RATE-LIMITED QUANTUM-TO-CLASSICAL OPTIMAL TRANSPORT

RATE-LIMITED QUANTUM-TO-CLASSICAL OPTIMAL TRANSPORT

By HAFEZ M. GARMAROUDI, M.Sc.

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for

the Degree Philosophical Doctrine

McMaster University

PHILOSOPHICAL DOCTRINE (2023)

Hamilton, Ontario, Canada (Electrical and Computer Engineering)

TITLE:	Rate-Limited Quantum-to-Classical Optimal Transport
AUTHOR:	Hafez M. Garmaroudi M.Sc. in Electrical and Computer Engineering, McMaster University, Hamilton, Canada
SUPERVISOR:	Prof. Jun Chen
NUMBER OF PAGES:	xv, 130

Lay Abstract

We establish a coding theorem for rate-limited quantum-classical optimal transport systems with limited classical common randomness. The coding theorem, referred to as *output-constrained rate-distortion theorem*, characterizes the rate region of measurement protocols on a product quantum source state for faithful construction of a given classical destination distribution while maintaining the source-destination distortion below a prescribed threshold with respect to a general distortion observable. This theorem provides a solution to the problem of rate-limited optimal transport, which aims to find the optimal cost of transforming a source quantum state to a destination distribution via a measurement channel with a limited classical communication rate. The coding theorem is further extended to cover Bosonic continuous-variable quantum systems. The analytical evaluation is provided for the case of a qubit measurement system with unlimited common randomness, as well as the case of Gaussian quantum systems.

Abstract

The goal of optimal transport is to map a source probability measure to a destination one with the minimum possible cost. However, the optimal mapping might not be feasible under some practical constraints. One such example is to realize a transport mapping through an information bottleneck. As the optimal mapping may induce infinite mutual information between the source and the destination, the existence of an information bottleneck forces one to resort to some suboptimal mappings. Investigating this type of constrained optimal transport problems is clearly of both theoretical significance and practical interest.

In this work we substantiate a particular form of constrained optimal transport in the context of quantum-to-classical systems by establishing an *Output-Constrained Rate-Distortion Theorem* similar to the classical case in [45]. This theorem develops a noiseless communication channel and finds the least required transmission rate R and common randomness R_c to transport a sufficiently large block of n i.i.d. source quantum states, to samples forming a perfectly i.i.d. classical destination distribution, while maintaining the distortion between them. The coding theorem provides operational meanings to the problem of *Rate-Limited Optimal Transport*, which finds the optimal transportation from source to destination subject to the rate constraints on transmission and common randomness.

We further provide an analytical evaluation of the quantum-to-classical rate-limited optimal transportation cost for the case of qubit source state and Bernoulli output distributions with unlimited common randomness. The evaluation results in a transcendental system of equations whose solution provides the rate-distortion curve of the transportation protocol.

We further extend this theorem to continuous-variable quantum systems by employing a clipping and quantization argument and using our discrete coding theorem. Moreover, we derive an analytical solution for rate-limited Wasserstein distance of 2nd order for Gaussian quantum systems with Gaussian output distribution. We also provide a Gaussian optimality theorem for the case of unlimited common randomness, showing that Gaussian measurement optimizes the rate in a system with Gaussian source and destination.

To my beloved Zahra

Acknowledgements

I do believe that this Thesis is not just the result of one's PhD program, but rather a long journey that helped me learn and better understand the world with many falls and rises, and it was not possible to bear by myself alone. Therefore, I would like to provide my sincere thanks and dedicate this work.

To my supervisor Professor Jun Chen with utmost sincerity and dedication, who started it all and made all this possible. Always there when I needed him and always pushed me forward to do better. Each simple meeting with him was indeed a course in classical information theory for me. His help and support were simply beyond the scope of a supervisor and were the role of a mentor.

To my external co-supervisor Professor Sandeep Pradhan with utmost sincerity and dedication, who dedicated many hours of his precious time to teach me and correct me in the field of quantum information. Each meeting with him was indeed a course for me on this subject instructed with passion, care, and patience beyond expectations.

To the committee members Professor Kon Max Wong, Professor Tim Field, and Professor Tim Davidson, and Professor Amin Aminzadeh Gohari for their valuable comments and corrections. And to Dr. Sadaf Saleh Kalaibar for the discussions and insights she provided. To Ms. Cheryl Gies, Ms. Tracey Coop Ms. Christine

Rich, and other staff of McMaster University for their administration, advice, and support.

To McMaster's Association of Iranian Muslims Club for night events and family gatherings that made us feel at home. And to my friends and coaches at the fencing club of McMaster, Coaches Allen, and Ryan, and especially my fencing instructors, Head Coach Dave who gave me my first sword, and Coach Tony who taught me how to wield it.

To my beloved friends Mohsen Shahriari and Helia Mohammadi whom I only had the chance to see for one day every two years and it was enough to get me through the next year. Special thanks to Mohsen for our discussions on the topic of measure theory and other mathematical subjects. Also to my brothers and sisters at Infinite8, whose happy moments spent together will always lift my spirits: Varnous Faderani, Razin Kamali, Mojtaba Rabiei and Reza Karkon.

To Mom and Dad, who have always been there for me and shared themselves with me piece by piece. To my brother Hasan for always being positive and giving me joy even at the darkest times of Covid. To my beloved mentor Prof. Hojjat Al-Hagh Hosseini with utmost respect and gratitude for his advice and constant support. And to Ms. Samira Ranjbar for her support during the pandemic crisis.

And above all, to my beloved wife Zahra Alilou. My companion through this journey, was there with me in each step at every turn. The kindle that whispered in the dark void of my mind. This work would never have been completed without her.

Table of Contents

List of Figures

List of Tables

Notation and Abbreviations

Notation

$\mathbb{E}[.]$	Expectation operator
$\mathbb{E}_X[.]$	Expectation with respect to RV X
$\mathrm{Tr}\{.\}$	Trace of a quantum operator
$\mathrm{Tr}_A\{.\}$	Trace of a quantum operator with respect to local state A
Sp (.)	Trace of a squared matrix
$\det\{.\}$	Determinant of a squared matrix
ρ^A	Density operator ρ of a quantum state in system A
$\otimes$	Tensor product
τ^{RA}	Density operator τ of a composite quantum state in the reference systems R and local system A
$\mathcal{H}_A$	Hilbert space of system A
$\dim \mathcal{H}_A$	Dimension of the Hilbert space of system A

$\mathcal{L}^2(\mathbb{R}^n)$	Two-dimensional space of functionals with inputs in n-dimensional real space $\mathbb{R}^n$
X^n	Sequence of n samples $(X_1, ..., X_n)$
$\lvert\mathcal{X}\rvert$	Cardinality of a set $\mathcal{X}$
$\mathcal{T}_W^{n,\delta}$	Typical set of sequences of size n from RV W with tolerance δ
$\lvert\phi\rangle$	Ket operator indicating state ϕ
$\langle\phi\rvert$	Bra operator indicating conjugate of state ϕ
$I(.;.)$	Mutual information
$I_g(.;.)$	Groenwold's information gain
$\boldsymbol{A}^T$	Matrix transpose applied to matrix $\boldsymbol{A}$
$X^{\mathcal{T}}$	Hilbert-space transpose of operator X with respect to canonical eigenbasis
$X^\dagger$	conjugate transpose (Hermitian) of the operator X in Hilbert space
$\lVert X\rVert_1$	Trace norm of the operator X
$d_{TV}(P_X, P_Y)$	Total variation distance between P_X and P_Y

Abbreviations

RV	Random Variable
QC	Quantum-to-Classical

OC	Output Constrained
IID	Independent and Identically Distributed
POVM	Positive Operator Valued Measure
KKT	Karush-Kuhn-Tucker
TV	Total Variation
MMSE	Minimum Mean Square Error
PMR	Post-Measured Reference
PMF	Probability Mass Function

Declaration of Academic Achievement

1. H. M. Garmaroudi, S. Pradhan and J. Chen ”A Coding Theorem for Rate-Limited Quantum-Classical Optimal Transport”. arXiv, 17 May 2023. arXiv.org, https://doi.org/10.48550/arXiv.2305.10004,
   (to be submitted to IEEE Transactions on Information Theory)

2. H. M. Garmaroudi, S. Pradhan and J. Chen ”Rate-limited Quantum-to-Classical Optimal Transport: A Lossy Source Coding Perspective,” 2023 IEEE International Symposium on Information Theory, Taipei, Taiwan, 2023.

Chapter 1

Introduction

The lossy source-coding problem in information theory aims to determine the minimum required rate for compressing a given source so that it can be reconstructed to meet a prescribed distortion constraint. The fundamental tradeoff between the compression rate and the reconstruction distortion is characterized by the *rate-distortion function* [11]. This subject has also drawn attention in the field of quantum information theory. In an early attempt, Barnum [4] conjectured a lower bound on the rate-distortion function for a quantum channel with entanglement fidelity as the distortion measure, based on the coherent information quantity. His lower bound was later proved to be not tight in [16], where the authors established the quantum rate-distortion theorems for both entanglement assisted and unassisted systems. The proofs in [16] rely on the reverse Shannon theorem [5], which addresses the problem of simulating a noisy channel with the help of a noiseless channel, or more generally, simulating one noisy channel with another noisy channel.

In a seminal paper [53], Winter introduced the notion of information in quantum measurement and established the *measurement compression theorem*, which delineates the required classical rate and common randomness to faithfully simulate a feedback measurement Λ for an input state ρ. In [38], variants of this measurement compression theorem were studied for the case of non-feedback simulation and the case with the presence of quantum side information. Further, in [17], Datta et. al. invoked this measurement compression theorem to give a proof of the quantum-classical rate-distortion theorem. This idea of measurement simulation was further extended to distributed measurement simulation for composite quantum states in [1, 2], where the required classical rates and common randomness to faithfully simulate a bipartite state ρ_{AB} using distributed measurements are characterized; this distributed measurement compression theorem was then leveraged to establish inner and outer bounds of the rate region for the distributed quantum-classical rate-distortion problem.

In the classical setting, Cuff introduced the notion of coordination capacity [13, 14] and the problem of *distributed channel synthesis* [12]. A closely related problem, known as output-constrained lossy source coding [45], has recently found many applications in different areas [6, 7, 55, 9, 37]. In contrast to distributed channel synthesis which attempts to simulate a fixed channel, in output-constrained lossy source coding only the output distribution is fixed, rendering the problem intimately connected to optimal transport.

