e}^{\phi(r)}\leq q(\omega)\mathrm{e}^r} p(\omega) + \sum_{\omega:p(\omega)\mathrm{e}^{\phi(r)}>q(\omega)\mathrm{e}^r} \mathrm{e}^{r-\phi(r)}q(\omega)\right) - \mathrm{e}^{r-\phi(r)}\mu \tag{11.58}$$

$$\geq \frac{1}{2}\left(\sum_{\omega\in\mathcal{U}_1(\nu)} p(\omega) + \sum_{\omega\in\mathcal{U}_2(\nu)} \mathrm{e}^{r-\phi(r)}q(\omega)\right) - \mathrm{e}^{r-\phi(r)}\mu, \tag{11.59}$$

where in the last line we introduce the decision regions for some $\nu > 0$:

$$\mathcal{U}_1(\nu) := \left\{\omega : \hat{q}_t(\omega)\mathrm{e}^{-\nu} < p(\omega)\mathrm{e}^{\phi(r)} \leq q(\omega)\mathrm{e}^r\right\}, \quad \mathcal{U}_2(\nu) := \left\{\omega : \hat{q}_t(\omega)\mathrm{e}^{-\nu} < q(\omega)\mathrm{e}^r < p(\omega)\mathrm{e}^{\phi(r)}\right\}, \tag{11.60}$$

doi:10.6342/NTU201800597

and $\hat{q}_t$ is the *tilted distribution* (see [32, Theorem 4]):

$$\hat{q}_t(\omega) = \frac{p^{1-t}(\omega)q^t(\omega)}{\sum_{\omega\in\mathcal{B}} p^{1-t}(\omega)q^t(\omega)}, \quad \omega \in \mathcal{B} \tag{11.61}$$

for some $t \in [0,1]$ such that $\hat{q}_t$ satisfies

$$D(\hat{q}_t\|p) = \phi(r) \quad \text{and} \quad D(\hat{q}_t\|q) = r. \tag{11.62}$$

In the following, we are going to lower bound the right-hand side of Eq. (11.59) in terms of $\hat{q}_t$. From Eq. (11.60), we find

$$\begin{aligned} \sum_{\omega\in\mathcal{U}_1(\nu)} p(\omega) &\geq \mathrm{e}^{-(\phi(r)+\nu)} \sum_{\omega\in\mathcal{U}_1(\nu)} \hat{q}_t(\omega); \\ \sum_{\omega\in\mathcal{U}_2(\nu)} q(\omega) &\geq \mathrm{e}^{-(r+\nu)} \sum_{\omega\in\mathcal{U}_2(\nu)} \hat{q}_t(\omega). \end{aligned} \tag{11.63}$$

Next, we estimate the error in the union: $\sum_{\omega\in\mathcal{U}_1(\nu)\cup\mathcal{U}_2(\nu)} \hat{q}_t(\omega)$. Let

$$\mathcal{U}_A := \left\{\omega : \hat{q}_t(\omega)\mathrm{e}^{-\nu} < q(\omega)\mathrm{e}^{r}\right\}, \quad \mathcal{U}_B := \left\{\omega : \hat{q}_t(\omega)\mathrm{e}^{-\nu} < p(\omega)\mathrm{e}^{\phi(r)}\right\}. \tag{11.64}$$

Observe that $\mathcal{U}_1(\nu)\cup\mathcal{U}_2(\nu) = \mathcal{U}_A \cap \mathcal{U}_B$ and

$$\sum_{\omega\in\mathcal{U}_A\cap\mathcal{U}_B} \hat{q}_t(\omega) \geq 1 - \sum_{\omega\in\mathcal{U}_A^{\mathrm{c}}} \hat{q}_t(\omega) - \sum_{\omega\in\mathcal{U}_B^{\mathrm{c}}} \hat{q}_t(\omega). \tag{11.65}$$

Denote by

$$\mathcal{U}_T := \left\{\omega : \left|\log\frac{\hat{q}_t(\omega)}{q(\omega)}\mathrm{e}^{-r}\right| \geq \nu\right\} \tag{11.66}$$

$$= \left\{\omega : \left|\log\frac{\hat{q}_t(\omega)}{q(\omega)} - \sum_{\omega\in\mathcal{B}} \hat{q}_t(\omega)\log\frac{\hat{q}_t(\omega)}{q(\omega)}\right| \geq \nu\right\}, \tag{11.67}$$

where the last equality follows from Eq. (11.62). Since $\mathcal{U}_A^{\mathrm{c}} \subseteq \mathcal{U}_T$, we apply Chebyshev's inequality to obtain

$$\sum_{\omega\in\mathcal{U}_A^{\mathrm{c}}} \hat{q}_t(\omega) \leq \sum_{\omega\in\mathcal{U}_T} \hat{q}_t(\omega) \leq \frac{V(\hat{q}_t\|q)}{\nu^2}. \tag{11.68}$$

Similarly,

$$\sum_{\omega\in\mathcal{U}_B^{\mathrm{c}}} \hat{q}_t(\omega) \leq \frac{V(\hat{q}_t\|p)}{\nu^2}. \tag{11.69}$$

Let $K = K(\rho,\sigma) := V(\hat{q}_t\|q) + V(\hat{q}_t\|p)$. Equation (11.65), along with (11.68) and (11.69) yields that

$$\sum_{\omega\in\mathcal{U}_1(\nu)\cup\mathcal{U}_2(\nu)} \hat{q}_t(\omega) = \sum_{\omega\in\mathcal{U}_A\cap\mathcal{U}_B} \hat{q}_t(\omega) \geq 1 - \frac{K}{\nu^2}. \tag{11.70}$$

doi:10.6342/NTU201800597

Hence, from Eqs. (11.59), (11.63), and (11.70), we obtain the lower bound of the type-I error:

$$\widehat{\alpha}_{\mu}\left(\rho\|\sigma\right) \geq \frac{1}{2}\left(\sum_{\omega\in\mathcal{U}_1(\nu)} p(\omega) + \sum_{\omega\in\mathcal{U}_2(\nu)} \mathrm{e}^{r-\phi(r)} q(\omega)\right) - \mathrm{e}^{r-\phi(r)}\mu, \tag{11.71}$$

$$\geq \frac{1}{2}\mathrm{e}^{-(\phi(r)+\nu)}\left(\sum_{\omega\in\mathcal{U}_1(\nu)} \hat{q}_t(\omega) + \sum_{\omega\in\mathcal{U}_2(\nu)} \hat{q}_t(\omega)\right) - \mathrm{e}^{r-\phi(r)}\mu \tag{11.72}$$

$$\geq \frac{1}{2}\mathrm{e}^{-(\phi(r)+\nu)}\left(\sum_{\omega\in\mathcal{U}_1(\nu)\cup\mathcal{U}_2(\nu)} \hat{q}_t(\omega)\right) - \mathrm{e}^{r-\phi(r)}\mu \tag{11.73}$$

$$\geq \frac{1}{2}\mathrm{e}^{-(\phi(r)+\nu)}\left(1-\frac{K}{\nu^2}\right) - \mathrm{e}^{r-\phi(r)}\mu. \tag{11.74}$$

Choose $\mu = \frac{1}{4}\exp\{-(r+\nu)\}$. Eq. (11.74) further gives

$$\widehat{\alpha}_{\frac{1}{4}\exp\{-(r+\nu)\}}\left(\rho\|\sigma\right) \geq \frac{1}{2}\mathrm{e}^{-(\phi(r)+\nu)}\left(1-\frac{K}{\nu^2}\right) - \frac{1}{4}\mathrm{e}^{-(\phi(r)+\nu)} \tag{11.75}$$

$$= \frac{1}{2}\left(\frac{1}{2}-\frac{K}{\nu^2}\right)\mathrm{e}^{-(\phi(r)+\nu)}, \tag{11.76}$$

which completes the proof. □

Applying Proposition 11.3 to product states yields the following result.

Proposition 11.4 (Chebyshev-Type Converse Bound for Classical-Quantum Channels). *Let $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, and let $R \in (R_\infty, C_{\mathcal{W}})$. Consider the binary hypothesis testing with sequences*

$$\mathsf{H}_0 : \rho^n = W_{\mathbf{x}^n}^{\otimes n}; \tag{11.77}$$

$$\mathsf{H}_1 : \sigma^n = \left(\sigma^{\star}_{R,P_{\mathbf{x}^n}}\right)^{\otimes n}, \tag{11.78}$$

where $\mathbf{x}^n \in \mathcal{X}^n$ and $\sigma^{\star}_{R,P} \in \arg\min_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{0<\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{W}\|\sigma|P_{\mathbf{x}^n}\right) - R\right)$. Then, for every $c > 0$, there exist $N_0 \in \mathbb{N}$ and $\kappa_1, \kappa_2 \in \mathbb{R}_{>0}$ such that for all $n \geq N_0$ we have

$$\widehat{\alpha}_{c\exp\{-nR\}}\left(\rho^n\|\sigma^n\right) \geq \kappa_1 \exp\left\{-\kappa_2\sqrt{n} - nE_{\mathrm{sp}}^{(2)}\left(R, P_{\mathbf{x}^n}\right)\right\}, \tag{11.79}$$

Remark 11.2. Consider independent and identically distributed (i.i.d.) extensions $\mathsf{H}_0 : \rho^{\otimes n}$ and $\mathsf{H}_1 : \sigma^{\otimes n}$. Proposition 11.4 then recovers the converse proof of the *quantum Hoeffding bound* (see [111] and [85, Section 5.4]): for $r \in (0, D(\rho\|\sigma))$,

$$\lim_{n\to+\infty} -\frac{1}{n}\log\widehat{\alpha}_{\exp\{-nr\}}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \leq \sup_{0<\alpha\leq 1}\frac{1-\alpha}{\alpha}\left(D_\alpha(\rho\|\sigma) - r\right). \tag{11.80}$$

◇

Proof of Proposition 11.4. Denote by $p^n = \bigotimes_{i=1}^n p_{x_i}$, $q^n = \bigotimes_{i=1} q_{x_i}$ Nussbaum-Szkoła distributions of ρ^n and σ^n [112] with joint supports $\mathcal{B}_{x_i} := \mathtt{supp}(p_{x_i}) \cap \mathtt{supp}(q_{x_i})$, $i \in [n]$. Let $R_n := R - \gamma_n$, where $\gamma_n := \frac{\nu+\log 4c}{n}$. Fix an arbitrary $R_0 \in (R_\infty, R)$. Choose an $N_0 \in \mathbb{N}$ such that $R_n \geq R_0$ for all $n \geq N_0$.

doi:10.6342/NTU201800597

Consider $n \geq N_0$ onwards. Then, Proposition 11.3 implies that

$$\widehat{\alpha}_{c\exp\{-nR\}}\left(\rho^n \| \sigma^n\right) \geq \frac{1}{2}\left(\frac{1}{2} - \frac{K(\rho^n, \sigma^n)}{\nu^2}\right) \exp\left\{-\nu - n\phi_n\left(R_n | \rho^n \| \sigma^n\right)\right\} \tag{11.81}$$

$$= \frac{1}{2}\left(\frac{1}{2} - \frac{K(\rho^n, \sigma^n)}{\nu^2}\right) \exp\left\{-\nu - nE_{\mathrm{sp}}^{(2)}\left(R_n, P_{\mathbf{x}^n}\right)\right\}, \tag{11.82}$$

where the second equality (11.82) follows from the saddle-point property, item (a) in Proposition 9.5. Since the coefficient $K(\rho^n, \sigma^n)$ in Eq. (11.55) is additive for product states, one has

$$K\left(\rho^n, \sigma^n\right) = V\left(\hat{q}_t^n \| p^n\right) + V\left(\hat{q}_t^n \| q^n\right) \tag{11.83}$$

$$= n \sum_{x \in \mathcal{X}} P_{\mathbf{x}^n}(x) \left[V\left(\hat{q}_{x,t} \| p_x\right) + V\left(\hat{q}_{x,t} \| q_x\right)\right], \tag{11.84}$$

where $P_{\mathbf{x}^n}$ is the empirical distribution for the sequence $\mathbf{x}^n$, and $\hat{q}_t^n := \bigotimes_{i=1}^n \hat{q}_{x_i,t}$ is the tilted distribution (see Eqs. (11.56) and (11.61)). Note that $\hat{q}_t^n \ll p^n$ and $\hat{q}_t^n \ll q^n$ for all $t \in [0,1]$. This guarantees that the quantity $K(\rho^n, \sigma^n)$ is finite.

Let

$$V_{\max} := \max_{t \in [0,1],\, P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X})} \sum_{x \in \mathcal{X}} P_{\mathbf{x}^n}(x) \left[V\left(\hat{q}_{x,t} \| p_x\right) + V\left(\hat{q}_{x,t} \| q_x\right)\right] \in \mathbb{R}_{>0}, \tag{11.85}$$

we obtain

$$K\left(\rho^n, \sigma^n\right) \leq n V_{\max}. \tag{11.86}$$

By choosing $\nu = \sqrt{4nV_{\max}}$, Eqs. (11.82) and (11.86) give

$$\widehat{\alpha}_{c\exp\{-nR\}}\left(\rho^n \| \sigma^n\right) \geq \frac{1}{8} \exp\left\{-\sqrt{4nV_{\max}} - nE_{\mathrm{sp}}^{(2)}\left(R - \gamma_n, P_{\mathbf{x}^n}\right)\right\}. \tag{11.87}$$

Finally, we will remove the rate back-off term γ_n in Eq. (11.87). Recall item (a) in Proposition 9.6 that the map $r \mapsto E_{\mathrm{sp}}^{(2)}(r, P_{\mathbf{x}^n})$ is convex and monotone decreasing. Further, we assume $E_{\mathrm{sp}}^{(2)}(R_0, P_{\mathbf{x}^n}) > 0$ and thus the $E_{\mathrm{sp}}^{(2)}(\cdot, P_{\mathbf{x}^n})$ is differentiable at R_0 by item (c) in Proposition 9.6. Otherwise, the monotone decreases imply that $E_{\mathrm{sp}}^{(2)}(R, P_{\mathbf{x}^n}) = E_{\mathrm{sp}}^{(2)}(R_0, P_{\mathbf{x}^n}) = 0$, which already completes the proof. Denoting by ∂_- the left derivative, the convexity then implies that

$$E_{\mathrm{sp}}^{(2)}(R - \gamma_n, P_{\mathbf{x}^n}) \leq E_{\mathrm{sp}}^{(2)}(R, P_{\mathbf{x}^n}) - \gamma_n \partial_- E_{\mathrm{sp}}^{(2)}(R - \gamma_n, P_{\mathbf{x}^n}), \tag{11.88}$$

$$\leq E_{\mathrm{sp}}^{(2)}(R, P_{\mathbf{x}^n}) - \gamma_n \left.\frac{\partial E_{\mathrm{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r}\right|_{r=R_0}, \tag{11.89}$$

where the last inequality (11.89) follows from the monotone decreases. Let

$$\Upsilon := \max_{P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X})} \left| \left.\frac{\partial E_{\mathrm{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r}\right|_{r=R_0} \right|. \tag{11.90}$$

Note that $\Upsilon \in \mathbb{R}_{\geq 0}$ due to $R_0 > R_\infty$ and item (d) of Proposition 9.6. Then, Eqs. (11.87), (11.89), and

doi:10.6342/NTU201800597

(11.90) lead to

$$\widehat{\alpha}_{c\exp\{-nR\}}\left(\rho^n\|\sigma^n\right) \geq \frac{1}{8}\exp\left\{-\sqrt{4nV_{\max}} - \gamma_n\Upsilon - nE_{\mathrm{sp}}^{(2)}\left(R, P_{\mathbf{x}^n}\right)\right\}. \tag{11.91}$$

Setting $\kappa_1 = 1/8$ and choosing a constant $\kappa_2 \in \mathbb{R}_{>0}$ such that $\sqrt{4nV_{\max}} + \gamma_n\Upsilon \leq \kappa_2\sqrt{n}$ for all $n \geq N_0$ conclude this corollary. □

11.3.3 A Sharp Converse Bound

Proposition 11.5 (Sharp Converse Hoeffding Bound). *Let $\mathcal{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, and let $R \in (R_\infty, C_{\mathcal{W}})$. Consider the following binary hypothesis testing problem with sequences*

$$\mathsf{H}_0 : \rho^n = W_{\mathbf{x}^n}^{\otimes n}; \tag{11.92}$$

$$\mathsf{H}_1 : \sigma^n = \left(\sigma^\star_{R,P_{\mathbf{x}^n}}\right)^{\otimes n}, \tag{11.93}$$

where $\mathbf{x}^n \in \mathcal{X}^n$, and $\sigma^\star_{R,P} := \arg\min_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{0<\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{W}\|\sigma|P_{\mathbf{x}^n}\right) - R\right)$ satisfying

$$E_{\mathrm{sp}}^{(2)}\left(R, P_{\mathbf{x}^n}\right) \in [\nu, +\infty) \tag{11.94}$$

for some positive $\nu > 0$. For every $c > 0$, there exists a constant $N_0 \in \mathbb{N}$, independent of the sequences ρ^n and σ^n, such that for all $n \geq N_0$ we have

$$\widehat{\alpha}_{c\exp\{-nR\}}\left(\rho^n\|\sigma^n\right) \geq \frac{A}{n^{\frac{1}{2}\left(1+s^\star_{R,P_{\mathbf{x}^n}}\right)}}\exp\left\{-nE_{\mathrm{sp}}^{(2)}\left(R, P_{\mathbf{x}^n}\right)\right\}, \tag{11.95}$$

where $s^\star_{R,P} := -\left.\frac{\partial E_{\mathrm{sp}}^{(2)}(r,P)}{\partial r}\right|_{r=R}$, and $A \in \mathbb{R}_{>0}$ is a finite constant depending on R, ν and $\mathcal{W}$.