The goal of optimal transport is to map a source probability measure into a destination one with the minimum possible cost [48]. Let X be a random variable in the source probability space $(\mathcal{X}, \mathcal{F}_{\mathcal{X}}, P_X)$, where $\mathcal{X}$ is the support, $\mathcal{F}_{\mathcal{X}}$ is the event space defined by the σ-algebra of the Borel sets on $\mathcal{X}$, and P_X is the probability distribution function. Let Y be a random variable from the target probability space $(\mathcal{Y}, \mathcal{F}_\mathcal{Y}, P_Y)$. The optimal transport problem aims at finding an optimal mapping $f : \mathcal{X} \to \mathcal{Y}$ that minimizes the expectation of the transportation cost $c(x, y)$ [39]. However, as such deterministic mappings do not exist in many cases, one has to resort to stochastic channels to transform the source distribution to the target distribution. Thus the problem boils down to finding the optimal *coupling* π^* of marginal distributions P_X and P_Y that minimizes the transportation cost [34]:

$$\pi^* = \min_{\pi} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\pi(dx, dy),$$

subject to

$$\int_{\mathcal{X}} \pi(dx, B_\mathcal{X}) = P_Y(B_\mathcal{X}), \quad \int_{\mathcal{Y}} \pi(B_\mathcal{Y}, dy) = P_X(B_\mathcal{Y}),$$

for any $B_\mathcal{X} \in \mathcal{F}_\mathcal{X}$ and $B_\mathcal{Y} \in \mathcal{F}_\mathcal{Y}$. In [3], the authors introduced the problem of *information-constrained optimal transport* by imposing an additional constraint on coupling π in the form of a threshold on the mutual information between X and Y, and established an upper bound on the information-constrained Wasserstein distance by generalizing Talagrand's transportation cost inequality. It is worth noting that the information-cost function in [3] is equivalent to the rate-distortion function of output-constrained lossy source coding with unlimited common randomness [45].

The quantum version of optimal transport has also been investigated in recent years [15, 33, 18, 10, 20]. In [18], the authors proposed a generalization of the quantum Wasserstein distance of order 2 and proved that it satisfies the triangle inequality. They further showed that the associated quantum optimal transport schemes are in one-to-one correspondence with the quantum channels, and in the case of quantum thermal states, the optimal transport schemes can be realized by quantum Gaussian attenuators/amplifiers. In [20], the quantum Wasserstein distance of order 1, together with the quantum counterparts of some classical inequalities, was introduced.

The present paper focuses on the rate-limited quantum-classical optimal transport systems. Specifically, we establish a single-letter characterization of the rate region of measurement protocols on a product source state for the construction of a prescribed destination classical distribution with the distortion below a given threshold (see Theorem 2.1), and further extend this result to the Bosonic continuous-variable quantum systems via quantum clipping on the continuous source state (see Theorem 3.2). Our work enables the generalization of quantum optimal transport to the rate-limited setting as well as the generalization of classical information-constrained optimal transport to the quantum setting.

In particular, we provide a detailed analysis of rate-limited quantum optimal transport for the case of qubit source state and entanglement fidelity distortion measure; the minimum transportation cost is explicitly characterized and is shown to be achievable with a finite transmission rate (see Theorem 4.1 and 4.2). We further provide the evaluation of the Gaussian QC systems. First, we develop a Gaussian measurement optimality theorem (see Theorem 5.3) which shows for a continuous QC system with a Gaussian quantum source and Gaussian destination distribution, the rate-limited optimal transport is a Gaussian measurement. Moreover, a detailed analytical formulation provides the parameters of the optimal Gaussian measurement and the corresponding rate-limited 2nd-order Wasserstein distance. The important contrast to the classical optimal transport appears in the Gaussian QC system for which the Wasserstein distance is achieved at a finite transmission rate, which is a direct result of the uncertainty principle and the fact that the measurement noise cannot be made smaller than a threshold.

The contents of the Thesis are organized as follows. Finite dimensional quantum-classical systems are addressed in Chapter 2 with the statement of the coding theorem, the proof of the achievability part, the proof of the converse part, and the proof of the cardinality bound given in Sections 2.2, 2.3, 2.4, and 2.5, respectively. We extend the coding theorem to cover infinite dimensional systems in Chapter 3. Specifically, we introduce the continuity theorems that are needed for generalizing the definitions of measurement systems to continuous Hilbert spaces in Section 3.1, state the coding theorem for continuous Hilbert spaces in Section 3.2, and prove the achievability part in Section 3.3 (the proof of the converse part is the same as that for finite dimensional systems). In Chapter 4, we consider the case of qubit measurement systems with unlimited common randomness, for which a detailed analysis of rate-limited optimal transport is provided. The Gaussian measurement systems are analyzed in 5 which introduced the rate-limited QC 2nd-order Wasserstein distance. Finally, the proof of the important theorems and lemmas are provided in Appendices A, B and C.

Chapter 2

Finite Dimensional Quantum Systems

The system comprises of an n-letter memoryless source with its product state $\rho^{\otimes n}$ as the input of an encoder on Alice's side, where ρ is a density operator defined on a Hilbert space $\mathcal{H}_A$. On Bob's side, we have the reconstruction Hilbert space $\mathcal{H}_X$ representing the classical outcomes as quantum registered classical states, with an orthonormal basis indexed by a finite set $\mathcal{X}$. We also let the quantum state R denote the reference of the source with the associated Hilbert space $\mathcal{H}_R$ with $\dim(\mathcal{H}_R) = \dim(\mathcal{H}_A)$.

2.1 Distortion Measure

The distortion measure between two systems R and X is defined in the general form using a distortion observable $\Delta_{RX} > 0$ defined on $\mathcal{H}_R \otimes \mathcal{H}_X$ for the single-letter composite state τ^{RX}, as described in [17]:

$$d(\Delta_{RX}, \tau^{RX}) := \mathrm{Tr}\{\Delta_{RX}\tau^{RX}\}. \tag{2.1.1}$$

Then, having an n-letter composite state $\tau^{R^nX^n}$, and the distortion observable for each i-th system defined as $\Delta_{R_iX_i}$, the average n-letter distortion is defined as

$$d_n(\Delta^{(n)}, \tau^{R^nX^n}) := \mathrm{Tr}\{\Delta^{(n)}\tau^{R^nX^n}\} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Tr}\{\Delta_{R_iX_i}\tau^{R_iX_i}\}, \tag{2.1.2}$$

where $\tau^{R_iX_i} = \mathrm{Tr}_{[n]\setminus i}\{\tau^{R^nX^n}\}$ is the localized i-th composite state, and $\Delta^{(n)}$ is the average n-letter distortion observable defined as

$$\Delta^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \Delta_{R_iX_i} \otimes I_{RX}^{\otimes[n]\setminus i}. \tag{2.1.3}$$

In the case of a discrete QC system, the composite state has the form $\tau^{RX} = \sum_x P_X(x)\hat{\rho}_x^R \otimes |x\rangle\langle x|^X$, where ρ_x is the post-measurement reference state and $P_X(.)$ is the pmf of outcomes. We further decompose the distortion observable as $\Delta_{RX} = \sum_{t=1}^{T} \Delta_R^t \otimes \Delta_X^t$ using the Kronecker product decomposition. Thus, we get

$$\begin{aligned}\mathrm{Tr}\{\tau^{RX}\Delta_{RX}\} &= \mathrm{Tr}_R\left[\mathrm{Tr}_X\left[\left(\sum_x P_X(x)\hat{\rho}_x^R \otimes |x\rangle\langle x|^X\right)\left(\sum_{t=1}^{T}\Delta_R^t \otimes \Delta_X^t\right)\right]\right] \\ &= \sum_x P_X(x)\,\mathrm{Tr}_R\left\{\hat{\rho}_x^R\Delta_R(x)\right\} = \mathbb{E}_X\left[\mathrm{Tr}_R\left\{\hat{\rho}_X^R\Delta_R(X)\right\}\right], \end{aligned} \tag{2.1.4}$$

where $\{\Delta_R(x) : x \in \mathcal{X}\}$ is a mapping from the outcome space $\mathcal{X}$ to operators in Hilbert space $\mathcal{H}_R$:

$$\Delta_R(x) := \sum_{t=1}^{T} \Delta_R^t \left\langle x|\Delta_X^t|x\right\rangle .$$

Next, for the continuous QC system, by defining $\{\Delta_R(x)\}_{x\in\mathbb{R}}$ in the general form, the distortion measure for the continuous QC system is formulated as

$$\int_{x\in\mathbb{R}} \mathrm{Tr}_R\left[\hat{\rho}_x^R\, \Delta_R(x)\right] \mu(dx), \tag{2.1.5}$$

with $\mu(.)$ the probability measure of the outcome space.

Note that in order to prove the achievability of the output IID in perfect realism, we strict the distortion observables to be uniformly integrable according to the following definition.

Definition 2.1. *Consider a QC system with a distortion observable Δ_{RX} with operator mapping $x \to \Delta_R(x), x \in \mathcal{X}$ and an input quantum state ρ forming (Δ_{RX}, ρ). The pair is called uniformly integrable if for any $\epsilon > 0$ there exists a $\delta > 0$ such that*

$$\sup_{\Pi}\sup_{M} \mathbb{E}_X\left[\mathrm{Tr}_R\left\{\Pi_X\rho_X^R\Pi_X\Delta_R(X)\right\}\right] \le \epsilon, \tag{2.1.6}$$

where the supremum is over all POVMs $M \equiv \{M\}_{x\in\mathcal{X}}$ and all projectors of the form $\Pi = \sum_x \Pi_x \otimes |x\rangle\langle x|$ such that $\mathbb{E}_X\left[\mathrm{Tr}(\rho_X\Pi_X)\right] \le \delta$, and ρ_x^R is the post-measurement reference state of ρ given the outcome x with respect to M.

Furthermore, in the case of continuous quantum systems, we assume that the distortion observable also satisfies the following condition: the operator mapping

$x \mapsto \Delta_R(x)$ is uniformly continuous with respect to the trace norm.

2.2 Achievable Rate Region for Discrete States

The system is comprised of an n-letter source coding scheme defined below.

Definition 2.2. ***(Discrete Source Coding Scheme)*** *An (n, R, R_c) source-coding scheme for this QC system is comprised of an encoder $\mathcal{E}_n$ on Alice's side and a decoder $\mathcal{D}_n$ on Bob's side, with the following elements. The encoder is a set of $|\mathcal{M}| = \lfloor 2^{nR_c} \rfloor$ collective n-letter measurement POVMs $\Upsilon^{(m)} \equiv \{\Upsilon_l^{(m)}\}_{l\in\mathcal{L}}$, each comprised of $|\mathcal{L}| = \lfloor 2^{nR} \rfloor$ POVM operators corresponding to $|\mathcal{L}|$ outcomes and the randomly selected shared (with Bob) common randomness value m determines which POVM will be applied to the source state. Bob receives the outcome L of the measurement through a classical channel and applies a randomized decoder to this input pair (L, M) to obtain the final sequence X^n stored in a quantum register. Thus, the composite state of the reference and output induced by this coding scheme is*

$$\tau_{ind}^{R^n X^n} = \sum_{x^n} \sum_{m,l} \frac{1}{|\mathcal{M}|} \operatorname{Tr}_{A^n} \left\{ (id \otimes \Upsilon_l^{(m)})[\psi_{RA}^{\rho}]^{\otimes n}) \right\} \otimes \mathcal{D}_n(x^n|l, m) \left|x^n\right\rangle \left\langle x^n\right| . \quad (2.2.1)$$

We define the average n-letter distortion for the source coding system with encoder/ decoder pair $\mathcal{E}_n$ and $\mathcal{D}_n$, distortion observable Δ_{RX} and source product state $\rho^{\otimes n}$ as

$$d_n(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) = \operatorname{Tr}\{\Delta^{(n)}(\mathrm{id}_{R^n} \otimes \mathcal{D}_n \circ \mathcal{E}_n)(\psi_{RA}^{\rho})^{\otimes n}\}, \quad (2.2.2)$$

where ψ^{ρ}_{RA} is a purification of the source state ρ. Note that the different purifications of the system to the composite Hilbert spaces of different references R_1, R_2 and Alice's space $\mathcal{H}_{R_i} \otimes \mathcal{H}_A$ are equivalent up to an isometry [51, 38] which will not affect the trace distance nor the distortion. The n-letter distortion in (2.2.2) can also be written in the form of average over localized distortions as in (2.1.2).

The goal is to prepare the destination quantum ensemble on Bob's side while maintaining the distortion limit from the input reference state. Consequently, the following definition of achievability is used throughout this paper.

Definition 2.3. *(**Achievable pair**) A desired PMF Q_X on the output space $\mathcal{X}$ and a maximum tolerable distortion level D are given. Assuming a product input state of $\rho^{\otimes n}$, a rate pair (R, R_c) is defined achievable if for any sufficiently large n and any positive value $\epsilon > 0$, there exists an (n, R, R_c) coding scheme comprising of a measurement encoder $\mathcal{E}_n$ and a decoder $\mathcal{D}_n$ that satisfy:*

$$X^n \sim Q_X^n, \quad d_n(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) \leq D + \epsilon. \tag{2.2.3}$$

The expression (2.2.3) indicates that the output sequence must be IID with fixed distribution Q_X and that the n-letter distortion between the input state and output state must be asymptotically less than a threshold D. Then using the above definition of achievable pair, we further define the achievable rate region as:

Definition 2.4. *(**Achievable Rate Region**) Given the output PMF Q_X, the input state ρ and a distortion threshold D, the achievable rate region $\mathcal{R}(D, \rho||Q_X)$ is defined as the closure of all achievable rate pairs with respect to the given ρ, Q_X and D.*

We are specifically interested in finding the value of the minimum achievable rate as a function of the distortion level for any fixed rate of common randomness, which we define as follows:

Definition 2.5. ***(Output-Constrained Rate-Distortion Function)*** *For any coding system with the achievable rate region $\mathcal{R}(D, \rho||Q_X)$, and given a fixed common randomness rate R_c, the output-constrained rate-distortion function is defined as*

$$R(D; R_c, \rho||Q_X) \equiv \inf \{R : (R, R_c) \in \mathcal{R}(D, \rho||Q_X)\} . \tag{2.2.4}$$

The inverse of this function which for any fixed R_c, is a mapping from the communication rates to their corresponding minimum transportation cost, is called the Rate-Limited Optimal Transport Cost function and expressed by $D(R; R_c, \rho||Q_X)$.

Based on the above definitions, we establish the main theorem which provides the single-letter characterization of the achievable rate region as follows:

Theorem 2.1. *Given the distortion threshold D, the output PMF Q_X and having a product input state $\rho^{\otimes n}$, a rate pair (R, R_c) is inside the rate region $\mathcal{R}(D, \rho||Q_X)$, if and only if there exists an intermediate state W with a corresponding measurement POVM M_w^A and randomized post-processing transformation $P_{X|W}$ which satisfies*

$$R \geq I(W; R)_\tau, \tag{2.2.5}$$

$$R + R_c \geq I(W; X)_\tau, \tag{2.2.6}$$

where W, with a Hilbert space $\mathcal{H}_W$ along with an orthonormal basis indexed by a finite set $\mathcal{W}$, constructs a quantum Markov chain $R - W - X$ with the overall post-measured *composite state*

$$\tau^{RWX} = \sum_{w,x} P_{X|W}(x|w) \left(\sqrt{\rho} M_w \sqrt{\rho}\right)^R \otimes |w\rangle\langle w|^W \otimes |x\rangle\langle x|^X ,$$

from the set

$$\mathcal{M}(\mathcal{D}) = \left\{ \tau^{RWX} \left| \begin{array}{l} \sum_w P_{X|W}(x|w) \operatorname{Tr}\{M_w^A \rho\} = Q_X(x) \quad \textit{for } x \in \mathcal{X} \\ \mathbb{E}_X \left[\operatorname{Tr}\{\Delta_{RX} \tau^{RX}\}\right] \leq D \\ |\mathcal{W}| \leq (\dim \mathcal{H}_A)^2 + |\mathcal{X}| + 1 \end{array} \right. \right\}. \tag{2.2.7}$$

The overall state can also be formulated in terms of the PMR states as

$$\tau^{RWX} = \sum_{w,x} \hat{\rho}_w \otimes P_W(w) \, |w\rangle\langle w|^W \otimes P_{X|W}(x|w) \, |x\rangle\langle x|^X ,$$

where $\hat{\rho}_w := \frac{1}{P_W(w)} \sqrt{\rho} M_w \sqrt{\rho}$, and $P_W(w) = \operatorname{Tr}\{M_w \rho\}$, are the conditional post-measurement reference state given the outcome w, and the probability of the outcome w, respectively.

2.3 Proof of Achievability for Finite Systems

In this section, we prove the achievability of Theorem 2.1, given that the pair (Q_X, D) is provided by the setting of the theorem.

2.3.1 Codebook and Encoder Construction

In this section, by following the random codebook construction of [53], we generate a codebook in the intermediate space $\mathcal{W}$ from the probability $P_W(w) = \text{Tr}\{M_w \rho\}$, which is derived from applying the measurement POVM $M \equiv \{M_w\}_{w\in\mathcal{W}}$ to source state ρ. Then a sequence of n independent outcomes W has the IID distribution $P_W^n(w^n)$. The pruned distribution is then defined by only selecting w^n from the typical set of W,

$$P_{W'^n}(w^n) = \begin{cases} P_W^n(w^n)/S & \text{if } w^n \in \mathcal{T}_W^{n,\delta} \\ 0 & \text{o.w.} \end{cases}, \tag{2.3.1}$$

where $S := \Pr\Big\{W^n \notin \mathcal{T}_W^{n,\delta}\Big\} \geq 1-\epsilon$ and ϵ is a presumed fixed parameter. Consequently, a total of $|\mathcal{M}| \times |\mathcal{L}|$ random codewords w^n are generated from the pruned distribution $P_{W'^n}$ and indexed with (m, l) pair, comprising a random codebook. We then repeat this process to generate $|\mathcal{K}|$ codebook realizations. The codewords in each codebook are indexed as $W^n(m, k, l)$. The random variable K is introduced as additional randomness for analytical purposes which will be de-randomized at the end.

Also, for each w^n sequence, the following set of typically projected post-measurement reference operators is defined

$$\hat{\rho}'_{w^n} := \Pi^n_{\rho,\delta}\, \Pi^n_{\hat{\rho}_{w^n},\delta}\, \hat{\rho}_{w^n}\, \Pi^n_{\hat{\rho}_{w^n},\delta}\, \Pi^n_{\rho,\delta},$$

where $\Pi^n_{\rho,\delta}$ and $\Pi^n_{\hat{\rho}_{w^n},\delta}$ are the typical set and conditional typical set projectors respectively [38]. We are also interested in the expectation of the above operators with respect to the typical sequence $W^n \in \mathcal{T}_W^{n,\delta}$, which is

$$\hat{\rho}'^n := \mathbb{E}_{W'^n}\left[\hat{\rho}'^n_{W'^n}\right] = \sum_{w^n \in \mathcal{T}_W^{n,\delta}} P_{W'^n}(w^n)\hat{\rho}'^n_{w^n}.$$

Further define a cut-off projector Π, which projects to the subspace spanned by the eigenstates of $\hat{\rho}'^n$ with eigenvalues larger than $\epsilon\alpha$, where $\alpha := 2^{-n[H(R)+\delta]}$. Then the cut-off version of the operators and the expected cut-off operator are given by

$$\hat{\rho}''_{w^n} := \Pi\ \hat{\rho}'_{w^n}\ \Pi, \quad \hat{\rho}''^n := \Pi\ \hat{\rho}'^n\ \Pi. \tag{2.3.2}$$

Consequently, similar to [38], for each (k, m) we define the POVM operators of the form

$$\Upsilon_l^{(k,m)} := \frac{1-\epsilon}{1+\eta}\frac{1}{|\mathcal{L}|}\omega^{-1/2}\hat{\rho}''_{W^n(m,k,l)}\omega^{-1/2}, \tag{2.3.3}$$

with $\omega := \rho^{\otimes n}$, and $\eta \in (0, 1)$ a parameter which can be determined later. One can alternatively define the POVM operators such that it directly outputs W^n:

$$\Gamma_{w^n}^{(m,k)} := \sum_l \mathbb{1}\{W^n(m,k,l) = w^n\}\Upsilon_l^{(k,m)} = \gamma_{w^n}^{(m,k)}\ \omega^{-1/2}\ \hat{\rho}''_{w^n}\ \omega^{-1/2}, \tag{2.3.4}$$

where

$$\gamma_{w^n}^{(k,m)} := \frac{1}{|\mathcal{L}|}\sum_{l=1}^{|\mathcal{L}|}\frac{1-\epsilon}{1+\eta}\mathbb{1}\left\{W^n(m,k,l) = w^n\right\}. \tag{2.3.5}$$