Proof. Let $p^n := \bigotimes_{i=1}^n p_{x_i}$ and $q^n := \bigotimes_{i=1}^n q_{x_i}$, where (p_{x_i}, q_{x_i}) are Nussbaum-Szkoła distributions [112] of $(W_{x_i}, \sigma^\star)$ for every $i \in [n]$. Since $D_\alpha(\rho_{x_i}\|\sigma_{x_i}) = D_\alpha(p_{x_i}\|q_{x_i})$, for $\alpha \in (0, 1]$, again we shorthand

$$\phi_n(r) := \phi_n\left(r|\rho^n\|\sigma^n\right) = \phi_n(r|p^n\|q^n) = E_{\mathrm{sp}}^{(2)}\left(r, W, P_{\mathbf{x}^n}\right), \tag{11.96}$$

where the last equality in Eq. (11.96) follows from the saddle-point property, item (i) in Proposition 9.5. Moreover, item (iii) in Proposition 9.5 implies that the state $\sigma^\star$ dominants all the states: $\sigma^\star \gg W_x$, for all $x \in \mathtt{supp}(P_{\mathbf{x}^n})$, Hence, we have $p^n \ll q^n$. In the following, we set zero all element of q_{x_i} that do not lie in the support of p_{x_i}, i.e. $q_{x_i}(\omega) = 0$, $\omega \notin \mathtt{supp}(p_{x_i})$, $i \in [n]$.

Repeating Nagaoka's argument [111] in Eq. (11.57) for any $0 \leq Q_n \leq \mathbb{1}$ and choosing $\delta = \exp\{nr - n\phi_n(r)\}$ yield:

$$\alpha\left(Q_n; \rho^n\right) + \delta\beta\left(Q_n; \sigma^n\right) \geq \frac{1}{2}\left(\alpha\left(\mathcal{U}; p^n\right) + \mathrm{e}^{nr-n\phi_n(r)}\beta\left(\mathcal{U}; q^n\right)\right), \tag{11.97}$$

doi:10.6342/NTU201800597

where

$$\alpha\left(\mathcal{U}; p^n\right) := \sum_{\omega \in \mathcal{U}^{\mathrm{c}}} p^n(\omega) \tag{11.98}$$

$$\beta\left(\mathcal{U}; q^n\right) := \sum_{\omega \in \mathcal{U}} q^n(\omega) \tag{11.99}$$

and

$$\mathcal{U} := \left\{\omega : p^n(\omega) \mathrm{e}^{n\phi_n(r)} > q^n(\omega) \mathrm{e}^{nr}\right\}. \tag{11.100}$$

In the following, we

In the following, we will relate the Fenchel-Legendre transform $\Lambda^*_{j,P_{\mathbf{x}^n}}(z)$ to the desired error-exponent function $\phi_n(r)$. Such a relationship was presented Lemma 2.14 in Section 2.2. Since the Lemma 2.14 (a) in Section 2.2 shows that the optimizer t in Eq. (4.48) always lies in the compact set $H := [0, 1]$, by invoking Eq. (11.173) we define the following quantities:

$$V_{\max}(r, \nu) := \max_{t \in H,\, P_{\mathbf{x}^n} \in \mathcal{P}_{r,\nu}} \Lambda''_{0,P_{\mathbf{x}^n}}(t); \tag{11.101}$$

$$V_{\min}(r, \nu) := \min_{t \in H,\, P_{\mathbf{x}^n} \in \mathcal{P}_{r,\nu}} \Lambda''_{0,P_{\mathbf{x}^n}}(t); \tag{11.102}$$

$$K_{\max}(r, \nu) := 15\sqrt{2\pi} M_{0,\max}; \tag{11.103}$$

$$M_{\max} := \max_{t \in H,\, P_{\mathbf{x}^n} \in \mathcal{P}_{r,\nu}} \frac{T_{0,P_{\mathbf{x}^n}}(t)}{\Lambda''_{0,P_{\mathbf{x}^n}}(t)}; \tag{11.104}$$

$$T_{0,P_{\mathbf{x}^n}}(t) := \sum_{x \in \mathcal{X}} P_{\mathbf{x}^n}(x) \mathbb{E}_{\hat{q}_{x,t}}\left[\left|\log \frac{q_x}{p_x} - \Lambda'_{0,x}(t)\right|^3\right], \tag{11.105}$$

where we define $\mathcal{P}_{r,\nu} := \{P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X}) : V\left(W \| \sigma | P_{\mathbf{x}^n}\right) \in [\nu, +\infty)\}$ for condition in Eq. (11.94) or define $\mathcal{P}_{r,\nu} := \{P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X}) : \phi_n(r) \in [\nu, +\infty)\}$ for condition in Eq. (11.94). Either way, $\mathcal{P}_{r,\nu}$ is a compact set. The uniform continuity, Proposition 11.6, in Section 11.3.4 below shows that $\Lambda''_{0,(\cdot)}(\cdot)$ and $T_{0,(\cdot)}(\cdot)$ are continuous functions in $(0, 1] \times P_{r,0}$. Hence, the maximization and minimization in the above definitions are well-defined and finite. Further, the quantity $V_{\min}(r, \nu)$ is bounded away from zero owing to Lemma 2.14 (a) in Section 2.2.

Now, we are ready to derive the lower bounds to $\alpha\left(\mathcal{U}; p^n\right)$ and $\beta\left(\mathcal{U}; q^n\right)$. If $n \in \mathbb{N}$ is sufficiently large such that

$$\sqrt{n} \geq N_0(r, \nu) := \frac{1 + (1 + K_{\max}(\nu, r))^2}{\sqrt{V_{\min}(r, \nu)}} \tag{11.106}$$

applying Bahadur-Randga Rao's inequality (Theorem 2.1) to $Z_i = \log q_i - \log p_i$ with probability

doi:10.6342/NTU201800597

measure $\lambda_i = p_i$, and $z = r - \phi_n(r)$ gives

$$\alpha\left(\mathcal{U}; p^n\right) := \sum_{\omega \in \mathcal{U}^{\mathrm{c}}} p^n(\omega) \tag{11.107}$$

$$= \Pr\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i \geq r - \phi_n(r)\right\} \tag{11.108}$$

$$\geq \frac{A(r,\nu)}{\sqrt{n}} \exp\left\{-n\Lambda^*_{0,P_{\mathbf{x}^n}}\left(\phi_n(r) - r\right)\right\} \tag{11.109}$$

where

$$A(r,\nu) := \frac{\mathrm{e}^{-K_{\max}(r,\nu)}}{2\sqrt{4\pi V_{\max}(r,\nu)}}. \tag{11.110}$$

Similarly, applying Theorem 2.1 to $Z_i = \log p_i - \log q_i$ with probability measure $\lambda_i = q_i$, and $z = \phi_n(r) - r$ yields

$$\beta\left(\mathcal{U}; q^n\right) := \sum_{\omega \in \mathcal{U}} q^n(\omega) \tag{11.111}$$

$$= \Pr\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i \geq \phi_n(r) - r\right\} \tag{11.112}$$

$$\geq \frac{A(r,\nu)}{\sqrt{n}} \exp\left\{-n\Lambda^*_{1,P_{\mathbf{x}^n}}\left(r - \phi_n(r)\right)\right\}. \tag{11.113}$$

Continuing from Eq. (11.109) and item (ii) in Lemma 2.14 gives

$$\alpha\left(\mathcal{U}; p^n\right) \geq \frac{A(r,\nu)}{\sqrt{n}} \mathrm{e}^{-n\phi_n(r)}. \tag{11.114}$$

Eq. (11.113) together with item (iii) in Lemma 2.14 yields

$$\beta\left(\mathcal{U}; q^n\right) \geq \frac{A(r,\nu)}{\sqrt{n}} \mathrm{e}^{-nr}. \tag{11.115}$$

Thus we can bound the left-hand side of Eq. (11.97) from below by $\frac{A(r,\nu)}{\sqrt{n}}\mathrm{e}^{-n\phi_n(r)}$. For any test $0 \leq Q_n \leq \mathbb{1}$ such that

$$\beta(Q_n; \sigma^n) \leq \frac{A(r,\nu)}{2\sqrt{n}} \mathrm{e}^{-nr}, \tag{11.116}$$

we have that

$$\alpha(Q_n; \rho^n) \geq \frac{A(r,\nu)}{2\sqrt{n}} \mathrm{e}^{-n\phi_n(r)}. \tag{11.117}$$

By letting $A'(r,\nu) = A(r,\nu)/2$ and $r' = r + \frac{1}{n}\log(\sqrt{n}/A'(r,\nu))$, we conclude that

$$\widehat{\alpha}_{\exp\{-nr'\}}\left(\rho^n \| \sigma^n\right) \geq \frac{A'(r,\nu)}{\sqrt{n}} \exp\left\{-n\phi_n\left(r' - \frac{1}{n}\log\frac{\sqrt{n}}{A'(r,\nu)}\middle|\middle|\rho^n\middle\|\sigma^n\right)\right\} \tag{11.118}$$

$$= \frac{A'(r,\nu)}{\sqrt{n}} \exp\left\{-nE^{(2)}_{\mathrm{sp}}\left(r' - \frac{1}{n}\log\frac{\sqrt{n}}{A'(r,\nu)}, W, P_{\mathbf{x}^n}\right)\right\}. \tag{11.119}$$

doi:10.6342/NTU201800597

□

11.3.4 Uniform Continuity

In this section, we prove a uniform continuity property, which is crucial to establish the finite blocklength bounds in error exponent analysis.

We first introduce necessary notation. Fix $R \in (C_{0,\mathcal{W}}, C_{1,\mathcal{W}})$, and denote by $(\alpha^\star_{R,P}, \sigma^\star_{R,P})$ the saddle-point of $F(R,P)$ for any $P \in \mathcal{P}_R(\mathcal{X})$. Define

$$B_\alpha(P,\mathcal{W}) := \sum_{x\in\mathcal{X}} P(x)\mathbb{E}_{v_{\alpha,x}}\left[\log\frac{p_x}{q_x}\right]; \tag{11.120}$$

$$V_\alpha(P,\mathcal{W}) := \sum_{x\in\mathcal{X}} P(x)\mathbb{E}_{v_{\alpha,x}}\left[\left|\log\frac{p_x}{q_x} - \mathbb{E}_{v_{\alpha,x}}\left[\log\frac{p_x}{q_x}\right]\right|^2\right]; \tag{11.121}$$

$$T_\alpha(P,\mathcal{W}) := \sum_{x\in\mathcal{X}} P(x)\mathbb{E}_{v_{\alpha,x}}\left[\left|\log\frac{p_x}{q_x} - \mathbb{E}_{v_{\alpha,x}}\left[\log\frac{p_x}{q_x}\right]\right|^3\right], \tag{11.122}$$

where (p_x, q_x) is the Nussbaum-Szkoła distribution of $(W_x, \sigma^\star_{R,P})$, and the *tilted distribution* is

$$v_{\alpha,x}(i,j) := \frac{p_x^\alpha(i,j)q_x^{1-\alpha}(i,j)}{\sum_{\imath,\jmath} p_x^\alpha(\imath,\jmath)q_x^{1-\alpha}(\imath,\jmath)}, \quad \alpha\in[0,1]. \tag{11.123}$$

Inspired by Ref. [12, Lemma 62], we show the following continuity property, which are crucial for establishing the large deviation bounds in finite blocklength regime.

Proposition 11.6 (Uniform Continuity)**.** *The functions $B_\alpha(P,\mathcal{W})$, $V_\alpha(P,\mathcal{W})$, and $T_\alpha(P,\mathcal{W})$ are jointly continuous on $(\alpha,P)\in[0,1]\times\mathcal{P}(\mathcal{X})$.*

Proof of Proposition 11.6. It is not hard to see that the quantities $B_\alpha(P,\mathcal{W})$, $V_\alpha(P,\mathcal{W})$, and $T_\alpha(P,\mathcal{W})$ are sums of finitely many terms. We thus show that each term is continuous. Fix an arbitrary $x\in\mathcal{X}$ onwards. Let $(\alpha_k,P_k)_{k\in\mathbb{K}}$ be an arbitrary sequence such that $(\alpha_k,P_k)\in[0,1]\times\mathcal{P}(\mathcal{X})$, and $\lim_{k\to+\infty}(\alpha_k,P_k) = (\alpha_0,P_0)\in[0,1]\times\mathcal{P}(\mathcal{X})$. Given the eigenvalue decompositions $W_x = \sum_i \lambda_i|e_i\rangle\langle e_i|$ and $\sigma^\star_{R,P_k} = \sum_j \mu_j(\sigma^\star_{R,P_k})|f_j^k\rangle\langle f_j^k|$, we have the Nussbaum-Szkoła distribution $p_x(i,j) = \lambda_i|\langle e_i|f_j^k\rangle|^2$ and $q_x(i,j) = \mu_j(\sigma^\star_{R,P_k})|\langle e_i|f_j^k\rangle|^2$. Here, we write f_j^k and $\mu_j(\sigma^\star_{R,P_k})$ to emphasize the dependence on P_k.

To prove the continuity of $B_\alpha(P,\mathcal{W})$, it suffices to show

$$\begin{aligned} &P_k(x)\frac{1}{\mathrm{Tr}\left[W_x^{\alpha_k}(\sigma^\star_{R,P_k})^{1-\alpha_k}\right]}\lambda_i^{\alpha_k}\mu_j^{1-\alpha_k}(\sigma^\star_{R,P_k})|\langle e_i|f_j^k\rangle|^2\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P_k})} \\ \to &P_0(x)\frac{1}{\mathrm{Tr}\left[W_x^{\alpha_0}(\sigma^\star_{R,P_0})^{1-\alpha_0}\right]}\lambda_i^{\alpha_0}\mu_j^{1-\alpha_0}(\sigma^\star_{R,P_0})|\langle e_i|f_j^0\rangle|^2\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P_0})}. \end{aligned} \tag{11.124}$$

If $\lambda_i = 0$, then it is obvious (recalling that the power function is only acting on the support). We assume $\lambda_i > 0$. If $P_0(x) > 0$, then $W_x \ll \sigma^\star_{R,P_k}$ for all sufficiently large $k\in\mathbb{N}$ and $k = 0$. Further, if $\mu_j(\sigma^\star_{R,P_k})|\langle e_i|f_j^k\rangle| = 0$, we have $\lambda_i|\langle e_i|f_j^k\rangle| = 0$ by the absolute continuity, which in turn implies the convergence of Eq. (11.124). Considering the other case, we can deduce that $\mu_j(\sigma^\star_{R,P_k})$ is bounded away from zero. Using the continuity of $P\mapsto\sigma^\star_{R,P}$ and logarithm, $\log\lambda_i/\mu_j(\sigma^\star_{R,P_k})$ tends to $\log\lambda_i/\mu_j(\sigma^\star_{R,P_0})$.

doi:10.6342/NTU201800597

It remains to show the case of $P_0(x) = 0$. To that end, we want to show $\log[\lambda_i/\mu_j(\sigma^\star_{R,P_k})] = O(\log 1/P_k)$. We may assume $P_k(x) > 0$ and $\mu_j(\sigma^\star_{R,P_k}) > 0$ for all $k \in \mathbb{N}$. The saddle-point property guarantees that $\sigma^\star_{R,P_k}$ must satisfy

$$\sigma^\star_{R,P_k} = \left(\sum_{\bar{x}\in\mathcal{X}} P_k(x) \frac{W_{\bar{x}}^{\alpha^\star_{R,P_k}}}{\mathrm{Tr}\left[W_{\bar{x}}^{\alpha^\star_{R,P_k}} (\sigma^\star_{R,P_k})^{1-\alpha^\star_{R,P_k}} \right]} \right)^{\frac{1}{\alpha^\star_{R,P_k}}}. \tag{11.125}$$

Further, noting that $\alpha^\star_{R,P} \in (0,1]$ for all $P \in \mathcal{P}(\mathcal{X})$, the continuity of $P \mapsto \alpha^\star_{R,P}$ and the compactness of $\mathcal{P}(\mathcal{X})$ imply that

$$\bar{\alpha}_R := \min_{P\in\mathcal{P}(\mathcal{X})} \alpha^\star_{R,P} > 0 \tag{11.126}$$

Therefore,

$$\begin{aligned}
\mu_j(\sigma^\star_{R,P_k}) &\geq \tilde{\lambda}_{\min}(\sigma^\star_{R,P_k}) && (11.127)\\
&\geq \tilde{\lambda}_{\min}^{\frac{1}{\alpha^\star_{R,P_k}}} \left(\sum_{\bar{x}} P_k(\bar{x}) \frac{W_{\bar{x}}^{\alpha^\star_{R,P_k}}}{\mathrm{Tr}\left[W_{\bar{x}}^{\alpha^\star_{R,P_k}} (\sigma^\star_{R,P_k})^{1-\alpha^\star_{R,P_k}} \right]} \right) && (11.128)\\
&\geq \tilde{\lambda}_{\min}^{\frac{1}{\alpha^\star_{R,P_k}}} \left(\sum_{\bar{x}} P_k(\bar{x}) W_{\bar{x}}^{\alpha^\star_{R,P_k}} \right) && (11.129)\\
&\geq \tilde{\lambda}_{\min}^{\frac{1}{\alpha^\star_{R,P_k}}} \left(P_k(x) W_x^{\alpha^\star_{R,P_k}} \right) && (11.130)\\
&= P_k^{\frac{1}{\alpha^\star_{R,P_k}}}(x) \tilde{\lambda}_{\min}(W_x) && (11.131)\\
&\geq P_k^{\frac{1}{\bar{\alpha}_{\min}}}(x) \tilde{\lambda}_{\min}(W_x). && (11.132)
\end{aligned}$$

where we denote by $\tilde{\lambda}_{\min}$ the smallest non-zero eigenvalue, and Eq. (11.129) holds because for any P_k,

$$\mathrm{Tr}\left[W_x^{\alpha^\star_{R,P_k}} (\sigma^\star_{R,P_k})^{1-\alpha^\star_{R,P_k}} \right] \in [0,1]. \tag{11.133}$$

Note that $\mu_j(\sigma^\star_{R,P_k}) \leq 1$. Eq. (11.132) then implies

$$\begin{aligned}
\left| \log \frac{\lambda_i}{\mu_j(\sigma^\star_{R,P_k})} \right| &\leq \log \frac{1}{\lambda_i} - \log P_k^{\frac{1}{\bar{\alpha}_{\min}}}(x) \tilde{\lambda}_{\min}(W_x) && (11.134)\\
&\leq 2 \log \frac{1}{\tilde{\lambda}_{\min}(W_x)} - \log P_k^{\frac{1}{\bar{\alpha}_{\min}}}(x). && (11.135)
\end{aligned}$$

doi:10.6342/NTU201800597

Using Eq. (11.135), we are able show that the left-hand side of Eq. (11.124) converges to 0:

$$P_k(x)\frac{1}{\operatorname{Tr}\left[W_x^{\alpha_k}(\sigma^\star_{R,P_k})^{1-\alpha_k}\right]}\lambda_i^{\alpha_k}\mu_j^{1-\alpha_k}(\sigma^\star_{R,P_k})|\langle e_i|f_j^k\rangle|^2\left|\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P_k})}\right| \tag{11.136}$$

$$\le P_k(x)\left|\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P_k})}\right| \tag{11.137}$$

$$\le 2P_k(x)\log\frac{1}{\tilde{\lambda}_{\min}(W_x)} - P_k(x)\log P_k^{\frac{1}{\tilde{\alpha}_{\min}}}(x)\tilde{\lambda}_{\min}(W_x) \tag{11.138}$$

$$\to 0, \tag{11.139}$$

which proves the continuity of $B_\alpha(P,\mathcal{W})$.

Next, we show the continuity of $V_\alpha(P,\mathcal{W})$ and $T_\alpha(P,\mathcal{W})$. Denote by $B_\alpha(W_x\|\sigma^\star_{R,P}) := \mathbb{E}_{v_{\alpha,x}}[\log p_x/q_x]$ for convenience. For $P_0(x)>0$, $\mu_j(\sigma^\star_{R,P_k})$ is bounded away from zero. Then, $\log\lambda_i/\mu_j(\sigma^\star_{R,P_k})$ tends to $\log\lambda_i/\mu_j(\sigma^\star_{R,P_0})$, and it is not hard to see that $B_{\alpha_k}(W_x\|\sigma\sigma^\star_{R,P_k}) \to B_{\alpha_0}(W_x\|\sigma\sigma^\star_{R,P_0})$. It suffices to prove the convergence when $P_k(x)\to 0$. Eq. (11.135) immediately implies that

$$B_{\alpha_k}(W_x\|\sigma^\star_{R,P_k}) = \sum_{i,j}\frac{1}{\operatorname{Tr}\left[W_x^{\alpha_k}(\sigma^\star_{R,P_k})^{1-\alpha_k}\right]}\lambda_i^{\alpha_k}\mu_j^{1-\alpha_k}(\sigma^\star_{R,P_k})|\langle e_i|f_j^k\rangle|^2\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P_k})} \tag{11.140}$$

$$\le 2\log\frac{1}{\tilde{\lambda}_{\min}(W_x)} - \log P_k^{\frac{1}{\tilde{\alpha}_{\min}}}(x). \tag{11.141}$$

Using the inequality $|a+b|^2\le 2(|a|^2+|b|^2)$, we obtain

$$P_k(x)\left|\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P})} - B_{\alpha_k}(W_x\|\sigma^\star_{R,P_k})\right|^2 \le 2P_k(x)\left|\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P_k})}\right|^2 + 2P_k(x)B^2_{\alpha_k}(W_x\|\sigma^\star_{R,P_k}). \tag{11.142}$$

Combining Eqs. (11.135), (11.141), and (11.142), we prove the continuity of $V_\alpha(P,\mathcal{W})$.

Similarly, using the inequality $|a+b|^3\le 4(|a|^3+|b|^3)$ gives

$$P_k(x)\left|\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P})} - B_{\alpha_k}(W_x\|\sigma^\star_{R,P})\right|^3 \le 4P_k(x)\left|\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P})}\right|^3 + 4P_k(x)B^3_{\alpha_k}(W_x\|\sigma^\star_{R,P}). \tag{11.143}$$

Further, Eq. (11.135) implies

$$\left|\log\frac{\lambda_i}{\mu_j(\sigma^\star_{R,P})}\right|^3 \le -4\log^3\tilde{\lambda}_{\min}(W_x) - 4\log^3 P_k^{\frac{1}{\tilde{\alpha}_{\min}}}(x). \tag{11.144}$$

Combining Eqs. (11.141), (11.143), and (11.144), proves the continuity of $T_\alpha(P,\mathcal{W})$. □

11.3.5 Proofs of Theorem 11.1 and Corollary 11.1

We are ready to prove our main result—the refined strong sphere-packing bound in Theorem 11.1 for constant composition codes and Corollary 11.1 for general codes.

doi:10.6342/NTU201800597

Proof of Theorem 11.1. Fix any rate $C_{0,\mathcal{W}} < R < C_{\mathcal{W}}$. First note that by Ref. [36, Proposition 10], we find

$$E_{\mathrm{sp}}(R) \in \mathbb{R}_{>0}. \tag{11.145}$$

By Proposition 11.2 and the standard expurgation method (see e.g. [30, p. 96], [32, Theorem 20], [23, p. 395]), it holds for every constant composition code $\mathcal{C}_n$ with a common composition $P_{\mathbf{x}^n}$ that

$$\begin{aligned}
\bar{\varepsilon}\left(\mathcal{C}_n\right) \geq \frac{1}{2}\varepsilon_{\max}\left(\mathcal{C}_n'\right) &\geq \max_{\sigma\in\mathcal{S}(\mathcal{H})} \frac{1}{2}\widehat{\alpha}_{1/|\mathcal{C}_n'|}\left(W_{\mathbf{x}^n}^{\otimes n}\|\sigma^{\otimes n}\right) && (11.146)\\
&\geq \max_{\sigma\in\mathcal{S}(\mathcal{H})} \frac{1}{2}\widehat{\alpha}_{2\exp\{-nR\}}\left(W_{\mathbf{x}^n}^{\otimes n}\|\sigma^{\otimes n}\right) && (11.147)\\
&\geq \frac{1}{2}\widehat{\alpha}_{2\exp\{-nR\}}\left(W_{\mathbf{x}^n}^{\otimes n}\|(\sigma^\star)^{\otimes n}\right), && (11.148)
\end{aligned}$$

where $\mathcal{C}_n'$ is an expurgated code with message size $|\mathcal{C}_n'| = \lceil |\mathcal{C}_n|/2 \rceil \geq \frac{1}{2}\exp\{nR\}$. Inequality (11.147) holds because the map $\mu \mapsto \widehat{\alpha}_\mu$ is monotone decreasing. In the last line (11.148) we denote by

$$\sigma^\star = \sigma^\star_{R,P_{\mathbf{x}^n}} := \operatorname*{arg\,min}_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{0<\alpha\leq 1}\left\{\frac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{W}\|\sigma|P_{\mathbf{x}^n}\right) - R\right)\right\} \tag{11.149}$$

a channel output state that depends on the coding rate R and the composition $P_{\mathbf{x}^n}$.

In the following, we deal with sequences of inputs that will yield different lower bounds. Fix an arbitrary $\delta \in (0, E_{\mathrm{sp}}(R))$. Let $\nu := E_{\mathrm{sp}}(R) - \delta > 0$, and define:

$$\mathcal{P}_{R,\nu}(\mathcal{X}) := \left\{P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X}) : \nu \leq E_{\mathrm{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \leq E_{\mathrm{sp}}(R) < +\infty\right\}. \tag{11.150}$$

The set $\mathcal{P}_{R,\nu}(\mathcal{X})$ ensures that the error exponents of the input sequences $\mathbf{x}^n$ with composition $P_{\mathbf{x}^n} \in \mathcal{P}_{R,\nu}(\mathcal{X})$ are close to the sphere-packing exponent $E_{\mathrm{sp}}(R)$.

For sequences $\mathbf{x}^n$ with $P_{\mathbf{x}^n} \notin \mathcal{P}_{R,\nu}(\mathcal{X})$, we infer that

$$E_{\mathrm{sp}}(R) - E_{\mathrm{sp}}^{(2)}\left(R, P_{\mathbf{x}^n}\right) = \delta > 0. \tag{11.151}$$

We then apply the Chebyshev-type bound, Proposition 11.4, with $c = 2$ to obtain, $\forall P_{\mathbf{x}^n} \notin \mathcal{P}_{R,\nu}(\mathcal{X})$,

$$\begin{aligned}
\widehat{\alpha}_{2\exp\{-nR\}}\left(W_{\mathbf{x}^n}^{\otimes n}\|(\sigma^\star)^{\otimes n}\right) &\geq \kappa_1 \exp\left\{-\kappa_2\sqrt{n} - nE_{\mathrm{sp}}^{(2)}\left(R, P_{\mathbf{x}^n}\right)\right\}, && (11.152)\\
&\geq \kappa_1 \exp\left\{-\kappa_2\sqrt{n} - n\left[E_{\mathrm{sp}}\left(R\right) - \delta\right]\right\}, && (11.153)
\end{aligned}$$

for all sufficiently large n, say $n \geq N_1 \in \mathbb{N}$. The equality in Eq. (11.152) follows from the saddle-point property, item (a) in Proposition 9.5, and the constants κ_1, κ_2 are positive and finite constants.

Next, we consider sequences $\mathbf{x}^n$ with $P_{\mathbf{x}^n} \in \mathcal{P}_{R,\nu}(\mathcal{X})$. Since such sequences satisfy Eq. (11.94), we apply the sharp lower bound, Proposition 11.5, with $c = 2$ to obtain, $\forall P_{\mathbf{x}^n} \in \mathcal{P}_{R,\nu}(\mathcal{X})$,

$$\widehat{\alpha}_{2\exp\{-nR\}}\left(W_{\mathbf{x}^n}^{\otimes n}\|(\sigma^\star)^{\otimes n}\right) \geq \frac{2A}{n^{\frac{1}{2}\left(1+s^\star_{R,P_{\mathbf{x}^n}}\right)}} \exp\left\{-nE_{\mathrm{sp}}^{(2)}\left(R, P_{\mathbf{x}^n}\right)\right\}, \tag{11.154}$$

for all sufficiently large n, say $n \geq N_2 \in \mathbb{N}$, and some $A \in \mathbb{R}_{>0}$. In the following, we will relate the

doi:10.6342/NTU201800597

term $s^{\star}_{R,P_{\mathbf{x}^n}}$ in Eq. (11.154) to $\left|E'_{\mathrm{sp}}(R)\right|$. The idea follows similar from [91, Eqs. (111)–(114)]. Let

$$\mathcal{P}^{\star}_{R}(\mathcal{X}) := \left\{P \in \mathcal{P}(\mathcal{X}) : E^{(2)}_{\mathrm{sp}}(R,P) = E_{\mathrm{sp}}(R)\right\}, \tag{11.155}$$

$$\mathcal{P}_{\theta}(\mathcal{X}) := \left\{P \in \mathcal{P}_{R,\nu}(\mathcal{X}) : \min_{Q\in\mathcal{P}^{\star}_{R}(\mathcal{X})} \|P-Q\|_1 \geq \theta\right\}. \tag{11.156}$$

Since $s^{\star}_{R,(\cdot)}$ is uniformly continuous on the compact set $P \in \mathcal{P}_{R,\nu}(\mathcal{X})$ (see item (d) of Proposition 9.6), one has

$$\forall \gamma \in \mathbb{R}_{>0},\ \exists f(\gamma) \in \mathbb{R}_{>0},\ \text{such that } \forall P,Q \in \mathcal{P}_{R,\nu}(\mathcal{X}),\ \|P-Q\|_1 < f(\gamma) \Rightarrow \left|s^{\star}_{R,P} - s^{\star}_{R,Q}\right| < \gamma. \tag{11.157}$$

By choosing $\gamma \in \mathbb{R}_{>0}$ that satisfies Eq. (11.157), it follows that

$$s^{\star}_{R,P_{\mathbf{x}^n}} \leq \left|E'_{\mathrm{sp}}(R)\right| + \gamma, \quad \forall P_{\mathbf{x}^n} \in \mathcal{P}_{R,\nu}(\mathcal{X}) \backslash \mathcal{P}_{f(\gamma)}(\mathcal{X}). \tag{11.158}$$

Hence, Eqs. (11.154) and (11.158) further lead to, $\forall P_{\mathbf{x}^n} \in \mathcal{P}_{R,\nu}(\mathcal{X}) \backslash \mathcal{P}_{f(\gamma)}(\mathcal{X})$,

$$\widehat{\alpha}_{2\exp\{-nR\}}\left(W^{\otimes n}_{\mathbf{x}^n} \| (\sigma^{\star})^{\otimes n}\right) \geq \frac{2A}{n^{\frac{1}{2}\left(1+\left|E'_{\mathrm{sp}}(R)\right|+\gamma\right)}} \exp\left\{-nE_{\mathrm{sp}}\left(R\right)\right\}. \tag{11.159}$$

For the case $P_{\mathbf{x}^n} \in \mathcal{P}_{R,\nu}(\mathcal{X}) \cap \mathcal{P}_{f(\gamma)}(\mathcal{X})$, we have

$$E_{\mathrm{sp}}(R) - \max_{P\in\mathcal{P}_{f(\gamma)}(\mathcal{X})} E^{(2)}_{\mathrm{sp}}(R,P_{\mathbf{x}^n}) =: \delta' > 0. \tag{11.160}$$

Then, Eqs. (11.154) and (11.160) give, $\forall P_{\mathbf{x}^n} \in \mathcal{P}_{R,\nu}(\mathcal{X}) \cap \mathcal{P}_{f(\gamma)}(\mathcal{X})$,

$$\widehat{\alpha}_{2\exp\{-nR\}}\left(W^{\otimes n}_{\mathbf{x}^n} \| (\sigma^{\star})^{\otimes n}\right) \geq \frac{2A}{n^{\frac{1}{2}\left(1+s^{\star}_{R,P_{\mathbf{x}^n}}\right)}} \exp\left\{-n\left[E_{\mathrm{sp}}\left(R\right) - \delta'\right]\right\}. \tag{11.161}$$

Finally, by comparing the bounds in Eqs. (11.153), (11.159) and (11.161), the first-order leading term in the right-hand side of Eq. (11.159) decays faster than that of Eqs. (11.153) and (11.161). Thus, for sufficiently large n, say $n \geq N_3 \in \mathbb{N}$, we combine the bounds to obtain, for all compositions $P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X})$,

$$\widehat{\alpha}_{2\exp\{-nR\}}\left(W^{\otimes n}_{\mathbf{x}^n} \| (\sigma^{\star})^{\otimes n}\right) \geq \frac{2A}{n^{\frac{1}{2}\left(1+\left|E'_{\mathrm{sp}}(R)\right|+\gamma\right)}} \exp\left\{-nE_{\mathrm{sp}}\left(R\right)\right\}. \tag{11.162}$$

By combining Eqs. (11.148), (11.162), we conclude our result: for any $\gamma > 0$ and every n-blocklength constant composition code $\mathcal{C}_n$,

$$\bar{\varepsilon}\left(\mathcal{C}_n\right) \geq \frac{A}{n^{\frac{1}{2}\left(1+\left|E'_{\mathrm{sp}}(R)\right|+\gamma\right)}} \exp\left\{-nE_{\mathrm{sp}}\left(R\right)\right\}, \tag{11.163}$$

for all sufficiently large $n \geq N_0 := \max\{N_1, N_2, N_3\}$. □

Proof of Corollary 11.1. For an n-blocklength code, there are at most $\binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1} < n^{|\mathcal{X}|}$ different compositions. Hence, for any code with $M = \exp\{nR\}$ codewords, there exists some codewords M' of

doi:10.6342/NTU201800597

the same composition such that $M' \geq M/n^{|\mathcal{X}|}$. Denote by $\mathcal{C}'_n$ such constant composition codes with composition $P_{\mathbf{x}^n}$.