Then the set of measurement operators for each $m \in \mathcal{M}$ and $k \in \mathcal{K}$ is given by

$$\tilde{M}_{\Gamma}^{(m,k),n} = \{\Gamma_{w^n}^{(m,k)} : w^n \in \mathcal{T}_W^{n,\delta}\}.$$

Examining each codebook $k \in \mathcal{K}$ generated randomly from pruned distribution $p_{W'^n}(w^n)$, using the operator Chernoff bound similar to [38], we claim that as long as $|\mathcal{L}| \geq 2^{n(I(R;W)+3\delta)}$, the following set of events $E_{m,k}$ happen with probability close to 1 for an arbitrary value $\eta \in (0,1)$:

$$E_{m,k}: \quad \frac{1}{|\mathcal{L}|}\sum_{l} \hat{\rho}''_{W^n(m,k,l)} \in [(1 \pm \eta)\hat{\rho}''^n] \quad \forall m \in \mathcal{M}. \tag{2.3.6}$$

Then following the analysis in [38] on the validity of POVM, we claim that the set $\tilde{M}_{\Gamma}^{(m,k),n}$ forms a sub-POVM, i.e.,

$$\sum_{w^n \in \mathcal{T}_W^{n,\delta}} \Gamma_{w^n}^{(m,k)} \leq I \quad \forall m \in \mathcal{M},\ k \in \mathcal{K}.$$

Thus, we can complete the sub-POVM by appending the following extra operator

$$\Gamma_{w_0^n}^{(m,k)} := I - \sum_{w^n} \Gamma_{w^n}^{(m,k)}.$$

We define the new set of operators as $[\tilde{M}_{\Gamma}^{(m,k),n}]$ which is a valid POVM.

Having the above construction, the intermediate POVM is established by randomly picking one of the $m \in |\mathcal{M}|$ POVMs according to the common randomness:

$$\tilde{\Lambda}_{w^n}^{(k)^A} := \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|} \Gamma_{w^n}^{(m,k)}, \qquad \forall k \in \mathcal{K}.$$

The decoder comprises of applying the $P_{X|W}$ classical memoryless channel to each element of w^n as $P^n_{X|W}$. Therefore, the form of the encoder/decoder is

$$\tilde{\tilde{\Lambda}}^{(k)}_{x^n} \equiv \sum_{w^n \in \mathcal{W}^n} P^n_{X|W}(x^n|w^n)\tilde{\Lambda}^{(k)^A}_{w^n}, \quad \forall x^n \in \mathcal{X}^n. \tag{2.3.7}$$

It should be noted that this is not the final decoder. We will later modify this decoder to yield a non-product batch decoder in section 2.3.5, which is required to ensure a perfectly IID output distribution. Using the above POVMs one can write the induced composite state of the reference and output for each random codebook realization $k \in \mathcal{K}$ as

$$\tau^{R^n X^n}_{\text{ind},k} = \sum_{x^n} \operatorname{Tr}_{A^n}\left\{(\mathrm{id}^{\otimes n}_R \otimes \tilde{\tilde{\Lambda}}^{(k)}_{x^n})(\psi^{RA}_\rho)^{\otimes n}\right\} \otimes |x^n\rangle\langle x^n|. \tag{2.3.8}$$

2.3.2 Proof of Near IID Output Distribution

It turns out that the proof of near IID output distribution does not depend on the codebook index $k \in \mathcal{K}$. Therefore, we hereby remove the index k from all expressions of this subsection, which means the following formulations apply to any fixed $k \in \mathcal{K}$. By tracing over the reference state in (2.3.8) we write the output state

$$\sigma^{X^n}_{\text{ind}} = \sum_{x^n} \operatorname{Tr}\left\{(\mathrm{id}^{\otimes n}_R \otimes \tilde{\tilde{\Lambda}}_{x^n})(\psi^{RA}_\rho)^{\otimes n}\right\} |x^n\rangle\langle x^n|. \tag{2.3.9}$$

Also, from the conditions of the $\mathcal{M}(\mathcal{D})$ feasible set (2.2.7), the output desired tensor state is in the following form,

$$\left(\sigma_{\text{des}}^{X}\right)^{\otimes n} = \sum_{x^n} Q_X^n(x^n)\, |x^n\rangle\langle x^n| = \sum_{x^n} \sum_{w^n \in \mathcal{W}^n} P_{X|W}^n(x^n|w^n) P_W^n(w^n) \tag{2.3.10}$$

$$= \sum_{x^n} \left[\prod_{i=1}^{n} \sum_{w \in \mathcal{W}} P_{X|W}(x_i|w_i)\, \mathrm{Tr}\left\{ M_{w_i}^{A} \rho \right\} \right] |x^n\rangle\langle x^n| \, .$$

Consequently, the trace distance between the induced output state and the desired product output state is,

$$\begin{aligned} &\left\| \left(\sigma_{\text{des}}^{X}\right)^{\otimes n} - \sigma_{\text{ind}}^{X^n} \right\|_1 \\ &= \sum_{x^n} \left| \sum_{w^n} P_{X|W}^n(x^n|w^n) P_W^n(w^n) - \mathrm{Tr}\left\{ (\mathrm{id}_R \otimes \tilde{\tilde{\Lambda}}_{x^n})(\psi_\rho^{RA})^{\otimes n} \right\} \right| \\ &= \sum_{x^n} \left| \sum_{w^n} P_{X|W}^n(x^n|w^n) P_W^n(w^n) - \frac{1}{|\mathcal{M}|} \sum_{w^n \in \mathcal{W}^n} P_{X|W}^n(x^n|w^n)\, \mathrm{Tr}\left\{ \left(\sum_{m=1}^{|\mathcal{M}|} \Gamma_{w^n}^{(m)} \right) \omega \right\} \right| := S, \end{aligned}$$

where we substitute (2.3.7) to get the second equality. Then we split and bound the above term by $S \le S_1 + S_2$, where we separate the extra operator from the rest of the POVM

$$S_1 \triangleq \sum_{x^n} \left| \sum_{w^n} P_{X|W}^n(x^n|w^n) P_W^n(w^n) - \sum_{w^n \neq w_0^n} P_{X|W}^n(x^n|w^n)\, \mathrm{Tr}\left\{ \left(\frac{1}{|\mathcal{M}|} \sum_{m=1} \Gamma_{w^n}^{(m)} \right) \omega \right\} \right|, \tag{2.3.11}$$

$$S_2 \triangleq \sum_{x^n} \left| P_{X|W}^n(x^n|w_0^n)\, \mathrm{Tr}\left\{ \frac{1}{|\mathcal{M}|} \sum_{m=1} \left(I - \sum_{w^n \neq w_0^n} \Gamma_{w^n}^{(m)} \right) \omega \right\} \right|. \tag{2.3.12}$$

We further simplify S_1 by substituting (2.3.4) into (2.3.11) and bound it again by

$S_1 \leq S_{11} + S_{12}$ by adding and subtracting a proper term and using triangle inequality:

$$S_{11} \triangleq \sum_{x^n} \left| \sum_{w^n} P^n_{X|W}(x^n|w^n) P^n_W(w^n) - \frac{1}{|\mathcal{M}||\mathcal{L}|} \frac{1-\epsilon}{1+\eta} \sum_{m,l} P^n_{X|W}(x^n|W^n(l,m)) \right|, \tag{2.3.13}$$

$$\begin{aligned} S_{12} &\triangleq \frac{1}{|\mathcal{M}||\mathcal{L}|} \frac{1-\epsilon}{1+\eta} \sum_{x^n} \left| \sum_{m,l} P^n_{X|W}(x^n|W^n(l,m)) \left(1 - \mathrm{Tr}\{\hat{\rho}''_{W^n(l,m)}\}\right) \right| \\ &= \frac{1}{|\mathcal{M}||\mathcal{L}|} \frac{1-\epsilon}{1+\eta} \sum_{x^n} \sum_{m,l} P^n_{X|W}(x^n|W^n(l,m)) \left(1 - \mathrm{Tr}\{\hat{\rho}''_{W^n(l,m)}\}\right) \\ &= \frac{1}{|\mathcal{M}||\mathcal{L}|} \frac{1-\epsilon}{1+\eta} \sum_{m,l} \left(1 - \mathrm{Tr}\{\hat{\rho}''_{W^n(l,m)}\}\right). \end{aligned} \tag{2.3.14}$$

For S_{11}, using the classical soft-covering lemma [Lemma 2 [12]] with the condition that $R+R_c > I(X;W)$ one can provide a decaying upper bound for its expectation as

$$\mathbb{E}\left[S_{11}\right] \leq \frac{3}{2} \exp\{-tn\}, \tag{2.3.15}$$

for some $t > 0$. Also, by taking the expectation of S_{12} we have

$$\mathbb{E}\left[S_{12}\right] = \frac{1-\epsilon}{1+\eta}(1 - \mathrm{Tr}\{\hat{\rho}''^n\}) \leq \frac{1-\epsilon}{1+\eta}(2\epsilon + 2\sqrt{\epsilon}) \triangleq \epsilon_2, \tag{2.3.16}$$

where the equality follows from (2.3.2) and the inequality appeals to the properties of the typical set and the Gentle Measurement Lemma [38, 52, 40]. Next, we bound and simplify the expectation of S_2 by substituting (2.3.4) into (2.3.12):

$$\begin{aligned}
\mathbb{E}\left[S_2\right] &\leq \mathbb{E}\left[\frac{1}{|\mathcal{M}|}\sum_m \sum_{x^n} P^n_{X|W}(x^n|w_0^n)\left|\mathrm{Tr}\left\{\omega - \sum_{w^n \neq w_0^n} \gamma^{(m)}_{w^n}\hat{\rho}''_{w^n}\right\}\right|\right] \\
&= \frac{1}{|\mathcal{M}|}\sum_m \mathbb{E}\left[\left|1 - \mathrm{Tr}\left\{\sum_{w^n \neq w_0^n} \gamma^{(m)}_{w^n}\hat{\rho}''_{w^n}\right\}\right|\right] \\
&\overset{a}{=} 1 - \frac{1}{|\mathcal{M}|}\sum_m \mathrm{Tr}\left\{\sum_{w^n \neq w_0^n} \mathbb{E}\left[\gamma^{(m)}_{w^n}\right]\hat{\rho}''_{w^n}\right\} \\
&\overset{b}{\leq} 1 - \frac{1-\epsilon}{1+\eta}(1 - 2\epsilon - 2\sqrt{\epsilon}) = \frac{\eta + \epsilon}{1+\eta} + \epsilon_2 \triangleq \epsilon_3,
\end{aligned} \tag{2.3.17}$$

where in (a) we remove the absolute sign because the trace is always less than or equal to one, and (b) uses the result from [38]. Hence, combining (2.3.17), (2.3.16) and (2.3.15) we proved that the expected distance between the output state induced by the random codebook and the product single-letter state is arbitrarily small for sufficiently large n:

$$\mathbb{E}\left[\left\|(\sigma^X_{\text{des}})^{\otimes n} - \sigma^{X^n}_{\text{ind}}\right\|_1\right] \leq \epsilon_2 + \epsilon_3 + c\exp\{-tn\} \triangleq \epsilon_{os}. \tag{2.3.18}$$

2.3.3 Proof of Distortion Constraint

The average distortion for a codebook $k \in \mathcal{K}$ is given by

$$\begin{aligned} {d_n}^{\{k\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) &= \mathrm{Tr}\{\Delta^{(n)} \tau_{\text{ind},k}^{R^n X^n}\} \\ &= \mathrm{Tr}\left\{\Delta^{(n)}\left(\sum_{x^n} \mathrm{Tr}_{A^n}\left\{(\mathrm{id}_R^{\otimes n} \otimes \tilde{\tilde{\Lambda}}_{x^n}^{(k)})(\psi_\rho^{RA})^{\otimes n}\right\} \otimes |x^n\rangle\langle x^n|\right)\right\} \\ &= \mathrm{Tr}\left\{\Delta^{(n)}\left(\sum_{m,l} \frac{1}{|\mathcal{M}|} \mathrm{Tr}_{A^n}\left\{(\mathrm{id}_R^{\otimes n} \otimes \Upsilon_l^{(k,m)})(\psi_\rho^{RA})^{\otimes n}\right\} \otimes \sigma_{W^n(m,k,l)}\right)\right\}. \end{aligned} \tag{2.3.19}$$

where $\sigma_{w^n} = \sum_{x^n} P_{X|W}^n(x^n|w^n)\,|x^n\rangle\langle x^n|$ is the classical decoder channel.

Recall from Section 2.3.2 that in order to have a faithful near IID output state, we need to satisfy the conditions of soft-covering lemma $|\mathcal{M}||\mathcal{L}| > 2^{nI(X;W)}$, which is needed for (2.3.15). On the other hand, according to the non-feedback measurement compression theorem [38], we need a sum rate of at least $I(XR;W)$ to have a faithful measurement simulation. Thus, by setting $|\mathcal{K}| > 2^{n(I(XR;W)-I(X;W))}$, we define an inter-codebook average state

$$\tau_{\text{avg}}^n \equiv \sum_{k,m,l} \frac{1}{|\mathcal{K}||\mathcal{M}|} \mathrm{Tr}_{A^n}\left\{\left(\mathrm{id}_R \otimes \Upsilon_l^{(k,m)}\right)(\psi_\rho^{\otimes n})^{RA}\right\} \otimes \sigma_{w^n(m,k,l)}. \tag{2.3.20}$$

Consequently, according to non-feedback measurement compression theorem [38], this inter-codebook average state is a faithful simulation of the ideal product measurement system; i.e., for any $\epsilon_{mc} > 0$ and for all sufficiently large n,

$$\mathbb{E}_c\left[\left\|\tau^{\otimes n} - \tau_{\text{avg}}^n\right\|_1\right] \le \epsilon_{mc}, \tag{2.3.21}$$

where the expectation is over all codebook realizations.

Then, we bound the expected average distortion as follows:

$$\begin{aligned}
&\mathbb{E}_K\left[d_n{}^{\{K\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\right] = \frac{1}{|\mathcal{K}|}\sum_k d_n{}^{\{k\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\\
&= \mathrm{Tr}\left\{\Delta^{(n)}\left(\sum_{k,m,l}\frac{1}{|\mathcal{K}||\mathcal{M}|}\mathrm{Tr}_{A^n}\left\{\left(\mathrm{id}_R^{\otimes n}\otimes\Upsilon_l^{(k,m)}\right)(\psi_\rho^{\otimes n})^{RA}\right\}\otimes\sigma_{W^n(m,k,l)}\right)\right\}\\
&= d_{\max}\,\mathrm{Tr}\left\{\frac{\Delta^{(n)}}{d_{\max}}(\tau_{\text{avg}}^{(n)} - \tau^{\otimes n})\right\} + \mathrm{Tr}\left\{\Delta^{(n)}\tau^{\otimes n}\right\}\\
&\leq d_{\max}\left\|\tau_{\text{avg}}^{(n)} - \tau^{\otimes n}\right\|_1 + \mathrm{Tr}\left\{\Delta^{(n)}\tau^{\otimes n}\right\}\\
&\leq d_{\max}\left\|\tau_{\text{avg}}^{(n)} - \tau^{\otimes n}\right\|_1 + D,
\end{aligned} \tag{2.3.22}$$

where $d_{\max}$ is the largest eigenvalue of the distortion observable. The first inequality holds by definition of the trace distance and the fact that $0 \leq \frac{\Delta^{(n)}}{d_{\max}} \leq I$. The second inequality holds because the average distortion of n identical copies of the single-letter system is the same as single-letter distortion. Next, we take the expectation of both sides with respect to all possible codebook realizations. Thus, for all sufficiently large n,

$$\mathbb{E}_c\left[\mathbb{E}_K\left[d_n{}^{\{k\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\right]\right] \leq d_{\max}\mathbb{E}_c\left[\left\|\tau_{\text{avg}}^{(n)} - \tau^{\otimes n}\right\|_1\right] + D \leq d_{\max}\epsilon_{mc} + D,$$

where for the first inequality we take the expectation of (2.3.22) and the second inequality follows from (2.3.21). Further, the LHS of above inequality can be rewritten as follows by changing the order of expectations,

$$\begin{aligned} \mathbb{E}_c\left[\mathbb{E}_K\left[{d_n}^{\{K\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\right]\right] &= \mathbb{E}_K\left[\mathbb{E}_c\left[{d_n}^{\{K\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\right]\right] \\ &= \mathbb{E}_c\left[{d_n}^{\{k\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\right], \end{aligned} \tag{2.3.23}$$

where the second equality holds for any codebook $k \in \mathcal{K}$ and follows because the expectation of the distance measure over all codebooks is independent of K. Then it is proved that the expected average distortion for any codebook $k \in \mathcal{K}$ is asymptotically bounded by D:

$$\mathbb{E}_c\left[{d_n}^{\{k\}}(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\right] \leq D + d_{\max}\epsilon_{mc}. \tag{2.3.24}$$

2.3.4 Intersection of the Constraints

In this section, we show that the previous bounds on the expected codebook realizations have an intersection with nonzero probability. i.e., there exists a codebook realization that can realize all events together. The following four cases are the required events in achievability proof which were proved to hold for the expected codebook realizations. Here we show that the union of the opposite of these events happens with a probability strictly less than one. This ensures there exists at least one codebook realization that satisfies all the constraints. Note that δ and ϵ are the parameters of the typical set and the probability of the non-typical set respectively.

- Firstly, it is shown that the $\Gamma_{w^n}^{(m,k)}$ form valid sub-POVM for all $m \in \mathcal{M}$ and

$k \in \mathcal{K}$. This is considered as event E_1. Using the Chernoff Bound technique, [38] shows that if $R > I(R;W)_\sigma$ then

$$\Pr\{\neg E_1\} \leq c \exp\{-2^{n\delta}\epsilon^3\}. \tag{2.3.25}$$

for some $c > 0$.

- Secondly, define event E_2 as when $S_{11} \leq \exp\{-\nu n\}$ for some $\nu > 0$. Then by applying Markov inequality to expression (2.3.15), we find the bound

$$\Pr\{\neg E_2 : S_{11} \geq \exp\{-\nu n\}\} \leq \frac{3}{2}\frac{\exp\{-tn\}}{\exp\{-\nu n\}}. \tag{2.3.26}$$

- Third, the bounds on expectations of S_{12} and S_2. Let E_{31} and E_{32} be the corresponding events for these random variables. Then applying the Markov inequality to these inequalities (2.3.16), (2.3.17), we have the following bound for a fixed value $\delta_3 > 0$:

$$\Pr\{\neg E_{31} : S_2 \geq 2\delta_3\} \leq \frac{\mathbb{E}\left[E_{31}\right]}{\delta_3} \leq \frac{\epsilon_3}{2\delta_3}, \tag{2.3.27}$$

$$\Pr\{\neg E_{32} : S_{12} \geq 2\delta_3\} \leq \frac{\mathbb{E}\left[E_{32}\right]}{\delta_3} \leq \frac{\epsilon_3}{2\delta_3}. \tag{2.3.28}$$

Note that we used the fact that $\epsilon_2 \leq \epsilon_3$.

- Fourth, define E_4 as the event when the average n-letter distortion constraint is satisfied. By applying Markov inequality to (2.3.24) we obtain for any fixed value $\delta_d > 0$ that

$$\Pr\big\{\neg E_4 : d(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) \geq D + \delta_d\big\} \leq \frac{\mathbb{E}_c\left[d(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n)\right]}{D + \delta_d} \leq \frac{D + d_{\max}\epsilon_{mc}}{D + \delta_d}. \tag{2.3.29}$$

Then the probability of not being in the intersection is bounded by using the union bound

$$\Pr\left\{\neg\bigcap_{i=1}^{4} E_i\right\} \leq \sum_{i=1}^{4} \Pr\{\neg E_i\} \leq c\exp\big\{-2^{n\delta}\epsilon^3\big\} + \frac{3}{2}\frac{\exp\{-kn\}}{\exp\{-\nu n\}} + \frac{\epsilon_3}{\delta_3} + \frac{D + d_{\max}\epsilon_{mc}}{D + \delta_d}. \tag{2.3.30}$$

By taking the limit of the above expression when $n \to \infty$, we have $\epsilon \to 0$. Then with proper choice of δ and $\nu \in (0, t)$, the first two terms decay exponentially, while $\epsilon_{mc}, \delta_3, \delta_d$ are fixed, thus,

$$\lim_{n\to\infty} \left(c\exp\big\{-2^{n\delta}\epsilon^3\big\} + \frac{3}{2}\frac{\exp\{-tn\}}{\exp\{-\nu n\}} + \frac{\epsilon_3}{\delta_3} + \frac{D + d_{\max}\epsilon}{D + \delta_d}\right) = \frac{\epsilon_3}{\delta_3} + \frac{D + d_{\max}\epsilon_{mc}}{D + \delta_d} < 1. \tag{2.3.31}$$

By choosing proper values for the ϵ_{mc}, δ_3 and δ_d parameters, we make sure the above inequality holds, which means there exists with nonzero probability, a valid quantum measurement coding scheme that satisfies all the above four conditions together.