Fix an arbitrary $R_0 \in (R_\infty, R)$, and choose N_1 be an integer such that $R - \frac{|\mathcal{X}|}{n}\log n \geq R_0$ for all $n \geq N_1$. Consider such $n \geq N_1$ onwards. By following the similar steps in Theorem 11.1, we obtain

$$\varepsilon^*(n,R) \geq \bar{\varepsilon}\left(\mathcal{C}'_n\right) \geq \frac{A}{n^{\frac{1}{2}\left(1+s^\star_{R,P^n_{\mathbf{x}}}\right)}} \exp\left\{-nE^{(2)}_{\mathrm{sp}}\left(R - \frac{|\mathcal{X}|}{n}\log n, P_{\mathbf{x}^n}\right)\right\}, \tag{11.164}$$

for all sufficiently large n, say $n \geq N_2 \in \mathbb{N}$, and some $s^\star_{R,P_{\mathbf{x}^n}} \in \mathbb{R}_{>0}$. Let

$$\Upsilon := \max_{P\in\mathcal{P}(\mathcal{X}):E^{(2)}_{\mathrm{sp}}(\bar{R},P)=E_{\mathrm{sp}}(\bar{R})} \left|\frac{\partial E^{(2)}_{\mathrm{sp}}(r,P)}{\partial r}\bigg|_{r=R_0}\right|. \tag{11.165}$$

Then, item (a) in Proposition 9.6 implies that

$$E^{(2)}_{\mathrm{sp}}\left(R - \frac{|\mathcal{X}|}{n}\log n, P_{\mathbf{x}^n}\right) \leq E^{(2)}_{\mathrm{sp}}(R, P_{\mathbf{x}^n}) + \Upsilon\cdot\frac{|\mathcal{X}|}{n}\log n \tag{11.166}$$

$$\leq E_{\mathrm{sp}}(R) + \Upsilon\cdot\frac{|\mathcal{X}|}{n}\log n, \quad \forall n \geq N_2 \tag{11.167}$$

Combining Eqs. (11.164) and (11.167) gives

$$\varepsilon^*(n,R) \geq \frac{A}{n^{\frac{1}{2}\left(1+s^\star_{R,P_{\mathbf{x}^n}}\right)+\Upsilon|\mathcal{X}|}} \exp\left\{-nE_{\mathrm{sp}}(R)\right\}, \quad \forall n \geq \max\{N_1,N_2\}. \tag{11.168}$$

By choosing $t \in \mathbb{R}_{>0}$ such that $n^{-t} \leq An^{-\frac{1}{2}\left(1+s^\star_{R,P_{\mathbf{x}^n}}\right)-\Upsilon|\mathcal{X}|}$, and letting $N_0 := \max\{N_1,N_2\}$, we conclude our claim. □

11.4 Symmetric Classical-Quantum Channels

In this section, we consider a symmetric c-q channels. By using the symmetric property of the channels, we show that the uniform distribution, denoted by $U_{\mathcal{X}}$, achieves the maximum of $E^{(1)}_{\mathrm{sp}}(R,\cdot)$ and $E^{(2)}_{\mathrm{sp}}(R,\cdot)$. Then, by choosing the optimal output state $\sigma^\star_R = \sigma^\star_{R,U_{\mathcal{X}}}$, every input sequence in the codebook is a good codeword and attains the sphere-packing exponent $E_{\mathrm{sp}}(R)$. Hence, we can remove the assumption of constant composition codes and apply Theorem 11.1 to obtain the exact pre-factor for the sphere-packing bound (Theorem 11.3).

A c-q channel $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$ is *symmetric* if it satisfies

$$W_x := V^{x-1}W_1(V^\dagger)^{x-1}, \quad \forall x \in \mathcal{X}, \tag{11.169}$$

where $W_1 \in \mathcal{S}(\mathcal{H})$ is an arbitrary density operator, and V satisfies $V^\dagger V = VV^\dagger = V^{|\mathcal{X}|} = \mathbb{1}_{\mathcal{H}}$.

doi:10.6342/NTU201800597

Theorem 11.3 (Exact Sphere-packing Bound for Symmetric Classical-Quantum Channels). *For any rate $R \in (R_\infty, C_{\mathrm{W}})$, there exist $A > 0$ and $N_0 \in \mathbb{N}$ such that for all codes $\mathcal{C}_n$ of length $n \geq N_0$ with message size $|\mathcal{C}_n| \geq \exp\{nR\}$, we have*

$$\varepsilon_{\max}(\mathcal{C}_n) \geq \frac{A}{n^{\frac{1}{2}\left(1+\left|E'_{\mathrm{sp}}(R)\right|\right)}} \exp\left\{-nE_{\mathrm{sp}}(R)\right\}. \tag{11.170}$$

Proof of Theorem 11.3. The proof consists of the following steps. First, we show that the distribution $U_{\mathcal{X}}$ satisfies $E_{\mathrm{sp}}^{(1)}(R, U_{\mathcal{X}}) = E_{\mathrm{sp}}^{(2)}(R, U_{\mathcal{X}}) = E_{\mathrm{sp}}(R)$. Second, we show that $E_{\mathrm{sp}}^{(2)}(R, P) = E_{\mathrm{sp}}(R)$ for all $P \in \mathcal{P}(\mathcal{X})$, which means that any codeword attains the sphere-packing exponent. Finally, we follow Theorem 11.1 to complete the proof.

Fix any $R \in (C_{0,\mathrm{W}}, C_{\mathrm{W}})$. From the definition of the symmetric channels in Eq. (11.169), it is not hard to verify that $U_{\mathcal{X}}\mathcal{W}^\alpha = VU_{\mathcal{X}}\mathcal{W}^\alpha V^\dagger$ for all $\alpha \in (0,1]$, where we denote by $P\mathcal{W}^\alpha := \sum_{x\in\mathcal{X}} P(x)W_x^\alpha$ for all $\alpha \in (0,1]$. Hence, it follows that

$$\begin{aligned}
\mathrm{Tr}[W_x^\alpha (U_{\mathcal{X}}\mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] &= \mathrm{Tr}[V^{x-1}W_1^\alpha V^{\dagger\, x-1}(U_{\mathcal{X}}\mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] && (11.171)\\
&= \mathrm{Tr}[W_1^\alpha (U_{\mathcal{X}}\mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] && (11.172)
\end{aligned}$$

for all $x \in \mathcal{X}$ and $\alpha \in (0,1]$. Summing Eq. (11.172) over all $x \in \mathcal{X}$ and dividing by M yields

$$\mathrm{Tr}[W_x^\alpha (U_{\mathcal{X}}\mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] = \mathrm{Tr}[(U_{\mathcal{X}}\mathcal{W}^\alpha)^{\frac{1}{\alpha}}], \tag{11.173}$$

for all $x \in \mathcal{X}$ and $\alpha \in (0,1]$. Recalling Proposition 11.7 below, the above equation shows that the distribution $U_{\mathcal{X}}$ indeed maximizes $E_0(s,P)$, $\forall s \in \mathbb{R}_{\geq 0}$. Then we have

$$E_{\mathrm{sp}}^{(1)}(R, U_{\mathcal{X}}) = \sup_{s\geq 0}\left\{\max_{P\in\mathcal{P}(\mathcal{X})} E_0(s,P) - sR\right\} = E_{\mathrm{sp}}(R).$$

Further, Jensen's inequality shows that $E_{\mathrm{sp}}^{(2)}(R, U_{\mathcal{X}}) \geq E_{\mathrm{sp}}^{(1)}(R, U_{\mathcal{X}}) = E_{\mathrm{sp}}(R)$, and thus, $E_{\mathrm{sp}}^{(2)}(R, U_{\mathcal{X}}) = E_{\mathrm{sp}}(R)$.

Next, let $(\alpha_R^\star, \sigma_R^\star)$ be the saddle-point of $F_{R,U_{\mathcal{X}}}(\cdot,\cdot)$ (see Eq. (9.142)). One can observe from the definition of $E_{\mathrm{sp}}^{(2)}$ and Eq. (11.173) that all the quantities $D_{\alpha_R^\star}(W_x\|\sigma_R^\star)$, $x \in \mathcal{X}$, are equal. Hence, quantum Sibson's identity given in Lemma 3.3 shows that

$$\sigma_R^\star = \frac{\left(U_{\mathcal{X}}\mathcal{W}^{\alpha_R^\star}\right)^{1/\alpha_R^\star}}{\mathrm{Tr}\left[\left(U_{\mathcal{X}}\mathcal{W}^{\alpha_R^\star}\right)^{1/\alpha_R^\star}\right]}, \tag{11.174}$$

which, in turn, implies that

$$E_{\mathrm{sp}}^{(2)}(R,P) = \sup_{\alpha\in(0,1]} F_{R,P}(\alpha, \sigma_R^\star) = \sup_{s\geq 0}\{E_0(s, U_{\mathcal{X}}) - sR\} = E_{\mathrm{sp}}(R), \quad \forall P \in \mathcal{P}(\mathcal{X}). \tag{11.175}$$

Further, we have

$$\left|E'_{\mathrm{sp}}(R)\right| = \frac{1-\alpha_R^\star}{\alpha_R^\star} = \left|\frac{\partial E_{\mathrm{sp}}^{(2)}(R,P)}{\partial R}\right|, \quad \forall P \in \mathcal{P}(\mathcal{X}). \tag{11.176}$$

doi:10.6342/NTU201800597

Since Eqs. (11.175) and (11.176) indicates that every input sequence attains the sphere-packing exponent, we apply the same arguments in the proof of Theorem 11.1 to conclude this theorem.

Proposition 11.7 ([35, Eq. (38)]). *Let $s \in \mathbb{R}_{\geq 0}$ be arbitrary. The Necessary and sufficient condition for the distribution $P^\star$ to maximize $E_0(s, P)$ is*

$$\mathrm{Tr}\left[W_x^{1/(1+s)} \cdot \left(\sum_{x\in\mathcal{X}} P^\star(x) W_x^{1/(1+s)}\right)^s\right] \geq \mathrm{Tr}\left[\left(\sum_{x\in\mathcal{X}} P^\star(x) W_x^{1/(1+s)}\right)^{1+s}\right], \ \forall x \in \mathcal{X} \tag{11.177}$$

with equality if $P^\star(x) \neq 0$.

□

doi:10.6342/NTU201800597

Chapter 12

Moderate Deviation Analysis (Channel Coding)

This section presents our main results—the error performance of classical-quantum channels satisfies the moderate deviation property, Eq. (1.6). The achievability part is stated in Theorem 12.1, and its proof is given in Section 12.1. Our proof strategy employs Hayashi's bound [88] and the properties of the modified auxiliary function (Proposition 9.2). Theorem 12.2 contains the converse part, and is proved in Section 12.2. The proof involves a weak sphere-packing bound (Theorem 11.2), a sharp converse lower bound (Theorem 11.1), and an approximation of the error-exponent function around capacity (Proposition 12.2).

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers satisfying

$$\begin{aligned} &\text{(i) } a_n \to 0, \quad \text{as} \quad n \to +\infty, \\ &\text{(ii) } a_n\sqrt{n} \to +\infty, \quad \text{as} \quad n \to +\infty. \end{aligned} \tag{12.1}$$

Unlike our proof techniques relying on error exponent analysis (the LDP regime), a recent and independent paper [146] obtained the same result, but proceeds from the second-order analysis (the CLT regime). Their achievability proof follows from the one-shot capacity by Hayashi and Nagaoka [87] (see also Hayashi [88], and Wang and Renner [147]); while the converse part reduces channel coding to hypothesis testing [148, 87, 44], followed by Strassen's Gaussian approximation [11] and a powerful inequality in probability [149] to the quantum scenario.

Theorem 12.1 (Achievability). *For any $\mathcal{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ with $V_{\mathcal{W}} > 0$ and any sequence $(a_n)_{n\geq 1}$ satisfying Eq. (12.1), there exists a sequence of codes $\{\mathcal{C}_n\}_{n\geq 1}$ with rates $R_n = C_{\mathcal{W}} - a_n$ so that*

$$\limsup_{n\to+\infty} \frac{1}{na_n^2} \log \bar{\varepsilon}\left(\mathcal{W}, \mathcal{C}_n\right) \leq -\frac{1}{2V_{\mathcal{W}}}. \tag{12.2}$$

The proof is given in Section 12.1.

Theorem 12.2 (Converse). *For any $W \in \mathcal{W}(\mathcal{X})$ with $V(W) > 0$, any sequence $\{a_n\}_{n\geq 1}$ satisfying Eq. (12.1), and any sequence of codes $\{\mathcal{C}_n\}_{n\geq 1}$ with rates $R_n = C(W) - a_n$, it holds that*

$$\liminf_{n\to+\infty} \frac{1}{na_n^2} \log \bar{\varepsilon}\left(W, \mathcal{C}_n\right) \geq -\frac{1}{2V(W)}. \tag{12.3}$$

doi:10.6342/NTU201800597

The proof is given in Section 12.2.

Remark 12.1. Altuğ and Wagner [43] proved Theorem 12.2 for discrete classical channels by a weak sphere-packing bound with the expression of $\widetilde{E}_{\mathrm{sp}}$. Although such a weak sphere-packing bound indeed holds for c-q channels (as we have shown in Theorem 11.2 and Remark 11.1 in Section 11.2), Proposition 12.2 in Section 12.2 shows that it will lead to

$$\limsup_{n\to+\infty} \frac{1}{na_n^2} \log \bar{\epsilon}\left(\mathcal{W}, \mathcal{C}_n\right) \geq -\frac{1}{2\widetilde{V}_{\mathcal{W}}}, \tag{12.4}$$

where $\widetilde{V}_{\mathcal{W}}$ is defined in Eq. (3.58). Since $\widetilde{V}(\rho\|\sigma) \leq V(\rho\|\sigma)$ [150, Theorem 1.2], it holds that $\widetilde{V}_{\mathcal{W}} \leq V_{\mathcal{W}}$ and the equality happens if and only if the channel reduces to classical. Hence, Altuğ and Wagner's method yields a weaker result in quantum regime; namely, a gap between the achievability and the converse. In Section 12.2, we will employ a sharp converse bound from strong large deviation theory to achieve our result, Theorem 12.2.

12.1 Proof of Achievability, Theorem 12.1

Let $\mathcal{W}: \mathcal{X} \to \mathcal{S}(\mathcal{H})$ satisfy $V_{\mathcal{W}} > 0$. Let $\{a_n\}_{n\geq 1}$ be any sequence of real numbers satisfying Eq. (12.1). Since $V_{\mathcal{W}} > 0$, Eq. (3.59) in Section 3.3 shows that $C_{\mathcal{W}} > 0$. Hence, we have $C_{\mathcal{W}} - a_n > 0$, for all sufficiently large n. Fix such an integer n onwards. The achievability bound, Theorem 10.1, in Chapter 10 implies that there exists a code $\mathcal{C}_n$ with $R_n = C_{\mathcal{W}} - a_n$ so that

$$\bar{\varepsilon}(\mathcal{W}, \mathcal{C}_n) \leq 6 \exp\left\{-n\left[\max_{0\leq s\leq 1}\left\{E_0^{\downarrow}(s, P, P\mathcal{W}) - sR_n\right\}\right]\right\}, \tag{12.5}$$

for all $P \in \mathcal{P}(\mathcal{X})$. In the following, we denote by $E_0^{\downarrow}(s, P) := E_0^{\downarrow}(s, P, P\mathcal{W})$ for notational convenience. Simple algebra yields

$$\frac{1}{na_n^2} \log \bar{\varepsilon}(\mathcal{W}, \mathcal{C}_n) \leq \frac{\log 6}{na_n^2} - \frac{1}{a_n^2} \max_{0\leq s\leq 1}\left\{E_0^{\downarrow}(s, P) - sR_n\right\}, \tag{12.6}$$

for all sufficiently large n and any $P \in \mathcal{P}(\mathcal{X})$.

Let $\widetilde{\mathcal{P}}(\mathcal{X})$ be the set of distributions that achieve the minimum in Eq. (3.57), and let $\widetilde{P} \in \widetilde{\mathcal{P}}(\mathcal{X})$. Note that Ref. [16, Lemma 3] implies that $\widetilde{\mathcal{P}}(\mathcal{X})$ is compact. Applying Taylor's theorem to $E_0^{\downarrow}(s, \widetilde{P})$ at $s = 0$ together with Proposition 9.2 gives

$$E_0^{\downarrow}\left(s, \widetilde{P}\right) = sC_{\mathcal{W}} - \frac{s^2}{2}V_{\mathcal{W}} + \frac{s^3}{6}\frac{\partial^3 E_0^{\downarrow}\left(s, \widetilde{P}\right)}{\partial s^3}\bigg|_{s=\bar{s}}, \tag{12.7}$$

for some $\bar{s} \in [0, s]$. Let $s_n = a_n/V_{\mathcal{W}}$. Then $s_n \leq 1$ for all sufficiently large n by the assumption in

doi:10.6342/NTU201800597

Eq. (12.1) and $V_{\mathcal{W}} > 0$. For all $s_n \leq 1$, Eq. (12.7) yields

$$\max_{0\leq s\leq 1}\left\{E_0^{\downarrow}\left(s,\widetilde{P}\right)-sR_n\right\} \geq E_0^{\downarrow}\left(s_n,\widetilde{P}\right)-s_nR_n \tag{12.8}$$

$$= \frac{a_n}{V_{\mathcal{W}}}\left(C_{\mathcal{W}}-R_n\right)-\frac{a_n^2}{2V_{\mathcal{W}}}+\frac{a_n^3}{6V_{\mathcal{W}}^3}\left.\frac{\partial^3 E_0^{\downarrow}\left(s,\widetilde{P}\right)}{\partial s^3}\right|_{s=\bar{s}_n} \tag{12.9}$$

$$= \frac{a_n^2}{2V_{\mathcal{W}}}+\frac{a_n^3}{6V_{\mathcal{W}}^3}\left.\frac{\partial^3 E_0^{\downarrow}\left(s,\widetilde{P}\right)}{\partial s^3}\right|_{s=\bar{s}_n}, \tag{12.10}$$

where $\bar{s}_n \in [0, s_n]$ and Eq. (12.10) holds since $R_n = C_{\mathcal{W}} - a_n$.

Define

$$\Upsilon = \max_{(s,P)\in[0,1]\times\widetilde{\mathcal{P}}(\mathcal{X})}\left|\frac{\partial^3 E_0^{\downarrow}(s,P)}{\partial s^3}\right|, \tag{12.11}$$

which is finite due to the compact set $[0,1]\times\widetilde{\mathcal{P}}(\mathcal{X})$ and item (a) in Proposition 9.2. Therefore, Eq. (12.10) implies that

$$\max_{0\leq s\leq 1}\left\{E_0^{\downarrow}\left(s,\widetilde{P}\right)-sR_n\right\} \geq \frac{a_n^2}{2V_{\mathcal{W}}}+\frac{a_n^3}{6V_{\mathcal{W}}^3}\left.\frac{\partial^3 E_0^{\downarrow}\left(s,\widetilde{P}\right)}{\partial s^3}\right|_{s=\bar{s}_n} \tag{12.12}$$

$$\geq \frac{a_n^2}{2V_{\mathcal{W}}}-\frac{a_n^3\left|\left.\frac{\partial^3 E_0^{\downarrow}\left(s,\widetilde{P}\right)}{\partial s^3}\right|_{s=\bar{s}_n}\right|}{6V_{\mathcal{W}}^3} \tag{12.13}$$

$$\geq \frac{a_n^2}{2V_{\mathcal{W}}}-\frac{a_n^3\Upsilon}{6V_{\mathcal{W}}^3}, \tag{12.14}$$

for all sufficiently large n.

Substituting Eq. (12.14) into Eq. (12.6) gives

$$\frac{1}{na_n^2}\log\bar{\varepsilon}(\mathcal{W},\mathcal{C}_n) \leq \frac{\log 4}{na_n^2}-\frac{1}{2V_{\mathcal{W}}}\left(1-\frac{a_n\Upsilon}{3V_{\mathcal{W}}^2}\right). \tag{12.15}$$

Recall Eq. (12.1) and let $n\to+\infty$, which completes the proof:

$$\limsup_{n\to+\infty}\frac{1}{na_n^2}\log\bar{\varepsilon}(\mathcal{W},\mathcal{C}_n) \leq -\frac{1}{2V_{\mathcal{W}}}. \tag{12.16}$$

□

12.2 Proof of Converse, Theorem 12.2

Our strategy consists of the following steps. First, we claim that it suffices to prove Eq. (12.3) for the maximal error probability of any code $\mathcal{C}_n$, i.e. $\varepsilon_{\max}(\mathcal{W},\mathcal{C}_n)$. Recall the standard expurgation method (see e.g. [30, p. 96], [32, Theorem 20], [23, p. 395]): by removing half codewords with highest error probability to arrive at $\bar{\varepsilon}(\mathcal{W},\mathcal{C}_n) \geq \frac{1}{2}\varepsilon_{\max}(\mathcal{W},\mathcal{C}_n')$ with $|\mathcal{C}_n'| = \lceil|\mathcal{C}_n|/2\rceil \geq \frac{1}{2}\exp\{nR_n\} = \exp\{n(R_n -$

doi:10.6342/NTU201800597

$\frac{1}{n}\log 2)\}$. Since the induced rate back-off is only $\frac{1}{n}\log 2 = o(a_n)$, one might define another sequence $a'_n := a_n - \frac{1}{n}\log 2$ satisfying Eq. (12.1). Hence, without of loss generality, we only need to prove the converse part for $\varepsilon_{\max}$.

Second, we employ the method of Ref. [26, Lemma 16] to relate the error probability $\varepsilon_{\max}$ to the minimum type-I error:

$$\frac{\log \varepsilon_{\max}(\mathcal{W}, \mathcal{C}_n)}{na_n^2} \geq \max_{\sigma^n \in \mathcal{S}(\mathcal{H}^{\otimes n})} \min_{\mathbf{x}^n \in \mathcal{X}^n} \frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}(W_{\mathbf{x}^n}^{\otimes n} \| \sigma^n)}{na_n^2} \tag{12.17}$$

$$\geq \min_{\mathbf{x}^n \in \mathcal{X}^n} \frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n})}{na_n^2}, \tag{12.18}$$

where $P^\star \in \mathcal{P}(\mathcal{X})$ is an arbitrary capacity-achieving distribution, i.e. $I(P^\star, \mathcal{W}) = C_{\mathcal{W}}$.

Third, we divide the set of codewords into two groups. Fix an arbitrary $\eta \in (0, \frac{1}{2})$. Let $A := \max_{\rho \in \mathcal{S}_\circ} V(\rho \| P^\star \mathcal{W})$ and let $\xi = \sqrt{2A/\eta}$. Define:

$$\Omega_{\text{good}} := \{\mathbf{x}^n \in \mathcal{X}^n : D(\mathcal{W} \| P^\star \mathcal{W} | P_{\mathbf{x}^n}) > R_n\}; \tag{12.19}$$

$$\Omega_{\text{bad}} := \mathcal{X}^n \backslash \Omega_{\text{good}}. \tag{12.20}$$

For the codes in Ω_{bad}, we employ a weak converse bound in Theorem 11.2, and apply a sharp converse bound, Proposition 12.1 below, for Ω_{good}. Furthermore, we can assume $a_n > 0$ for all sufficiently large $n \in \mathbb{N}$ owing to the assumption $\lim_{n \to +\infty} a_n \sqrt{n} = +\infty$. Subsequently, we will consider such n onwards. We remark that Proposition 12.1 follows the same argument as Proposition 11.5 in Section 11.3.3, and Chaganty-Sethuraman's concentration inequality, Theorem 2.2 in Section 2.2. Thus, we skip the proof.

Proof of Theorem 12.2. We start the proof with the case Ω_{bad}, and further consider two different cases:

$$\Omega_{\text{bad}}^{(1)} := \left\{\mathbf{x}^n \in \mathcal{X}^n : D(\mathcal{W} \| P^\star \mathcal{W} | P_{\mathbf{x}^n}) \leq R_n - \frac{2\xi}{\sqrt{n}}\right\}; \tag{12.21}$$

$$\Omega_{\text{bad}}^{(2)} := \left\{\mathbf{x}^n \in \mathcal{X}^n : R_n - \frac{2\xi}{\sqrt{n}} < D(\mathcal{W} \| P^\star \mathcal{W} | P_{\mathbf{x}^n}) \leq R_n\right\}. \tag{12.22}$$

We apply the weak converse bound, Theorem 11.2, in Section 11.2 with $\sigma = P^\star \mathcal{W}$ to further lower bound the right-hand side of Eq. (12.18).

Let η and ξ be defined as above, and let N_1 be an integer satisfying Eq. (11.13). Then Eq. (11.14) gives, for all $n \geq N_1$,

$$\frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n})}{na_n^2} \geq -\frac{\widetilde{E}_{\text{sp}}\left(R_n - \frac{2\xi}{\sqrt{n}}, P_{\mathbf{x}^n}, P^\star \mathcal{W}\right)}{a_n^2(1-\eta)} + \frac{\log f(\eta)}{na_n^2}. \tag{12.23}$$

Further, Eq. (9.18) implies that for all $\mathbf{x}^n \in \Omega_{\text{bad}}^{(1)}$,

$$\widetilde{E}_{\text{sp}}\left(R_n - \frac{2\xi}{\sqrt{n}}, P_{\mathbf{x}^n}, P^\star \mathcal{W}\right) = 0. \tag{12.24}$$

doi:10.6342/NTU201800597

Hence, we have for all $\mathbf{x}^n \in \Omega_{\text{bad}}^{(1)}$,

$$\frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n})}{na_n^2} \geq \frac{\log f(\eta)}{na_n^2} \tag{12.25}$$

$$\geq -\frac{1}{2V_{\mathcal{W}}} + \frac{\log f(\eta)}{na_n^2}, \tag{12.26}$$

where the last inequality follows from $V_{\mathcal{W}} > 0$. Since $f(\eta) < +\infty$, taking the infimum limit of $n \to +\infty$ and using Eq. (12.1) give, for all $\mathbf{x}^n \in \Omega_{\text{bad}}^{(1)}$,

$$\liminf_{n\to+\infty} \frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n}\right)}{na_n^2} \geq -\frac{1}{2V_{\mathcal{W}}}. \tag{12.27}$$

Next, we move on to $\mathbf{x}^n \in \Omega_{\text{bad}}^{(2)}$. In this case, $\widetilde{E}_{\text{sp}}$ in Eq. (12.23) is not equal to zero for any finite n, we employ Eq. (12.45) in Proposition 12.2 below with $\delta_n = a_n + 2\xi/\sqrt{n}$ and $b_n = a_n$ to arrive at

$$\liminf_{n\to+\infty} \frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n}\right)}{na_n^2} \geq -\lim_{n\to+\infty} \frac{4\xi^2}{n\left(a_n + \frac{2\xi}{\sqrt{n}}\right)^2} \cdot \frac{1}{2\widetilde{V}_{\mathcal{W}}(1-\eta)} \tag{12.28}$$

$$= 0 \tag{12.29}$$

$$\geq -\frac{1}{2V_{\mathcal{W}}}, \tag{12.30}$$

where the equality follows since $\lim_{n\to+\infty} na_n^2 = +\infty$.

In the last case of $\mathbf{x}^n \in \Omega_{\text{good}}$, we employ a tighter bound, Proposition 12.1, to lower bound the right-hand side of Eq. (12.18).

Proposition 12.1 (A Sharp Converse Bound). *Consider a classical-quantum channel $\mathcal{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ and a state $\sigma \in \mathcal{S}(\mathcal{H})$. Suppose the sequence $\mathbf{x}^n \in \mathcal{X}^n$ satisfies*

$$\nu \leq V\left(\mathcal{W} \| \sigma | P_{\mathbf{x}^n}\right) < +\infty \tag{12.31}$$

for some $\nu > 0$, and suppose the sequence of rates $(R_n)_{n\in\mathbb{N}}$ satisfies[a] $D_0(\mathcal{W}\|\sigma|P_{\mathbf{x}^n}) < R_n < D(\mathcal{W}\|\sigma|P_{\mathbf{x}^n})$. Then, there exists an $N_0 \in \mathbb{N}$ such that, for all $n \geq N_0$,

$$\widehat{\alpha}_{\exp\{-nR_n\}}(W_{\mathbf{x}^n}^{\otimes n} \| \sigma^{\otimes n}) \geq \frac{A}{s_n^\star \sqrt{n}} \exp\left\{-nE_{\text{sp}}^{(2)}\left(R_n - c_n, P_{\mathbf{x}^n}, \sigma\right)\right\}, \tag{12.32}$$

where $c_n = \frac{K \log n}{n}$ and $A, K > 0$ are finite constants independent of the sequence $\mathbf{x}^n$, and

$$s_n^\star := \arg\max_{s\geq 0} \left\{E_{\text{h}}(s, P_{\mathbf{x}^n}, \sigma) - sR_n\right\}. \tag{12.33}$$

[a] Note that $D_0(\mathcal{W}\|\sigma|P) = D(\mathcal{W}\|\sigma|P)$ implies $W_x = \sigma$ for all $x \in \mathsf{supp}(P)$ [8, Collorary 4.1]. This further gives $V(\mathcal{W}\|\sigma|P) = 0$. However, the assumption in Eq. (12.31) ensures that $\liminf_{n\in\mathbb{N}} D(\mathcal{W}\|\sigma|P_{\mathbf{x}^n}) - D_0(\mathcal{W}\|\sigma|P_{\mathbf{x}^n}) > 0$. Hence, the intervals $[D_0(\mathcal{W}\|\sigma|P_{\mathbf{x}^n}), D(\mathcal{W}\|\sigma|P_{\mathbf{x}^n})]$ for all $\mathbf{x}^n$ satisfying Eq. (12.31) are not measure zero.

Before applying Proposition 12.1, we verify that the condition, Eq. (12.31), is satisfied. Define

$$v(\delta) := \min_{P\in\mathcal{P}(\mathcal{X})} \left\{V\left(\mathcal{W}\|P^\star\mathcal{W}|P\right) : D(\mathcal{W}\|P^\star\mathcal{W}|P) \geq C_{\mathcal{W}} - \delta\right\}. \tag{12.34}$$

doi:10.6342/NTU201800597

Note that the map $\delta \mapsto v(\delta)$ is monotone decreasing and continuous at 0 from above, i.e. $\lim_{\delta \downarrow 0} v(\delta) = v(0) = V_{\mathcal{W}}$ [16, Lemma 22]. For any $\kappa \in (0,1)$, we can choose a sufficiently small $\gamma > 0$ independent of the sequence $\mathbf{x}^n$ such that $v(\gamma) \geq (1-\kappa)V_{\mathcal{W}} =: \nu > 0$. Further, let $N_2 \in \mathbb{N}$ such that $a_n \leq \gamma$ for all $n \geq N_2$. Then, one finds, for all $\mathbf{x}^n \in \Omega_{\text{good}}$ and $n \geq N_2$,

$$V\left(\mathcal{W} \| P^\star \mathcal{W} | P_{\mathbf{x}^n}\right) \geq v(\gamma) \geq \nu > 0. \tag{12.35}$$

Moreover, since $V_{\mathcal{W}} > 0$ implies that $C_{\mathcal{W}} = \max_{P \in \mathcal{P}(\mathcal{X})} D(\mathcal{W} \| P^\star \mathcal{W} | P) > \max_{P \in \mathcal{P}} D_0(\mathcal{W} \| P^\star \mathcal{W} | P)$, one can choose a sufficiently large n, say $N_3 \in \mathbb{N}$, such that $R_n > D_0(\mathcal{W} \| P^\star \mathcal{W} | P_{\mathbf{x}^n})$ for all $n \geq N_3$. Now, we have for all $\mathbf{x}^n \in \Omega_{\text{good}}$ and $n \geq \max\{N_2, N_3\}$ that

$$\max_{P \in \mathcal{P}(\mathcal{X})} D_0(\mathcal{W} \| P^\star \mathcal{W} | P) < R_n < D(\mathcal{W} \| P^\star \mathcal{W} | P_{\mathbf{x}^n}); \tag{12.36}$$

$$0 < \nu \leq V(\mathcal{W} \| P^\star \mathcal{W} | P_{\mathbf{x}^n}). \tag{12.37}$$

Together with Eqs. (12.18) and (12.35) and letting $\sigma = P^\star \mathcal{W}$, Proposition 12.1 yields, for all $\mathbf{x}^n \in \Omega_{\text{good}}$ and all sufficiently large n, say $n \geq N_4 \in \mathbb{N}$,

$$\frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n}\right)}{n a_n^2} \geq -\frac{E_{\text{sp}}^{(2)}\left(R_n - c_n, P_{\mathbf{x}^n}, P^\star \mathcal{W}\right)}{a_n^2} - \frac{\log s_n^\star \sqrt{n}}{n a_n^2} + \frac{\log A}{n a_n^2}. \tag{12.38}$$

Recall Eq. (12.46) in Proposition 12.2 below with $b_n = 0$ and $\delta_n = a_n + c_n$ that $\limsup_{n \to +\infty} \frac{s_n^\star}{a_n + c_n} \leq \frac{1}{V_{\mathcal{W}}}$. Hence, one can fix an arbitrary $\zeta > 0$ and there exists an $N_5 \in \mathbb{N}$ such that $\frac{s_n^\star \sqrt{n}}{(a_n + c_n)\sqrt{n}} \leq \frac{1}{V_{\mathcal{W}}} + \zeta$ for all $n \geq N_5$. This then leads to for all sufficiently large $n \geq \max\{N_2, N_3, N_4, N_5\}$ and all $\mathbf{x}^n \in \Omega_{\text{good}}$,

$$\frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n}\right)}{n a_n^2} \geq -\frac{E_{\text{sp}}^{(2)}\left(R_n - c_n, P_{\mathbf{x}^n}, P^\star \mathcal{W}\right)}{a_n^2} - \frac{\log(a_n + c_n)\sqrt{n}}{n a_n^2} + \frac{\log \frac{A}{\frac{1}{V_{\mathcal{W}}} + \zeta}}{n a_n^2}. \tag{12.39}$$

Taking $n \to +\infty$, the second and the third terms on the right-hand side of Eq. (12.39) vanish since $c_n = K \frac{\log n}{n} = o(a_n)$ and the assumption $\lim_{n \to +\infty} a_n \sqrt{n} = +\infty$.

Next, we apply Eq. (12.44) in Proposition 12.2 again to bound the error-exponent function $E_{\text{sp}}^{(2)}$ in Eq. (12.38): for all $\mathbf{x}^n \in \Omega^{(3)}$

$$\liminf_{n \to +\infty} \frac{\log \widehat{\alpha}_{\exp\{-nR_n\}}\left(W_{\mathbf{x}^n}^{\otimes n} \| (P^\star \mathcal{W})^{\otimes n}\right)}{n a_n^2} \geq -\limsup_{n \to +\infty} \frac{E_{\text{sp}}^{(2)}\left(C_{\mathcal{W}} - \delta_n, P_{\mathbf{x}^n}, P^\star \mathcal{W}\right)}{a_n^2} \tag{12.40}$$

$$= -\limsup_{n \to +\infty} \frac{E_{\text{sp}}^{(2)}\left(C_{\mathcal{W}} - \delta_n, P_{\mathbf{x}^n}, P^\star \mathcal{W}\right)}{\delta_n^2} \tag{12.41}$$

$$\geq -\frac{1}{2V_{\mathcal{W}}}. \tag{12.42}$$

Finally, combining Eqs. (12.18), (12.27), (12.30) and (12.42) concludes the desired Eq. (12.3).

doi:10.6342/NTU201800597

Proposition 12.2 (Error Exponent around Capacity). *Let $(b_n)_{n\in\mathbb{N}}$ be a sequence of real numbers with $\lim_{n\to+\infty} b_n = 0$ and let $(\delta_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers with $\lim_{n\to+\infty}\delta_n = 0$. Suppose the sequence of distributions $(P_n)_{n\in\mathbb{N}}$ satisfies*

$$C_{\mathcal{W}} - \delta_n < D(\mathcal{W}\|P^\star\mathcal{W}|P_n) \le C_{\mathcal{W}} - b_n. \tag{12.43}$$

The following hold:

$$\limsup_{n\to+\infty} \frac{E_{\mathrm{sp}}^{(2)}\left(C_{\mathcal{W}} - \delta_n, P_n, P^\star\mathcal{W}\right)}{\delta_n^2} \le \limsup_{n\to+\infty} \frac{(\delta_n - b_n)^2}{2V_{\mathcal{W}}\delta_n^2}; \tag{12.44}$$

$$\limsup_{n\to+\infty} \frac{\widetilde{E}_{\mathrm{sp}}\left(C_{\mathcal{W}} - \delta_n, P_n, P^\star\mathcal{W}\right)}{\delta_n^2} \le \limsup_{n\to+\infty} \frac{(\delta_n - b_n)^2}{2\widetilde{V}_{\mathcal{W}}\delta_n^2}; \tag{12.45}$$

$$\limsup_{n\to+\infty} \frac{s_n^\star}{\delta_n} \le \frac{1}{V_{\mathcal{W}}}, \tag{12.46}$$

where

$$s_n^\star := \arg\max_{s\ge 0} \left\{E_{\mathrm{h}}(s, P_n, P^\star\mathcal{W}) - s\left(C_{\mathcal{W}} - \delta_n\right)\right\}. \tag{12.47}$$

The proof of Proposition 12.2 is provided in Section 12.3 below. □

12.3 Asymptotic Expansions of Error-Exponent around Capacity

Proof of Proposition 12.2. We only prove Eqs. (12.44) and (12.46), since Eq. (12.45) follows from the same argument and Proposition 9.4.