2.3.5 Exactly Satisfying IID Output Distribution

It remains to prove that the perfect IID output distribution can be achieved from the near-perfect one with an arbitrarily small increase in distortion level. The desired perfect IID distribution and the near-perfect induced output distribution for this source coding scheme are expressed by

$$P^{\text{des}}_{X^n}(x^n) := \big\langle x^n \big| (\sigma^X_{\text{des}})^{\otimes n} | x^n \big\rangle = Q^n_X(x^n) = \sum_{w^n} P^n_{X|W}(x^n|w^n) P^n_W(w^n), \tag{2.3.32}$$

$$\begin{aligned} P^{\text{ind}}_{X^n}(x^n) &:= \big\langle x^n \big| \sigma^{X^n}_{\text{ind}} | x^n \big\rangle = \operatorname{Tr}\Big\{ (\text{id}_R \otimes \tilde{\tilde{\Lambda}}_{x^n})(\psi^\rho_{RA})^{\otimes n} \Big\} \\ &= \sum_{w^n \in W^n} P^n_{X|W}(x^n|w^n) \operatorname{Tr}\Big\{ (\text{id}_R \otimes \tilde{\Lambda}^A_{w^n})(\psi^\rho_{RA})^{\otimes n} \Big\}, \end{aligned} \tag{2.3.33}$$

where, $(\sigma^X_{\text{des}})^{\otimes n}$ and $\sigma^{X^n}_{\text{ind}}$ are defined in (2.3.10) and (2.3.9), respectively.

Then [Theorem 1 [49]] shows that by fixing the measurement while changing only the IID post-processing unit $P_{X|W}$ to batch decoder $\tilde{P}_{X^n|W^n}$ we can satisfy the perfect IID condition from the near-perfect one. Define the alternative decoder as the conditional probability of any event $A \subseteq \mathcal{X}^n$ given w^n as

$$\tilde{P}_{\hat{X}^n|W^n}(A|w^n) = \sum_{x^n \in A \cap \mathcal{X}^n_+} \theta_{x^n} P^n_{X|W}(x^n|w^n) + P^n_{X|W}(A \setminus \mathcal{X}^n_+|w^n) + \phi_{w^n} Z(A), \tag{2.3.34}$$

where the given expressions are defined as

$$\mathcal{X}_+^n := \left\{ x^n \in \mathcal{X}^n \middle| P_{X^n}^{\text{ind}}(x^n) > P_{X^n}^{\text{des}}(x^n) \right\}, \tag{2.3.35}$$

$$\theta_{x^n} := \frac{P_{X^n}^{\text{des}}(x^n)}{P_{X^n}^{\text{ind}}(x^n)} \quad x^n \in \mathcal{X}_+^n, \tag{2.3.36}$$

$$\phi_{w^n} := \sum_{x^n \in \mathcal{X}_+^n} (1-\theta_{x^n}) P_{X|W}^n(x^n|w^n), \tag{2.3.37}$$

$$Z(A) := \frac{\sum_{x^n \in A} \left[P_{X^n}^{\text{des}}(x^n) - P_{X^n}^{\text{ind}}(x^n) \right]^+}{\sum_{x^n \in \mathcal{X}^n \setminus \mathcal{X}_+^n} \left[P_{X^n}^{\text{des}}(x^n) - P_{X^n}^{\text{ind}}(x^n) \right]^+} = \frac{P_{X^n}^{\text{des}}(A \setminus \mathcal{X}_+^n) - P_{X^n}^{\text{ind}}(A \setminus \mathcal{X}_+^n)}{d_{TV}(P_{X^n}^{\text{ind}}, P_{X^n}^{\text{des}})}. \tag{2.3.38}$$

where $[x]^+ := \max(x, 0)$ for $x \in \mathbb{R}$. The validity and admissibility of the new post-processing decoder can be verified with simple calculus, which states that $\tilde{P}_{\hat{X}^n|W^n}(\mathcal{X}^n|w^n) = 1 \quad \forall w^n \in W^n$, and the new induced output distribution satisfies the desired IID condition

$$\tilde{P}_{\hat{X}^n}^{\text{ind}}(A) := \sum_{w^n} \tilde{P}_{\hat{X}^n|W^n}(A|w^n) \operatorname{Tr}\left\{ (\text{id}_R \otimes \tilde{\Lambda}_{w^n}^A)(\psi_{RA}^\rho)^{\otimes n} \right\} = P_{X^n}^{\text{des}}(A). \tag{2.3.39}$$

Also, using the definition of Batch decoder (2.3.34), the following set of equalities hold for any $w^n \in W^n$:

$$
\begin{aligned}
&d_{TV}\Big(\tilde{P}_{\hat{X}^n|W^n}(.|w^n), P^n_{X|W}(.|w^n)\Big)\\
&= \frac{1}{2}\sum_{x^n\in\mathcal{X}^n_+}\Big|\tilde{P}_{\hat{X}^n|W^n}(x^n|w^n) - P^n_{X|W}(x^n|w^n)\Big| + \frac{1}{2}\sum_{x^n\in\mathcal{X}^n\setminus\mathcal{X}^n_+}\Big|\tilde{P}_{\hat{X}^n|W^n}(x^n|w^n) - P^n_{X|W}(x^n|w^n)\Big|\\
&= \frac{1}{2}\sum_{x^n\in\mathcal{X}^n_+}\big|(\theta_{x^n}-1)P^n_{X|W}(x^n|w^n) + \phi_{w^n}Z(x^n)\big| + \frac{1}{2}\sum_{x^n\in\mathcal{X}^n\setminus\mathcal{X}^n_+}|\phi_{w^n}Z(x^n)|\\
&\overset{a}{=} \frac{1}{2}\sum_{x^n\in\mathcal{X}^n_+}\big|(\theta_{x^n}-1)P^n_{X|W}(x^n|w^n)\big| + \frac{1}{2}\sum_{x^n\in\mathcal{X}^n\setminus\mathcal{X}^n_+}\left|\phi_{w^n}\frac{P^{\text{des}}_{X^n}(x^n)-P^{\text{ind}}_{X^n}(x^n)}{d_{TV}(P^{\text{ind}}_{X^n},P^{\text{des}}_{X^n})}\right|\\
&= \frac{1}{2}\phi_{w^n} + \frac{1}{2}\phi_{w^n} = \phi_{w^n},
\end{aligned}
\tag{2.3.40}
$$

where (a) is because using the definition of $Z(A)$, we know that $Z(x^n) = 0$ for all $x^n \in \mathcal{X}^n_+$ and the last line is from the definition of total variation distance. Thus, by definition, there exists a coupling such that,

$$
\Pr\Big(X^n \neq \hat{X}^n\Big|w^n\Big) \leq \phi_{w^n} \quad \forall w^n \in \mathcal{W}^n. \tag{2.3.41}
$$

Then from the above inequality, using the argument in [49], the probability of outputs not being equal is bounded by

$$
\Pr\Big(X^n \neq \hat{X}^n\Big) \leq d_{TV}(P^{\text{ind}}_{X^n}, P^{\text{des}}_{X^n}) \leq \epsilon_{os}, \tag{2.3.42}
$$

and the second inequality appeals to (2.3.18) and the union bound argument in section 2.3.4.

Next, we bound the n-letter distortion for the new decoder using the above bound. First, note that $\tau_{R_i\hat{X}_i}$ is the local i-th reference-output state of the system, given by

$$\begin{aligned}
\tau^{R_i\hat{X}_i} &= \mathrm{Tr}_{R^{n\setminus\{i\}}X^{n\setminus\{i\}}}\{\tau^{R^n\hat{X}^n}\} \\
&= \sum_{x^n} \mathrm{Tr}_{R^{n\setminus\{i\}}A^n}\left\{(\mathrm{id}^{\otimes n}\otimes\hat{\Lambda}_{x^n})(\psi^{RA})^{\otimes n}\right\}\otimes|x_i\rangle\langle x_i| \\
&= \sum_{x_i} \mathrm{Tr}_{R^{n\setminus\{i\}}A^n}\left\{\left(\mathrm{id}^{\otimes n}\otimes\left(\sum_{x^{[n]\setminus i}}\hat{\Lambda}_{x^n}\right)\right)(\psi^{RA})^{\otimes n}\right\}\otimes|x_i\rangle\langle x_i| \\
&= \sum_{x_i} Q_X(x_i)\zeta_{x_i}^{R_i}\otimes|x_i\rangle\langle x_i|\,. \qquad (2.3.43)
\end{aligned}$$

The $\zeta_{x_i}^{R_i}$ is the post-measurement reference state of the i-th local state given the outcome x_i, given by

$$\begin{aligned}
\zeta_{x_i}^{R_i} &= \frac{1}{Q_X(x_i)}\langle\mathrm{id}\otimes x_i|\tau_{R_iX_i}|\mathrm{id}\otimes x_i\rangle \\
&= \frac{1}{Q_X(x_i)}\mathrm{Tr}_{R^{n\setminus\{i\}}A^n}\left\{\left(\mathrm{id}^{\otimes n}\otimes\left(\sum_{x^{[n]\setminus i}}\hat{\Lambda}_{x^n}\right)\right)(\psi^{RA})^{\otimes n}\right\}
\end{aligned}$$

where $\hat{\Lambda}\equiv\{\hat{\Lambda}_{x^n}\}_{x^n\in\mathcal{X}^{\otimes n}}$ is the combined encoder/decoder collective POVM taking into account the batch decoder.