Recall the error-exponent function $E_{\mathrm{sp}}^{(2)}$:

$$E_{\mathrm{sp}}^{(2)}\left(C_{\mathcal{W}} - \delta_n, P, P^\star\mathcal{W}\right) = \sup_{s\ge 0}\left\{-s\left(C_{\mathcal{W}} - \delta_n\right) + E_{\mathrm{h}}(s, P, P^\star\mathcal{W})\right\}. \tag{12.48}$$

In the following, we fix $\sigma = P^\star\mathcal{W}$ in the definition of E_{h} (Eq. (9.7)) and denote by

$$E_{\mathrm{h}}(s, P) := E_{\mathrm{h}}(s, P, P^\star\mathcal{W}) = sD_{\frac{1}{1+s}}\left(\mathcal{W}\|P^\star\mathcal{W}|P\right). \tag{12.49}$$

for notational convenience. We define a *critical rate* for a c-q channel $\mathcal{W}$ to be

$$r_{\mathrm{cr}} := \max_{P\in\mathcal{P}(\mathcal{X})} \left.\frac{\partial E_{\mathrm{h}}(s, P)}{\partial s}\right|_{s=1}. \tag{12.50}$$

Let N_0 be the smallest integer such that $C_{\mathcal{W}} - \delta_n > r_{\mathrm{cr}}$, $\forall n \ge N_0$. Since the map $r \mapsto E_{\mathrm{sp}}^{(2)}(r,\cdot,\cdot)$ is non-increasing [86, Section 5], the maximization over s in Eq. (12.48) can be restricted to the set $[0,1]$ for any rate above r_{cr}, i.e.,

$$E_{\mathrm{sp}}^{(2)}\left(C_{\mathcal{W}} - \delta_n, P_n, P^\star\mathcal{W}\right) = \max_{0\le s\le 1}\left\{-s\left(C_{\mathcal{W}} - \delta_n\right) + E_{\mathrm{h}}(s, P_n)\right\}. \tag{12.51}$$

For every $n \in \mathbb{N}$, let $s_n^\star$ attain the maxima in Eq. (12.51) at a rate of $C_{\mathcal{W}} - \delta_n \ge 0$. In the following lemma, we discuss the asymptotic behavior of $\{s_n^\star\}_{n\in\mathbb{N}}$.

doi:10.6342/NTU201800597

Lemma 12.1. *Let $s_n^\star$ attain the maxima in Eq. (12.51) and P_n satisfy Eq. (12.43). We have*

(a) The limit point of $\{P_n\}_{n\in\mathbb{N}}$ is capacity achieving.

(b) $s_n^\star > 0$ for all $n \in \mathbb{N}$ and $\lim_{n\to+\infty} s_n^\star = 0$.

Proof of Lemma 12.1. Let $(P_{n_k})_{k\in\mathbb{N}}$ and $(s_{n_k}^\star)_{k\in\mathbb{N}}$ be arbitrary subsequences. Since $\mathcal{P}(\mathcal{X})$ and $[0,1]$ are compact, we may assume that

$$\lim_{k\to+\infty} P_{n_k} = P_o, \quad \lim_{k\to\infty} s_{n_k}^\star = s_o, \tag{12.52}$$

for some $P_o \in \mathcal{P}(\mathcal{X})$ and $s_o \in [0,1]$.

(12.1-(a)) Let $k \to +\infty$. Eq. (12.43) implies that

$$D(\mathcal{W}\|P^\star\mathcal{W}|P_o) = C_{\mathcal{W}}, \tag{12.53}$$

which guarantees that P_o is capacity-achieving by the dual representation of the information radius, see e.g. [151], [17, Theorem 2].

(12.1-(b)) One can observe from Eq. (12.51) that $s_n^\star = 0$ if and only if $C_{\mathcal{W}} - \delta_n \geq D(\mathcal{W}\|P^\star\mathcal{W}|P_n)$. However, this violates the assumption in Eq. (12.43). Hence, we have $s_n^\star > 0$ for all $n \in \mathbb{N}$.

Since P_o is capacity achieving, the uniqueness of the divergence center implies that $P_o\mathcal{W} = P^\star\mathcal{W}$. Item (c) in Proposition 9.3 shows that

$$\left.\frac{\partial^2 E_{\mathrm{h}}(s,P_o)}{\partial s^2}\right|_{s=0} = -V(\mathcal{W}\|P^\star\mathcal{W}|P_o) = -V(P_o,\mathcal{W}) \leq -V_{\mathcal{W}} < 0, \tag{12.54}$$

where the last inequality follows since $V_{\mathcal{W}} > 0$. Then, Eq. (12.54) implies that the first-order derivative $\partial E_{\mathrm{h}}(s,P_o)/\partial s$ is strictly decreasing around $s = 0$. Moreover, item (d) in Proposition 9.3 gives

$$\left.\frac{\partial E_{\mathrm{h}}(s,P_o)}{\partial s}\right|_{s=s_o} \leq D(\mathcal{W}\|P^\star\mathcal{W}|P_o) = C_{\mathcal{W}}, \tag{12.55}$$

This, together with items (b) and (c) in Proposition 9.3, shows that the first inequality in Eq. (12.55) becomes an equality if and only if $s_o = 0$. Since the subsequence was arbitrary, item (b) is shown.

□

Now we are ready to prove this proposition. We start with proving Eq. (12.46). Since $s \mapsto E_{\mathrm{h}}(s,\cdot)$ is concave from item (b) in Proposition 9.3, the maximizer $s_n^\star$ must satisfy

$$\left.\frac{\partial E_{\mathrm{h}}(s,P_{n_k})}{\partial s}\right|_{s=s_{n_k}^\star} = C_{\mathcal{W}} - \delta_{n_k}. \tag{12.56}$$

doi:10.6342/NTU201800597

Further, item (c) in Proposition 9.3 gives

$$\left.\frac{\partial E_{\mathrm{h}}\left(s, P_{n_k}^{\star}\right)}{\partial s}\right|_{s=0}=D\left(\mathcal{W}\|P^{\star}\mathcal{W}|P_{n_k}^{\star}\right). \tag{12.57}$$

The mean value theorem states that there exists a number $\hat{s}_{n_k} \in \left(0, s_{n_k}^{\star}\right)$, for each $k \in \mathbb{N}$, such that

$$\begin{aligned} -\left.\frac{\partial^2 E_{\mathrm{h}}\left(s, P_{n_k}\right)}{\partial s^2}\right|_{s=\hat{s}_{n_k}} &= \frac{D\left(\mathcal{W}\|P^{\star}\mathcal{W}|P_{n_k}\right)-C_{\mathcal{W}}+\delta_{n_k}}{s_{n_k}^{\star}} && (12.58)\\ &\leq \frac{\delta_{n_k}}{s_{n_k}^{\star}}, && (12.59)\end{aligned}$$

where the last inequality is again due to $D\left(\mathcal{W}\|P^{\star}\mathcal{W}|P_{n_k}^{\star}\right) \leq C_{\mathcal{W}}$. When k approaches infinity, items (a) and (e) in Proposition 9.3 give

$$\lim_{k\to+\infty}\left.\frac{\partial^2 E_{\mathrm{h}}\left(s, P_{n_k}\right)}{\partial s^2}\right|_{s=\hat{s}_{n_k}}=\left.\frac{\partial^2 E_{\mathrm{h}}\left(s, P_o\right)}{\partial s^2}\right|_{s=0}=-V(P_o, \mathcal{W}) \leq -V_{\mathcal{W}}. \tag{12.60}$$

Combining Eqs. (12.59) and (12.60) leads to

$$\limsup_{k\to+\infty}\frac{s_{n_k}^{\star}}{\delta_{n_k}} \leq \frac{1}{V_{\mathcal{W}}}. \tag{12.61}$$

Since the subsequence was arbitrary, the above result establishes Eq. (12.46).

Next, for any sufficiently large $n \geq N_0$, we apply Taylor's theorem to the map $s_n^{\star} \mapsto E_{\mathrm{h}}\left(s_n^{\star}, P_n\right)$ at the original point to obtain

$$\begin{aligned} &E_{\mathrm{sp}}^{(2)}\left(C_{\mathcal{W}}-\delta_n, P_n, P^{\star}\mathcal{W}\right)\\ &=-s_n^{\star}\left(C_{\mathcal{W}}-\delta_n\right)+E_{\mathrm{h}}\left(s_n^{\star}, P_n\right) && (12.62)\\ &=s_n^{\star}\left(\delta_n+D(\mathcal{W}\|P^{\star}W|P_n)-C_{\mathcal{W}}\right)-\frac{\left(s_n^{\star}\right)^2}{2}V\left(P_n, \mathcal{W}\right)+\frac{\left(s_n^{\star}\right)^3}{6}\left.\frac{\partial^3 E_{\mathrm{h}}(s, P_n)}{\partial s^3}\right|_{s=\bar{s}_n} && (12.63)\end{aligned}$$

for some $\bar{s}_n \in [0, s_n^{\star}]$. Let

$$\Upsilon=\max_{(s,P)\in[0,1]\times\mathcal{P}(\mathcal{X})}\left|\frac{\partial^3 E_{\mathrm{h}}\left(s, P\right)}{\partial s^3}\right|. \tag{12.64}$$

Continuing from Eq. (12.63) gives

$$\begin{aligned} E_{\mathrm{sp}}^{(2)}\left(C_{\mathcal{W}}-\delta_n, P_n, P^{\star}\mathcal{W}\right) &\leq s_n^{\star}(\delta_n-b_n)-\frac{\left(s_n^{\star}\right)^2}{2}V\left(P_n, \mathcal{W}\right)+\frac{\left(s_n^{\star}\right)^3\Upsilon}{6} && (12.65)\\ &\leq \sup_{s\geq 0}\left\{s(\delta_n-b_n)-\frac{s^2}{2}V(P_n, \mathcal{W})\right\}+\frac{\left(s_n^{\star}\right)^3\Upsilon}{6} && (12.66)\\ &=\frac{(\delta_n-b_n)^2}{2V(P_n, \mathcal{W})}+\frac{\left(s_n^{\star}\right)^3\Upsilon}{6}, && (12.67)\end{aligned}$$

where the first line follows from the assumption $D\left(\mathcal{W}\|P^{\star}\mathcal{W}|P_n\right) \leq C_{\mathcal{W}}-b_n$ in Eq. (12.43) and

doi:10.6342/NTU201800597

Eq. (12.64). Finally, Eq. (12.67), along with item (b) in Lemma 12.1 and Eq. (12.61), implies that

$$\limsup_{n\to+\infty} \frac{E_{\mathrm{sp}}^{(2)}\left(C_{\mathsf{W}}-\delta_n, P_n, P^\star\mathsf{W}\right)}{\delta_n^2} \le \limsup_{n\to+\infty} \frac{(\delta_n - b_n)^2}{2V(P_n,\mathsf{W})\delta_n^2} \tag{12.68}$$

$$\le \limsup_{n\to+\infty} \frac{(\delta_n - b_n)^2}{2V_{\mathsf{W}}\delta_n^2}, \tag{12.69}$$

where the last inequality follows from the continuity of $V(\cdot,\mathsf{W})$ on $\mathcal{P}(\mathcal{X})$ (Eq. (3.55)); the fact that $\{P_n\}_{n\in\mathbb{N}}$ is capacity achieving (item (a) in Lemma 12.1); and the definition of V_{W} in Eq. (3.57).

□

doi:10.6342/NTU201800597

Chapter 13

Conclusions and Open problems

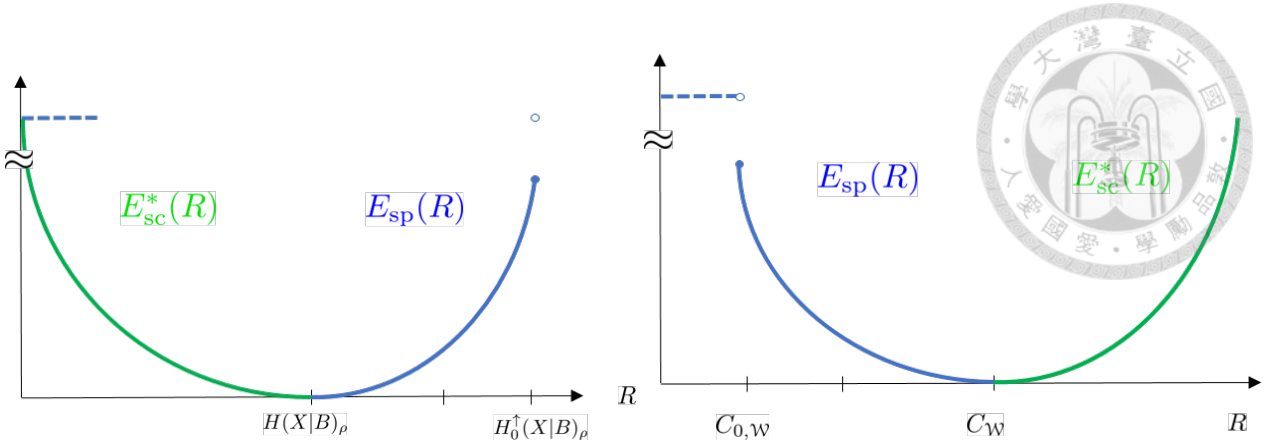
This thesis targets at characterizing the decoding error probability as a function of the coding block-englth. We study two fundamental quantum information processing protocols—the classical data compression (i.e. Slepian-Wolf coding) with quantum side information, and the classical-quantum channel coding. We have proven varieties of properties for the error exponent functions, which enables us to better understand the error behaviors of these information tasks. Then, we established numerous finite blocklength bounds for the optimal probability of error. Our results are not only of theoretical interests but also of practical values—they serve as the performance benchmark for designing the next generation quantum information technology. Lastly, we extend the derived finite blocklength results in the large deviation regime to the moderate deviation regime. We show that the optimal probability error vanishes asymptotically as the rate approaches the Slepian-Wolf limit/channel capacity slowly.

It is interesting to observe that there is an elegant duality between the two tasks when expressing the error exponent functions as conditional Rényi entropy and Rényi capacity. By exploiting this duality, we are able to unify the technical proofs these two tasks under the same framework of quantum hypothesis testing. Finally, we illustrate such relationship in Table 13.1 below, and depict the error exponent functions in Figure 13.1.

Bounds\Settings	Slepian-Wolf Coding with Quantum Side Information	Classical-Quantum Channel Coding
Achievability $(R < C_{\mathcal{W}}$ or $R > H(X\|B)_\rho)$	$E_{\mathrm{r}}(R) := \max_{0\le s\le 1}\{E_0(s) - sR\}$ $= \max_{1/2\le\alpha\le 1}\left\{\frac{1-\alpha}{\alpha}\left(R - H_\alpha^{\uparrow}(X\|Y)_\rho\right)\right\}$	$E_{\mathrm{r}}(R) := \max_{0\le s\le 1}\left\{\max_{P\in\mathcal{P}(\mathcal{X})} E_0(s,P) - sR\right\}$ $= \max_{1/2\le\alpha\le 1}\left\{\frac{1-\alpha}{\alpha}\left(C_{\alpha,\mathcal{W}} - R\right)\right\}$
Optimality $(R < C_{\mathcal{W}}$ or $R > H(X\|B)_\rho)$	$E_{\mathrm{sp}}(R) := \sup_{s\le 0}\{E_0(s) - sR\}$ $= \sup_{0\le\alpha\le 1}\left\{\frac{1-\alpha}{\alpha}\left(R - H_\alpha^{\uparrow}(X\|Y)_\rho\right)\right\}$	$E_{\mathrm{sp}}(R) := \sup_{s\le 0}\left\{\max_{P\in\mathcal{P}(\mathcal{X})} E_0(s,P) - sR\right\}$ $= \sup_{0\le\alpha\le 1}\left\{\frac{1-\alpha}{\alpha}\left(C_{\alpha,\mathcal{W}} - R\right)\right\}$
Strong Converse $(R > C_{\mathcal{W}}$ or $R < H(X\|B)_\rho)$	$E_{\mathrm{sc}}^*(R) := \sup_{-1<s<0}\left\{E_0^*(s) - sR\right\}$ $= \sup_{\alpha>1}\left\{\frac{1-\alpha}{\alpha}\left(R - H_\alpha^{*,\uparrow}(X\|Y)_\rho\right)\right\}$	$E_{\mathrm{sc}}^*(R) := \sup_{-1<s<0}\left\{\max_{P\in\mathcal{P}(\mathcal{X})} E_0^*(s,P) - sR\right\}$ $= \sup_{\alpha>1}\left\{\frac{1-\alpha}{\alpha}\left(C_{\alpha,\mathcal{W}}^* - R\right)\right\}$
Auxiliary Function	$E_0(s) := -\log \mathrm{Tr}_B\left[\left(\mathrm{Tr}_X(\rho_{XB})^{1/(1+s)}\right)^{1+s}\right]$	$E_0(s,P) := -\log \mathrm{Tr}\left[\left(\sum_{x\in\mathcal{X}} P(x)\cdot W_x^{1/(1+s)}\right)^{1+s}\right]$

Table 13.1: The comparison of the error exponent analysis for Slepian-Wolf coding with quantum side information and classical-quantum channel coding. We note that we only obtained suboptimal achievability results (i.e. with the exponent $E_{\mathrm{r}}^{\downarrow}(R)$ instead of $E_{\mathrm{r}}(R)$).

doi:10.6342/NTU201800597



(a) Slepian-Wolf coding with quantum side information (b) Classical-quantum channel coding

Figure 13.1: Sphere-packing exponents in two quantum information processing protocols.

13.1 Open Problems

There are still many open problems in the error exponent analysis. We divide them into the following categories: (a) Properties of the error exponent functions and auxiliary functions; (b) Random coding bound; (c) Sphere-packing bound; and (d) Moderate Deviation Analysis.

13.1.1 Properties of Error Exponent Functions and Auxiliary Functions

Problem 1 (Concavity). *For any classical-quantum channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$, define the* sandwiched auxiliary function*:*

$$E_0^*(s,P) := \min_{\sigma\in\mathcal{S}(\mathcal{H})} sD^*_{\frac{1}{1+s}}\left(P\circ W\|P\otimes\sigma\right), \quad (s,P)\in(-1,+\infty)\times\mathcal{P}(\mathcal{X}), \tag{13.1}$$

where we denote by

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log \mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right], \quad \forall\alpha\geq 0 \tag{13.2}$$

the sandwiched α-Rényi divergence [152, 64, 8].

Then, the that map $s\mapsto E_0^*(s,P)$ is concave for all $s\in(-1,0)$.

Remark 13.1. We are able to show that the map $s\mapsto E_0(s,P)$ is concave for all $s\in(-1,0)$, where $E_0(s,P)$ is defined via Petz's Rényi divergence. However, the sandwiched α-Rényi divergence has been shown the tightest entropic quantity in the strong converse domain [64, 58]. Hence, the concavity of the sandwiched auxiliary function is the most relevant. ◇

Problem 2 (Continuity of the Sphere-Packing Exponent). *Let $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ be a classical-quantum channel, and fix $R\in(C_{0,\mathcal{W}},C_{1,\mathcal{W}})$. For every $\nu>0$, there exists a constant $c>0$ such that for all $P\in\mathcal{P}(\mathcal{X})$ with $E_{\rm sp}(R,P)\geq\nu$ and,*

$$E^{(2)}_{\rm sp}(R,P)\leq E_{\rm sp}(R)-c\|\sigma_{\alpha_{R,P},P}-\sigma_{\alpha_R,\mathcal{W}}\|_1^2, \tag{13.3}$$

doi:10.6342/NTU201800597

where

$$E_{\mathrm{sp}}^{(2)}(R,P) := \sup_{0\leq\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(I_\alpha^{(2)}(P,\mathcal{W}) - R\right); \tag{13.4}$$

$$I_\alpha^{(2)}(P,\mathcal{W}) := \inf_{\sigma\in\mathcal{S}(\mathcal{H})} D_\alpha\left(\mathcal{W}\|\sigma|P\right); \tag{13.5}$$

and $\sigma_{\alpha,P}$, $\sigma_{\alpha,\mathcal{W}}$, $\alpha_{R,P}$, α_R are the optimizers such that

$$I_\alpha^{(2)}(P,\mathcal{W}) = D_\alpha\left(\mathcal{W}\|\sigma_{\alpha,P}|P\right); \tag{13.6}$$

$$C_{\alpha,\mathcal{W}} = \sup_{P\in\mathcal{P}(\mathcal{X})} D_\alpha\left(\mathcal{W}\|\sigma_{\alpha,\mathcal{W}}|P\right); \tag{13.7}$$

$$E_{\mathrm{sp}}^{(2)}(R,P) = \frac{1-\alpha_{R,P}}{\alpha_{R,P}}\left(I_{\alpha_{R,P}}^{(2)}(P,\mathcal{W}) - R\right); \tag{13.8}$$

$$E_{\mathrm{sp}}(R) = \frac{1-\alpha_R}{\alpha_R}\left(C_{\alpha_R,\mathcal{W}} - R\right). \tag{13.9}$$

13.1.2 Achievaibility: Random Coding Bound

We shorthand $\mathsf{P}_{\mathrm{RC}}(n) := \mathbb{E}_{\mathcal{C}_n}\left[\bar{\varepsilon}(\mathcal{W},\mathcal{C}_n)\right]$ the average probability of error for a n-blocklength random codes with distribution $P\in\mathcal{P}(\mathcal{X})$ on the input alphabet $\mathcal{X}$. Moreover, the following conditional Rényi entropies and Rényi divergences are defined via Petz's version [59]; see Eq. (3.5).

Problem 3 (Random Coding Bound for Slepian-Wolf Coding with Quantum Side Information)**.** *Consider a Slepian-Wolf coding with a joint classical-quantum state $\rho_{XB}\in\mathcal{S}(XB)$ with $H(X|B)_\rho > 0$. Let $R < H(X|B)_\rho$. The following holds for every $n\in\mathbb{N}$,*

$$\varepsilon^*(n,R) \leq \mathrm{e}^{-nE_{\mathrm{r}}(R)}, \tag{13.10}$$

where

$$E_{\mathrm{r}}(R) := \sup_{\frac{1}{2}\leq\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(R - H_\alpha^{\uparrow}(X|B)_\rho\right); \tag{13.11}$$

$$H_\alpha^{\uparrow}(X|B)_\rho := \sup_{\sigma_B\in\mathcal{S}(B)} -D_\alpha(\rho_{XB}\|\mathbb{1}_X\otimes\sigma_B). \tag{13.12}$$

Problem 4 (Random Coding Bound for Classical-Quantum Channels)**.** *For any classical-quantum channel $\mathcal{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, rate $R < C_{\mathcal{W}}$, and any $n\in\mathbb{N}$,*

$$\mathsf{P}_{\mathrm{RC}}(n) \leq \mathrm{e}^{-nE_{\mathrm{r}}(R,P)}, \tag{13.13}$$

where

$$E_{\mathrm{r}}(R,P) := \sup_{\frac{1}{2}\leq\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(I_\alpha^{(1)}(P,\mathcal{W}) - R\right); \tag{13.14}$$

$$I_\alpha^{(1)}(P,\mathcal{W}) := \inf_{\sigma\in\mathcal{S}(\mathcal{H})} D_\alpha(P\circ\mathcal{W}\|P\otimes\sigma). \tag{13.15}$$

doi:10.6342/NTU201800597

Moreover, the optimal probability of error can be upper bounded as

$$\varepsilon^*(n,R) \leq \mathrm{e}^{-nE_{\mathrm{r}}(R)}, \tag{13.16}$$

where $E_{\mathrm{r}}(R) := \sup_{P\in\mathcal{P}(\mathcal{X})} E_{\mathrm{r}}(R,P)$.

Problem 5 (Exact Asymptotics of Random Coding Bound for Classical-Quantum Channels)**.** *For any classical-quantum channel* $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ *and any* n*-blocklength block codes,*

$$\mathsf{P}_{\mathrm{RC}}(n) = \frac{1+o(1)}{\sqrt{n}}\mathrm{e}^{-nE_{\mathrm{r}}(R,P)}, \quad R \leq C_{1/2,\mathcal{W}} \tag{13.17}$$

$$\mathsf{P}_{\mathrm{RC}}(n) = \frac{1+o(1)}{n^{\frac{1}{2}\left(1+\left|\frac{\partial E_{\mathrm{r}}(R,P)}{\partial R}\right|\right)}}\mathrm{e}^{-nE_{\mathrm{r}}(R,P)}, \quad C_{1/2,\mathcal{W}} < R < C_{\mathcal{W}}. \tag{13.18}$$

Problem 6 (Random Coding Bound for Entanglement-Assisted Codes)**.** *Let* $\mathcal{N} : \mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathcal{B})$ *be a quantum channel. Fix any rate below the entanglement-assisted classical capacity, i.e.* $R < C_{\mathrm{ea}}(\mathcal{N})$ *The optimal probability of error over all* n*-blocklength entanglement-assisted codes can be upper bounded as*

$$\varepsilon^*_{\mathrm{ea}}(n,R) \leq \mathrm{e}^{-nE_{\mathrm{r,ea}}(R)}, \tag{13.19}$$

where

$$E_{\mathrm{r,ea}}(R) := \sup_{\frac{1}{2}\leq\alpha\leq 1}\sup_{\psi_{AA'}}\inf_{\sigma_B\in\mathcal{S}(B)} \frac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{N}_{A\to B}(\psi_{AA'})\|\rho_{A'}\otimes\sigma_B\right) - R\right), \tag{13.20}$$

and $\psi_{AA'}$ *denotes the purification of* ρ_A.

13.1.3 Optimality: Sphere-Packing Bound

We remark that the exact asymptotics of the sphere-packing for general codes in classical channels is still open. We do believe that the following Eq. (13.21) holds for both classical and c-q channels.

Problem 7 (Exact Asymptotics of Sphere-Packing Bound for Classical-Quantum Channels)**.** *For any classical-quantum channel* $\mathcal{W} : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ *and any* n*-block codes (not necessary constant composition codes)*[1]*,*

$$\varepsilon^*\left(n,R\right) \geq \frac{1}{n^{\frac{1}{2}\left(1+\left|\frac{\partial E_{\mathrm{sp}}(R)}{\partial R}\right|\right)}}\mathrm{e}^{-nE_{\mathrm{sp}}(R)}, \quad \forall R < C. \tag{13.21}$$

where,

$$E_{\mathrm{sp}}(R) := \sup_{0\leq\alpha\leq 1} \frac{1-\alpha}{\alpha}\left(C_{\alpha,\mathcal{W}} - R\right). \tag{13.22}$$

The sphere-packing bound beyond c-q channels are still unknown. We conjecture that it holds for an *entanglement-breaking channel* $\mathcal{N}_{\mathrm{EB}}$, whose classical capacity is additive, i.e $C(\mathcal{N}_{\mathrm{EB}}^{\otimes n}) = nC(\mathcal{N}_{\mathrm{EB}})$ [64, Theorem 18]. For general quantum channels, we might need regularized Rényi capacity.

[1] We note that the sphere-packing exponent is not necessarily differentiable. Throughout this section, we write $\partial E_{\mathrm{sp}}(R)/\partial R$ to be the left derivative.

doi:10.6342/NTU201800597

Problem 8 (Sphere-Packing Bound beyond Classical-Quantum Channels)**.** *For any entanglement-breaking channel $\mathcal{N}_{\mathrm{EB}}$, and any n-block codes*

$$\varepsilon^*\left(n,R\right)\geq\frac{1}{n^{\frac{1}{2}\left(1+\left|\frac{\partial E_{\mathrm{sp}}(R)}{\partial R}\right|\right)}}\mathrm{e}^{-nE_{\mathrm{sp}}(R)},\quad\forall R<C, \tag{13.23}$$

where

$$E_{\mathrm{sp}}(R):=\sup_{0\leq\alpha\leq1}\frac{1-\alpha}{\alpha}\left(C_\alpha(\mathcal{N}_{\mathrm{EB}})-R\right). \tag{13.24}$$

Moreover, for any quantum channel $\mathcal{N}$, and any n-block codes

$$\varepsilon^*\left(n,R\right)\geq\frac{1}{n^{\frac{1}{2}\left(1+\left|\frac{\partial E^\infty_{\mathrm{sp}}(R)}{\partial R}\right|\right)}}\mathrm{e}^{-nE^\infty_{\mathrm{sp}}(R)},\quad\forall R<C, \tag{13.25}$$

where

$$E^\infty_{\mathrm{sp}}(R):=\sup_{0\leq\alpha\leq1}\frac{1-\alpha}{\alpha}\left(\lim_{n\to+\infty}\frac{1}{n}C_\alpha(\mathcal{N}^{\otimes n})-R\right). \tag{13.26}$$

Problem 9 (Sphere-Packing Bound for Entanglement-Assisted Codes)**.** *Let $\mathcal{N}:\mathcal{S}(\mathcal{A})\to\mathcal{S}(\mathcal{B})$ be a quantum channel. Fix any rate below the entanglement-assisted classical capacity, i.e. $R<C_{\mathrm{ea}}(\mathcal{N})$. Then for any n-block codes*

$$\varepsilon^*_{ea}\left(n,R\right)\geq\frac{1}{n^{\frac{1}{2}\left(1+\left|\frac{\partial E_{\mathrm{sp,ea}}(R)}{\partial R}\right|\right)}}\mathrm{e}^{-nE_{\mathrm{sp,ea}}(R)},\quad\forall R<C, \tag{13.27}$$

where

$$E_{\mathrm{sp,ea}}(R):=\sup_{0\leq\alpha\leq1}\sup_{\psi_{AA'}}\inf_{\sigma_B\in\mathcal{S}(B)}\frac{1-\alpha}{\alpha}\left(D_\alpha\left(\mathcal{N}_{A\to B}(\psi_{AA'})\|\rho_{A'}\otimes\sigma_B\right)-R\right), \tag{13.28}$$

13.1.4 Moderate Deviation Analysis

Problem 10 (Moderate Deviation Analysis for Entanglement-Breaking Channels)**.** *Prove that any quantum entanglement-breaking channel $\mathcal{N}_{\mathrm{EB}}$ satisfies moderate deviation principle, i.e.*

$$\lim_{n\to+\infty}\frac{1}{na_n^2}\log\varepsilon^*\left(n,R\right)=-\frac{1}{2V(\mathcal{N}_{\mathrm{EB}})}, \tag{13.29}$$

where the sequence $(a_n)_{n\in\mathbb{N}}$ satisfy Eq. (12.1).

doi:10.6342/NTU201800597

Bibliography

[1] B. C. H and B. G, "Quantum cryptography: Public key distribution and coin tossing," *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, 1984.

[2] B. C. H, ". quantum cryptography using any two nonorthogonal states," *Physical Review Letters*, vol. 68, p. 3121–3124, 1992.

[3] B. C. H, B. G, C. C, Peres, and Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Physical Review Letters*, vol. 70, p. 1895, 1993.

[4] C. I. L. Gottesman D, "Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations," *Nature*, vol. 402, p. 390, 1999.

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2009.

[6] E. Gibney, "Chinese satellite is one giant step for the quantum internet," *Nature*, vol. 535, no. 7613, pp. 478–479, jul 2016.

[7] M. Tomamichel, M. Berta, and J. M. Renes, "Quantum coding with finite resources," *Nature Communications*, vol. 7, p. 11419, may 2016.

[8] M. Tomamichel, *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016.

[9] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[10] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, jul 1973.

[11] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," *Transactions of the Third Prague Conference on Information Theory*, pp. 689–723, 1962.

[12] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, may 2010.

[13] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, nov 2009.

doi:10.6342/NTU201800597

[14] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7693–7710, nov 2013.

[15] K. Li, "Second-order asymptotics for quantum hypothesis testing," *The Annals of Statistics*, vol. 42, no. 1, pp. 171–189, feb 2014.

[16] M. Tomamichel and V. Y. F. Tan, "Second-order asymptotics for the classical capacity of image-additive quantum channels," *Communications in Mathematical Physics*, vol. 338, no. 1, pp. 103–137, may 2015.

[17] ——, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7041–7051, nov 2013.

[18] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 10, no. 4, pp. 1–184, 2014.

[19] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2430–2438, may 2015.

[20] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, may 1959.

[21] R. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968. [Online]. Available: http://as.wiley.com/WileyCDA/WileyTitle/productCd-0471290483.html

[22] R. M. Fano, *Transmission of Information, A Statistical Theory of Communications*. The MIT Press, 1961.

[23] R. E. Blahut, *Principles and practice of information theory*. Addison-Wesley, 1987.

[24] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, "Reliability criteria in information theory and in statistical hypothesis testing," *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 2–3, pp. 97–263, 2007.

[25] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press (CUP), 2011.

[26] H.-C. Cheng, M.-H. Hsieh, and M. Tomamichel, "Quantum sphere-packing bounds with polynomial prefactors," 2017.

[27] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.

[28] A. Feinstein, "Error bounds in noisy channels without memory," *IEEE Transactions on Information Theory*, vol. 1, no. 2, pp. 13–14, sep 1955.

[29] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transaction on Information Theory*, vol. 11, no. 1, pp. 3–18, jan 1965.

doi:10.6342/NTU201800597

[30] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, jan 1967.

[31] E. A. Haroutunian, "Estimates of the error exponents for the semicontinuous memoryless channel," *Problemy Peredachi Informatsii*, vol. 4, no. 4, pp. 37–48, 1968, (in Russian). [Online]. Available: http://mi.mathnet.ru/eng/ppi1871

[32] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Transaction on Information Theory*, vol. 20, no. 4, pp. 405–417, jul 1974.

[33] A. Ben-Tal, M. Teboulle, and A. Charnes, "The role of duality in optimization problems involving entropy functionals with applications to information theory," *Journal of Optimization Theory and Applications*, vol. 58, no. 2, pp. 209–223, aug 1988.