We expand the n-letter average distortion as

$$\begin{aligned}
\mathrm{Tr}\Big\{\Delta^{(n)}\tau^{R^n\hat{X}^n}\Big\} &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{Tr}\left\{\Delta_{R_iX_i}\tau^{R_i\hat{X}_i}\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\hat{X}_i}\left[\mathrm{Tr}_R\left\{\zeta_{\hat{X}_i}^{R_i}\Delta_{R_i}(\hat{X}_i)\right\}\mathbb{1}_{\hat{X}_i=X_i}\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\hat{X}_i}\left[\mathrm{Tr}_R\left\{\zeta_{\hat{X}_i}^{R_i}\Delta_{R_i}(\hat{X}_i)\right\}\mathbb{1}_{\hat{X}_i\neq X_i}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{Tr}\big\{\Delta_{R_iX_i}\tau^{R_iX_i}\big\} + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\hat{X}_i}\left[\mathrm{Tr}_R\left\{\zeta_{\hat{X}_i}^{R_i}\Delta_{R_i}(\hat{X}_i)\right\}\mathbb{1}_{\hat{X}_i\neq X_i}\right] \\
&\leq D+\epsilon+\sup_{A}\mathbb{E}_X\left[\mathrm{Tr}_R\left\{\zeta_X^R\Delta_R(X)\right\}\mathbb{1}_A\right].
\end{aligned} \tag{2.3.44}$$

with A being any event with probability $P(A) = P(\hat{X}_i \neq X_i) \leq \epsilon_{os}$. By accepting the assumption that the system is uniformly integrable (which is always true for the systems with finite dimensions), it is derived from (2.3.44) that

$$\mathrm{Tr}\big\{\Delta^{(n)}\tau_{R^n\hat{X}^n}\big\} \leq D + \epsilon_4. \tag{2.3.45}$$

2.4 Proof of the Converse

Assume that there exists an achievable (n, R, R_c) coding scheme and a set of n-letter collective measurements $\Upsilon^{(m)}$ selected by a shared random number m on Alice's side. The measurement results in outcome L which is sent to Bob. Finally, Bob uses a batch decoder $P_{X^n|L,M}(x^n|l, m)$, to generate the final state. Thus, the n-letter encoding quantum measurement composite state is

$$\omega^{R^n LM} = \sum_{l,m} \mathrm{Tr}_{A^n} \left\{ (\mathrm{id} \otimes \Upsilon_l^{(m)})(\psi_\rho^{RA})^{\otimes n} \right\} \otimes \frac{1}{|\mathcal{M}|} |m\rangle\langle m| \otimes |l\rangle\langle l| \,. \tag{2.4.1}$$

2.4.1 Rate Inequalities

Let us assume that the above system is achievable in the sense of definition 2.3, similar to [38], for the communication rate, we have

$$\begin{aligned} nR \geq H(L)_\omega &\geq I(L;MR^n)_\omega = I(LM;R^n)_\omega + I(L;M)_\omega - I(M;R^n)_\omega \\ &\overset{a}{\geq} I(LM;R^n)_\omega = H(R^n)_\omega - H(R^n|LM)_\omega \\ &\overset{b}{\geq} \sum_k [H(R_k)_\omega - H(R_k|LM)_\omega] \\ &= \sum_k I(LM;R_k)_\omega \\ &= nI(LM;R_K|K)_\sigma \\ &= nI(LMK;R_K)_\sigma. \end{aligned} \tag{2.4.2}$$

The first two lines are well-known properties of mutual information. Also, (a) follows because common randomness M is independent of the source. (b) holds by the sub-additivity of conditional quantum entropy and the fact that the source is a product state. For Eq. (2.4.2), define K as a uniform random variable over the set $\{1, 2, ..., n\}$ which represents the index to the selected system. Then the overall state of the system can be redefined with K being a random index as

$$\sigma^{RLMK} = \sum_{k,l,m,} \mathrm{Tr}_{R^{[n]\setminus k} A^n} \left\{ (\mathrm{id} \otimes \Upsilon_l^{(m)})(\psi_\rho^{RA})^{\otimes n} \right\} \otimes \frac{1}{|\mathcal{M}|} |m\rangle\langle m| \otimes |l\rangle\langle l| \otimes \frac{1}{n} |k\rangle\langle k| \,. \tag{2.4.3}$$

Also, the last equality holds because the reference and the index state are independent, i.e., $I(R;K)_\sigma = 0$. This can be easily verified by tracing out other terms in (2.4.3):

$$\begin{aligned}
\sigma^{RK} &= \mathrm{Tr}_{LMX} \left\{ \sigma^{RLMK} \right\} \\
&= \sum_{k,l,m} \frac{1}{|\mathcal{M}|} \mathrm{Tr}_{R^{[n]\setminus k} A^n} \left\{ (\mathrm{id} \otimes \Upsilon_l^{(m)})(\psi_\rho^{RA})^{\otimes n} \right\} \otimes \frac{1}{n} |k\rangle\langle k| \\
&= \sum_{k,l,m} \frac{1}{|\mathcal{M}|} \mathrm{Tr}_{R^{[n]\setminus k}} \left\{ \sqrt{\rho^{\otimes n}} \Upsilon_l^{(m)} \sqrt{\rho^{\otimes n}} \right\} \otimes \frac{1}{n} |k\rangle\langle k| \\
&= \sum_{k} \mathrm{Tr}_{R^{[n]\setminus k}} \left\{ \rho^{\otimes n} \right\} \otimes \frac{1}{n} |k\rangle\langle k| = \rho \otimes \left(\frac{1}{n} \sum_k |k\rangle\langle k| \right),
\end{aligned}$$

which proves that this is a product state. Thus, R, K are independent.

For the second bound, the decoder is a classical channel, therefore, the following inequalities hold for the classical entropy and Shannon's mutual information:

$$\begin{aligned} n(R+R_c) \geq H(LM) &\geq I(LM;X^n) = H(X^n) - H(X^n|LM) \\ &\geq \sum_k H(X_k) - H(X_k|LM) \\ &= \sum_k I(X_k;LM) = nI(X_K;LM|K) \\ &\geq nI(X_K;LM|K) + nI(X_K;K) \\ &= nI(LMK;X_K). \end{aligned} \tag{2.4.4}$$

The arguments for the above inequalities are the same as before. However, here we simply have $I(X_K;K) = 0$ as X^n are exactly IID and given by the constraint of the problem.

Therefore, we can define $W := (L, M, K)$ and observe that a quantum Markov chain of the form $R - (L, M, K) - X$ exists. Also, the single-letter decoder is given by $P_{X|W}(x|w) := P_{X_K|LMK}(x|l, m, k)$, and the encoder is the below set of measurements $\mathcal{M}_w$ on a state ξ^A:

$$\xi^A \xrightarrow{\mathcal{M}_w} \frac{1}{n|\mathcal{M}|} \mathrm{Tr}_{A_1^{k-1}AA_{k+1}^n} \left\{ (\mathrm{id} \otimes \Upsilon_l^{(m)})^{A^n} (\psi_\rho^{RA})^{\otimes k-1} \otimes \xi^A \otimes (\psi_\rho^{RA})^{\otimes n-k} \right\}.$$

2.4.2 Distortion Constraint

For the proof of satisfying the distortion constraint, we utilize a continuity analysis similar to [45]. According to Lemma 2.2, the minimum achievable rate described by the OC rate-distortion function, is a convex function of distortion D. Therefore, it is continuous in its domain.

Lemma 2.2. *$R(D; R_c, \rho||Q_X)$ is a convex function of D on $0 < D < \infty$.*

Proof. See Appendix C.1. □

This ϵ-continuity implies that for any (R, R_c) in the interior of the rate region; i.e. $R > R(D; R_c, \rho||Q_X)$, there exists an $\epsilon > 0$ that still satisfies $R > R(D - \epsilon; R_c, \rho||Q_X)$. As a result, having (2.2.3), there exists an (n, R, R_c) coding scheme which satisfies

$$X^n \sim Q_X^n, \quad d_n(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) \leq D. \tag{2.4.5}$$

Using the above tight distortion bound, we provide the following bound on the single-letter distortion which completes the proof for the converse of the theorem 2.1:

$$\begin{aligned} d(\rho, \Delta_{RX}) &= \mathbb{E}_X \left[\mathrm{Tr}\{\rho_X \Delta_R(X)\}\right] \\ &= \mathbb{E}_K \left[\mathbb{E}_X \left[\mathrm{Tr}\{\rho_X \Delta_R(X)\}|K\right]\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_X \left[\mathrm{Tr}\{\rho_X \Delta_R(X)\}|K=k\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_{X_k} \left[\mathrm{Tr}\{\rho_{X_k} \Delta_{R_k}(X_k)\}\right] \\ &= \mathbb{E}_{X^n} \left[\mathrm{Tr}\left\{\rho_{X^n} \Delta^{(n)}(X^n)\right\}\right] \leq D. \end{aligned} \tag{2.4.6}$$

2.5 Cardinality Bound

In this section, to provide the Cardinality bound, we first introduce a Quantum-Classical version of the original support lemma in [21]:

Lemma 2.3. ***(Quantum-Classical Support Lemma)*** *Let $\mathcal{P} \subseteq \mathcal{D}(\mathcal{H}_A)$ be a compact and connected subset of the set of density operators $\mathcal{D}(\mathcal{H}_A)$ in the finite-dimensional Hilbert space $\mathcal{H}_A$. Also let, $\mathcal{W}$ be an arbitrary set. Further, let $\rho_w \in \mathcal{P}$ indexed by $w \in \mathcal{W}$ be a collection of (conditional) density operators. Suppose having $g_j(\rho), j = 1, ..., J$ a group of real-valued continuous functions (observables) of $\rho \in \mathcal{P}$ and having $\Pi_k \in \mathcal{O}(\mathcal{H}_A), k = 1, ..., K$ a group of projections in Hilbert space $\mathcal{H}_A$ each with rank $\mathrm{Tr}\{\Pi_k\} = d_K \leq \dim(\mathcal{H}_A)$. Then for every $W \sim F(w)$ defined on $\mathcal{W}$, there exist a random variable $W' \sim p(w')$ with $|W'| \leq J + \sum_k (d_k^2 - 1)$ and a collection of conditional density operators $\tilde{\rho}_{w'} \in \mathcal{P}$ indexed by $w' \in \mathcal{W}'$ such that:*

$$\int_{\mathcal{W}} g_j(\rho_w) dF(w) = \sum_{w' \in \mathcal{W}'} g_j(\tilde{\rho}_{w'}) p(w'), \quad \textit{for } j = 1, ..., J \tag{2.5.1}$$

$$\int_{\mathcal{W}} \Pi_k \rho_w \Pi_k dF(w) = \sum_{w' \in \mathcal{W}'} \Pi_k \tilde{\rho}_{w'} \Pi_k p(w'), \quad \textit{for } k = 1, ..., K \tag{2.5.2}$$

Proof. First note that the set of density operators $\mathcal{D}(\mathcal{H}_A)$ is compact and connected. It is compact because it is defined over a complex space that is closed and bounded. Further, it is connected because it is a convex set.