[34] M. V. Burnashev and A. S. Holevo, "On the reliability function for a quantum communication channel," *Problems of information transmission*, vol. 34, no. 2, pp. 97–107, 1998.

[35] A. Holevo, "Reliability function of general classical-quantum channel," *IEEE Transaction on Information Theory*, vol. 46, no. 6, pp. 2256–2261, 2000.

[36] H.-C. Cheng and M.-H. Hsieh, "Concavity of the auxiliary function for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5960 – 5965, 2016.

[37] A. Winter, "Coding theorems of quantum information theory," 1999, (PhD Thesis, Universität Bielefeld).

[38] M. Dalai, "Lower bounds on the probability of error for classical and classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8027–8056, dec 2013.

[39] M. Dalai and A. Winter, "Constant compositions in the sphere packing bound for classical-quantum channels," *IEEE Transactions on Information Theory*, pp. 1–1, 2017.

[40] I. Devetak and A. Winter, "Classical data compression with quantum side information," *Physical Review A*, vol. 68, no. 4, oct 2003.

[41] J. M. Renes and R. Renner, "One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1985–1991, mar 2012.

[42] Y. Altuğ and A. B. Wagner, "Moderate deviation analysis of channel coding: Discrete memoryless case," in *2010 IEEE International Symposium on Information Theory*. Institute of Electrical & Electronics Engineers (IEEE), jun 2010.

[43] ——, "Moderate deviations in channel coding," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4417–4426, aug 2014.

[44] Y. Polyanskiy and S. Verdu, "Channel dispersion and moderate deviations limits for memoryless channels," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Institute of Electrical & Electronics Engineers (IEEE), sep 2010.

doi:10.6342/NTU201800597

[45] I. Sason, "Moderate deviations analysis of binary hypothesis testing," in *2012 IEEE International Symposium on Information Theory Proceedings*. Institute of Electrical & Electronics Engineers (IEEE), jul 2012.

[46] ——, "On refined versions of the Azuma-Hoeffding inequality with applications in information theory," 2011.

[47] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Foundations and Trends® in Communications and Information Theory*, vol. 10, no. 1-2, pp. 1–247, 2013.

[48] R. R. Bahadur and R. R. Rao, "On deviations of the sample mean," *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 1015–1027, dec 1960.

[49] N. R. Chaganty and J. Sethuraman, "Strong large deviation and local limit theorems," *The Annals of Probability*, vol. 21, no. 3, pp. 1671–1690, jul 1993.

[50] M. M. Wilde, *Quantum Information Theory*. Cambridge University Press, 2016.

[51] T. Heinosaari and M. Ziman, *The Mathematical language of Quantum Theory*. Cambridge University Press, 2009.

[52] H.-C. Cheng and M.-H. Hsieh, "Moderate deviation analysis for classical-quantum channels and quantum hypothesis testing," 2017.

[53] H.-C. Cheng, N. Datta, E. P. Hansen, and M.-H. Hsieh, "Non-asymptotic classical data compression with quantum side information," (in preparation).

[54] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Transactions on Information Theory*, vol. 26, no. 2, pp. 155–165, mar 1980.

[55] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, jul 1982.

[56] R. G. Gallager, "Source coding with side information and universal coding," 1976, technical Report. [Online]. Available: http://web.mit.edu/gallager/www/papers/paper5.pdf

[57] T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Transaction on Information Theory*, vol. 46, no. 7, pp. 2428–2433, 2000.

[58] M. Mosonyi and T. Ogawa, "Strong converse exponent for classical-quantum channel coding," *Communications in Mathematical Physics*, vol. 355, no. 1, pp. 373–426, jun 2017.

[59] D. Petz, "Quasi-entropies for finite quantum systems," *Reports on Mathematical Physics*, vol. 23, no. 1, pp. 57–65, feb 1986.

[60] S. Golden, "Lower bounds for the Helmholtz function," *Physical Review*, vol. 137, no. 4B, pp. B1127–B1128, feb 1965.

[61] C. J. Thompson, "Inequality with applications in statistical mechanics," *Journal of Mathematical Physics*, vol. 6, no. 11, p. 1812, 1965.

doi:10.6342/NTU201800597

[62] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transaction on Information Theory*, vol. 45, no. 7, pp. 2481–2485, 1999. [Online]. Available: http://dx.doi.org/10.1109/18.796385

[63] T. Ogawa and H. Nagaoka, "Strong converse to the quantum channel coding theorem," *IEEE Transaction on Information Theory*, vol. 45, no. 7, pp. 2486–2489, 1999.

[64] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy," *Communications in Mathematical Physics*, vol. 331, no. 2, pp. 593–622, jul 2014.

[65] M. S. Berger, *Nonlinearity and Functional Analysis.* Academic Press, 1977.

[66] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces.* Pergamon Press, New York, 1982.

[67] R. Bhatia, *Matrix Analysis.* Springer New York, 1997.

[68] K. Atkinson and W. Han, *Theoretical Numerical Analysis: A Functional Analysis Framework.* Springer International Publishing, 2009.

[69] N. J. Higham, *Functions of Matrices: Theory and Computation.* Society for Industrial & Applied Mathematics (SIAM), Jan 2008. [Online]. Available: http://dx.doi.org/10.1137/1.9780898717778

[70] V. V. Peller, "Hankel operators in the perturbation theory of unitary and self-adjoint operators," *Functional Analysis and Its Applications*, vol. 19, no. 2, pp. 111–123, 1985. [Online]. Available: http://dx.doi.org/10.1007/bf01078390

[71] K. Bickel, "Differentiating matrix functions," *Operators and Matrices*, no. 1, pp. 71–90, 2007. [Online]. Available: http://dx.doi.org/10.7153/oam-07-03

[72] M. Z. Nashed, "Some remarks on variations and differentials," *The American Mathematical Monthly*, vol. 73, no. 4, p. 63, Apr 1966. [Online]. Available: http://dx.doi.org/10.2307/2313752

[73] F. HIAI, "Matrix analysis: Matrix monotone functions, matrix means, and majorization," *Interdisciplinary Information Sciences*, vol. 16, no. 2, pp. 139–246, 2010. [Online]. Available: http://dx.doi.org/10.4036/iis.2010.139

[74] L. M. Graves, "Riemann integration and Taylor's theorem in general analysis," *Transactions of the American Mathematical Society*, vol. 29, no. 1, pp. 163–163, Jan 1927. [Online]. Available: http://dx.doi.org/10.1090/s0002-9947-1927-1501382-x

[75] K. Makherjea, *Differential Calculus in Normed Linear Spaces*, 2nd ed. Hindustan Book Agency, New Delhi, India, 2007.

[76] J. Dieudonne, *Foundations of Modern Analysis*, 2nd ed. Academic Press, New York, 1969.

[77] N. J. Higham and S. D. Relton, "Higher order fréchet derivatives of matrix functions and the level-2 condition number," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 3, pp. 1019–1037, Jan 2014. [Online]. Available: http://dx.doi.org/10.1137/130945259

doi:10.6342/NTU201800597

[78] F. Hansen, "Operator convex functions of several variables," *Publications of the Research Institute for Mathematical Sciences*, vol. 33, no. 3, pp. 443–463, 1997. [Online]. Available: http://dx.doi.org/10.2977/prims/1195145324

[79] J. R. Rice, "A theory of condition," *SIAM Journal on Numerical Analysis*, vol. 3, no. 2, pp. 287–310, Jun 1966. [Online]. Available: http://dx.doi.org/10.1137/0703023

[80] N. J. Higham and S. D. Relton, "Estimating the condition number of the fréchet derivative of a matrix function," *SIAM Journal on Scientific Computing*, vol. 36, no. 6, pp. C617–C634, Jan 2014. [Online]. Available: http://dx.doi.org/10.1137/130950082

[81] E. Carlen, "Trace inequalities and quantum entropy: an introductory course," in *Contemporary Mathematics*. American Mathematical Society (AMS), 2010, vol. 529, pp. 73–140. [Online]. Available: http://dx.doi.org/10.1090/conm/529/10428

[82] F. Hiai and D. Petz, *Introduction to Matrix Analysis and Applications*. Springer International Publishing, 2014.

[83] J. S. Matharu and J. S. Aujla, "Some inequalities for unitarily invariant norms," *Linear Algebra and its Applications*, vol. 436, pp. 1623–1631, 2012.

[84] H. Araki, "On an inequality of Lieb and Thirring," *Letters in Mathematical Physics*, vol. 19, no. 2, pp. 167–170, feb 1990.

[85] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, "Discriminating states: The quantum Chernoff bound," *Physical Review Letters*, vol. 98, p. 160501, apr 2007.

[86] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," *Communications in Mathematical Physics*, vol. 279, no. 1, pp. 251–283, feb 2008.

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

[88] M. Hayashi, "Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding," *Physical Review A*, vol. 76, no. 6, dec 2007.

[89] T. Ando, "Comparison of norms $|||f(a) - f(b)|||$ and $|||f(|a - b|)|||$," *Mathematische Zeitschrift*, vol. 197, no. 3, pp. 403–409, sep 1988.

[90] J.-C. Bourin and E.-Y. Lee, "Matrix inequalities from a two variables functional," *International Journal of Mathematics*, vol. 27, no. 09, p. 1650071, aug 2016.

[91] Y. Altuğ and A. B. Wagner, "Refinement of the sphere-packing bound: Asymmetric channels," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1592–1614, mar 2014.

[92] A. Rényi, "On measures of entropy and information," *Proc. 4th Berkeley Symp. on Math. Statist. Probability*, vol. 1, pp. 547–561, 1962.

doi:10.6342/NTU201800597

[93] M. Hayashi, *Quantum Information: An Introduction*. Springer.

[94] H. Umegaki, "Conditional expectation in an operator algebra. IV. entropy and information," *Kodai Mathematical Seminar Reports*, vol. 14, no. 2, pp. 59–85, 1962.

[95] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99–114, dec 1991.

[96] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., apr 2005.

[97] E. Hellinger, "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen," *Journal für die reine und angewandte Mathematik*, vol. 1909, no. 136, 1909.

[98] L. L. Cam and G. L. Yang, *Asymptotics in Statistics*. Springer US, 1990.

[99] E. H. Lieb, "Convex trace functions and the Wigner-Yanase-Dyson conjecture," *Advances in Mathematics*, vol. 11, no. 3, pp. 267–288, dec 1973.

[100] F. Hiai, "Concavity of certain matrix trace and norm functions. II," *Linear Algebra and its Applications*, vol. 496, pp. 193–220, may 2016.

[101] A. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problems of Information Transmission*, vol. 9, no. 3, pp. 177–183, 1973.

[102] ——, "Problems in the mathematical theory of quantum communication channels," *Reports on Mathematical Physics*, vol. 12, no. 2, pp. 273–278, oct 1977.

[103] ——, "The capacity of the quantum channel with general signal states," *IEEE Transaction on Information Theory*, vol. 44, no. 1, pp. 269–273, 1998.

[104] M. Hayashi and M. Tomamichel, "Correlation detection and an operational interpretation of the Rényi mutual information," 2014.

[105] B. Nakiboğlu, "Augustin's method - part i: The Renyi center."

[106] U. Augustin, "Noisy channels," 1978, habilitation thesis, Universitat Erlangen.

[107] B. Nakiboğlu, "The Augustin center and the sphere packing bound for memoryless channels," 2017.

[108] N. Sharma and N. A. Warsi, "On the strong converses for the quantum channel capacity theorems," 2012.

[109] C. Berge, *Topological Spaces*. Oliver & Boyd, 1963.

[110] B. Pshenichnyi, *Necessary Conditions for an Extremum Pshenichnyi*. CRC Press, 1971.

[111] H. Nagaoka, "The converse part of the theorem for quantum Hoeffding bound," 2006.

[112] M. Nussbaum and A. Szkoła, "The Chernoff lower bound for symmetric quantum hypothesis testing," *Annals of Statistics*, vol. 37, no. 2, pp. 1040–1057, apr 2009.

doi:10.6342/NTU201800597

[113] M. Junge and Q. Zeng, "Noncommutative Bennett and Rosenthal inequalities," *The Annals of Probability*, vol. 41, no. 6, pp. 4287–4316, nov 2013.

[114] S. Watanabe and M. Hayashi, "Finite-length analysis on tail probability for markov chain and application to simple hypothesis testing," *The Annals of Applied Probability*, vol. 27, no. 2, pp. 811–845, apr 2017.

[115] C. Rouzé and N. Datta, "Finite blocklength and moderate deviation analysis of hypothesis testing of correlated quantum states and application to classical-quantum channels with memory," 2016.

[116] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*. American Mathematical Society, 1992.

[117] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, mar 2006.

[118] G. Pisier and Q. Xu, "Non-commutative martingale inequalities," *Communications in Mathematical Physics*, vol. 189, no. 3, pp. 667–698, nov 1997.

[119] M. Tomamichel, M. Berta, and M. Hayashi, "Relating different quantum generalizations of the conditional rényi entropy," *Journal of Mathematical Physics*, vol. 55, no. 8, p. 082206, aug 2014.

[120] S. M. Lin and M. Tomamichel, "Investigating properties of a family of quantum rényi divergences," *Quantum Information Processing*, vol. 14, no. 4, pp. 1501–1512, feb 2015.

[121] R. M. Dudley, *Real Analysis and Probability*. Cambridge University Press (CUP), 2002.

[122] R. T. Rockafellar, "Minimax theorems and conjugate saddle-functions." *Mathematica Scandinavica*, vol. 14, p. 151, jun 1964.

[123] ——, *Convex Analysis*. Walter de Gruyter GmbH, jan 1970.

[124] S. Weis, "Quantum convex support," *Linear Algebra and its Applications*, vol. 435, no. 12, pp. 3168–3188, dec 2011.

[125] N. Sharma and N. A. Warsi, "Fundamental bound on the reliability of quantum information transmission," *Physical Review Letters*, vol. 110, no. 8, feb 2013.

[126] M. Mosonyi and T. Ogawa, "Two approaches to obtain the strong converse exponent of quantum hypothesis testing for general sequences of quantum states," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6975–6994, dec 2015.

[127] R. Bhatia and J. Holbrook, "Riemannian geometry and matrix geometric means," *Linear Algebra and its Applications*, vol. 413, no. 2-3, pp. 594–618, mar 2006.

[128] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer Nature, 2001.

[129] M. Hayashi, "Universal coding for classical-quantum channel," *Communications in Mathematical Physics*, vol. 289, no. 3, pp. 1087–1098, may 2009.

doi:10.6342/NTU201800597

[130] M. Dalai, 2017, (in preparation).

[131] J. Scarlett, "Reliable communication under mismatched decoding," 2014, phD thesis (University of Cambridge).

[132] Y. Altuğ and A. B. Wagner, "The third-order term in the normal approximation for singular channels," in *2014 IEEE International Symposium on Information Theory*. Institute of Electrical and Electronics Engineers (IEEE), jun 2014.

[133] J. K. Omura, "A lower bounding method for channel and source coding probabilities," *Information and Control*, vol. 27, no. 2, pp. 148–177, feb 1975.

[134] N. Elkayam and M. Feder, "Sphere packing bound for constant composition," 2016, (in preparation). [Online]. Available: http://www.eng.tau.ac.il/~elkayam/SPB_Abstract.pdf

[135] U. Augustin, "Error estimates for low rate codes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 1, pp. 61–88, 1969.

[136] B. Nakiboğlu, "Augustin's method - part ii: The sphere packing bound."

[137] A. Valembois and M. P. Fossorier, "Sphere-packing bounds revisited for moderate block lengths," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 2998–3014, dec 2004.

[138] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," in *2008 Information Theory and Applications Workshop*. Institute of Electrical & Electronics Engineers (IEEE), jan 2008.

[139] R. Ahlswede, *Storing and Transmitting Data*, A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds. Springer International Publishing, 2014.

[140] J. H. B. Kemperman, "Studies in coding theory I. technical report," *University of Rochester, NY*, 1962.

[141] ——, "On the optimum rate of transmitting information," *The Annals of Mathematical Statistics*, vol. 40, no. 6, pp. 2156–2177, dec 1969.

[142] U. Augustin, "Gedächtnisfreie kanäle für diskrete zeit," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 6, no. 1, pp. 10–61, 1966.

[143] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, vol. 1, no. 4, pp. 591–606, 1957. [Online]. Available: http://projecteuclid.org/euclid.ijm/1255380682

[144] N. Elkayam and M. Feder, "Achievable and converse bounds over a general channel and general decoding metric," in *2015 IEEE Information Theory Workshop (ITW)*. Institute of Electrical & Electronics Engineers (IEEE), apr 2015.

[145] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2576–2595, may 2013.

doi:10.6342/NTU201800597

[146] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel, "Moderate deviation analysis for classical communication over quantum channels," *Communications in Mathematical Physics*, vol. 355, no. 3, pp. 1283–1315, Nov 2017. [Online]. Available: https://doi.org/10.1007/s00220-017-2971-1

[147] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Physical Review Letters*, vol. 108, no. 20, may 2012.

[148] H. Nagaoka, "Strong converse theorems in quantum information theory," 2001, p. 33.

[149] L. V. Rozovsky, "Estimate from below for large-deviation probabilities of a sum of independent random variables with finite variances," *Journal of Mathematical Sciences*, vol. 109, pp. 2192–2209, 2002.

[150] J.-C. Bourin, "Matrix versions of some classical inequalities," *Linear Algebra and its Applications*, vol. 416, no. 2-3, pp. 890–907, jul 2006.

[151] B. Schumacher and M. D. Westmoreland, "Optimal signal ensembles," *Physical Review A*, vol. 63, no. 2, jan 2001.

[152] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: A new generalization and some properties," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122203, 2013.

doi:10.6342/NTU201800597