Next, a density operator $\rho \in \mathcal{D}(\mathcal{H}_A)$ can be expressed on an arbitrary orthonormal basis $\{|i\rangle\}_{i=1:d}$ as

$$\rho = \sum_{i,j}^{d} \alpha_{i,j} |i\rangle\langle j| , \tag{2.5.3}$$

where d is the rank of the density operator. We further have $\alpha_{i,j} \in \mathbb{C}, i \neq j$, and $\alpha_{i,i} \in \mathbb{R}, i = 1, ..., d$, and

$$\alpha_{i,j} = \alpha^*_{j,i}, \quad \text{for } i \neq j, \tag{2.5.4}$$

$$0 \leq \alpha_{i,i} \leq 1, \quad \text{for } i = 1, ..., d, \tag{2.5.5}$$

$$\sum_i \alpha_{i,i} = 1. \tag{2.5.6}$$

Thus, any density operator with rank d can be interpreted as a point in a real vector space with $d^2 - 1$ dimensions (Also in [54]). Then the lemma follows directly from the Fenchel-Eggleston-Caratheodory Theorem. □

Assume having a probability distribution $P_W(w)$ for the intermediate variable, we define a function $f : \mathbb{R} \to \mathbb{R}^5$:

$$f : P_W(w) \longrightarrow \Big(\rho^A,\ Q_X(x),\ I(X;W),\ I(R;W),\ \mathrm{Tr}\big\{\Delta_{RX}\tau^{RX}\big\}\Big).$$

We then find the number of affine functions required to implement the above function with a convex combination of conditionals on W,

$$f_{\text{affine}} : P_W(w) \longrightarrow \Big(\rho^A, Q_X(x), H(X|W), H(R|W), \mathrm{Tr}\big\{\Delta_{RX}\tau^{RX}\big\}\Big).$$

The first condition ρ^A requires at most $(\dim \mathcal{H}_A)^2 - 1$ affine functions as discussed in lemma 2.3. The two classical states have simple classical conditional representations as follows. Note that there is at most need for $|\mathcal{X}| - 1$ separate functions to represent the distribution Q_X and one for the conditional entropy:

$$Q_X(x_i) = \sum_w Q_{X|W}(x_i|w)P_W(w), \qquad \forall i \in [|\mathcal{X}|-1], x_i \in \mathcal{X} \tag{2.5.7}$$

$$H(X|W) = \sum_w H(X|W=w)P_W(w). \tag{2.5.8}$$

Also, the other functions have the following representations,

$$\begin{aligned} H(R|W) &= \sum_w H(R|W=w)P_W(w) \\ &= \sum_w H(\rho_w)P_W(w), \end{aligned} \tag{2.5.9}$$

$$\begin{aligned} \mathrm{Tr}\{\Delta_{RX}\tau^{RX}\} &= \mathrm{Tr}\left\{\Delta_{RX}\left(\sum_{w,x}\mathrm{Tr}_A\{(\mathrm{id}_R\otimes M_w)\psi_\rho^{RA}\}\otimes P_{X|W}(x|w)\,|x\rangle\langle x|^X\right)\right\} \\ &= \mathrm{Tr}\left\{\Delta_{RX}\left(\sum_{w,x}\rho_w^A\,P_W(w)\otimes P_{X|W}(x|w)\,|x\rangle\langle x|^X\right)\right\} \\ &= \sum_w P_W(w)\,\mathrm{Tr}\left\{\Delta_{RX}\left(\sum_x \rho_w^A\otimes P_{X|W}(x|w)\,|x\rangle\langle x|^X\right)\right\} \\ &= \sum_w \mathrm{Tr}\{\Delta_{RX}\tau^{RX}|W=w\}P_W(w). \end{aligned} \tag{2.5.10}$$

Therefore, using the QC support lemma 2.3, we claim that there exists a random variable W' with cardinality at most $|\mathcal{W}'| \leq (\dim\mathcal{H}_A)^2+|\mathcal{X}|+1$, with the composite state $\nu^{RW'X}$ which forms the following quantum Markov chain $R-W'-X$ with the same marginals and satisfies the entropic and distortion equalities,

$$\nu^X = \tau^X \equiv Q_X,\;\; \nu^R = \tau^R \equiv \rho,\;\; \mathrm{Tr}\{\Delta_{RX}\nu^{RX}\} = \mathrm{Tr}\{\Delta_{RX}\tau^{RX}\}, \tag{2.5.11}$$

$$I(X;W')_\nu = I(X;W)_\tau,\;\; I(R;W')_\nu = I(R;W)_\tau. \tag{2.5.12}$$

Chapter 3

Continuous-Variable Quantum System

3.1 Generalized Definitions of Continuous Quantum Systems

In this section, we investigate the measurement coding for the Bosonic continuous-variable quantum systems [Chapters 11 ,12 of [29]]. The proof of the achievability of the random coding argument in previous sections does not directly apply to the continuous quantum systems. The first reason is that the Chernoff bound [38] (which is the main theorem for the validity of the measurement POVMs in random coding argument) is available only for a finite-dimensional Hilbert space.

Secondly, in infinite-dimensional systems, it is not possible to represent the outcome space using quantum registers defined on separable Hilbert spaces. This is in contrast to the finite-dimensional system for which, the set of all outcome states forms a complete orthonormal set. As a result, the quantum mutual information is not defined for such continuous measurement systems. Instead, we keep the output system as classical and use the generalized ensemble representation. In order to properly define the continuous system model, we first provide the following generalized definitions.

Definition 3.1 ([29] Definition 11.22)**.** *The generalized ensemble is defined as a Borel probability measure π on the subspace of density operators $\mathcal{G}(\mathcal{H}_A)$. Then the average state of the ensemble is defined as*

$$\bar{\psi}_\pi = \int \psi\pi(d\psi). \tag{3.1.1}$$

In contrast to the finite Hilbert space for which the POVM is defined for all possible outcomes, in the continuous quantum measurement systems, the generalized POVM is defined over the subset of σ-algebra of Borel subsets.

Definition 3.2 ([29] Definition 11.29.)**.** *A POVM is generally defined on a measurable space X with a σ-algebra of measurable subsets $\mathcal{B}$, as a set of Hermitian operators $M = \{M(B), B \in \mathcal{B}\}$ satisfying the following conditions:*

1. $M(B) \geq 0, B \in \mathcal{B}$,
2. $M(\mathcal{X}) = I$,
3. *For any countable (not necessarily finite) decomposition of mutually exclusive subsets $B = \cup B_j$ ($B_i \cap B_j = \emptyset,\ i \neq j$), the sum of the measures converge in weak operator sense to measure of the combined set; i.e. $M(B) = \sum_j M(B_j)$.*

Then having an observable POVM M acting on a state ψ, with outcomes in measurable space $\mathcal{X}$, results in the following probability measure

$$\mu_\rho^M(B) = \mathrm{Tr}\{\rho M(B)\}, \quad B \in \mathcal{B}. \tag{3.1.2}$$

It is also necessary to have a proper definition of post-measurement states. The a posteriori average density operator for a subset $B \in \mathcal{B}$ is defined in [41] for a general POVM M as

$$\rho_B = \frac{\sqrt{M(B)}\rho\sqrt{M(B)}}{\mathrm{Tr}\{\rho M(B)\}}. \tag{3.1.3}$$

Based on that, Ozawa defines the post-measurement state for a continuous quantum system, as given in the following theorem.

Theorem 3.1 (Theorem 3.1. [41]). *For any observable M and input density operator ρ, there exists a family of a posteriori density operators $\{\rho_x; x \in \mathbb{R}\}$, defined with the following properties:*

1. *For any $x \in \mathbb{R}$, ρ_x is a density operator in $\mathcal{H}_A$;*
2. *The mapping $x \to \rho_x$ is strongly Borel measurable;*
3. *For any arbitrary observable N, and any Borel sets A and B, the joint probability*

$$\begin{aligned} P(X \in B, Y \in A) &= \mathrm{Tr}\Big\{\sqrt{M(B)}\rho\sqrt{M(B)}N(A)\Big\} \\ &= \int_B \mathrm{Tr}\{\rho_x N(A)\}\mu(dx), \end{aligned}$$

where $\mu(B) = \mathrm{Tr}\{M(B)\rho\},\ B \in \mathcal{B}$ is the probability measure of the outcome space.

Compare this theorem to [[30], Section IV], which defines the PMR ensemble as

$$\mathcal{E}' : \qquad \pi'(B) = \mathrm{Tr}\{\rho M(B)\}, \quad \rho'_y = \frac{\rho^{1/2} m(y) \rho^{1/2}}{\mathrm{Tr}\{\rho m(y)\}}, \quad M(B) = \int_B m(y)\mu(dy),$$

where $m(y)$ is a weakly measurable function with values in the cone of bounded positive operators of $\mathcal{H}$, expressing the Radon-Nikodym derivative of the M POVM. This simpler representation only applies to the POVMs that have this well-defined $m(y)$, and is not necessarily available for the general POVM. However, the post-measured state in its general form ρ_x is available for the general POVM as described by Ozawa's theorem.

The above theorem provides the necessary requirements to define the proper information quantity. Thus, using the post-measured state ensemble one can define the information gain introduced by Groenwold [27] for an input state ρ and output ensemble $\{\rho(B), \mu_\rho^M(B)\}_{B\in\mathcal{B}}$ as [29]:

$$I_g(\rho, X) = H(\rho) - \int_{\mathcal{X}} H(\rho_x)\mu_\rho^M(dx). \tag{3.1.4}$$

It is worth mentioning that the information gain equals the quantum mutual information in finite-dimensional measurement systems.

3.2 Achievable Rate Region for Continuous-Variable Quantum Systems

We first redefine the Definitions 2.2 and 2.3 to match with the continuous quantum system. The definition of achievability for a continuous quantum source coding scheme is as follows:

Definition 3.3. *An* (n, R, R_c) *source-coding scheme for the continuous quantum-classical system is comprised of an encoder* $\mathcal{E}_n$ *on Alice's side and a decoder* $\mathcal{D}_n$ *on Bob's side, with the detailed description provided in the Definition 2.2. The final output sequence* X^n *is generated by the decoder in the output space* $\mathcal{X}^n$ *with the probability measure* $\{P_{X^n}(B),\ B \in \mathcal{B}(\mathcal{X}^n)\}$*. Thus, the average PMR state and its corresponding Borel subset of the output sample space, form an ensemble of the form* $\{\hat{\rho}_B^{R^n}, B \in \mathcal{B}(\mathcal{X}^n)\}$*, where*

$$\hat{\rho}_B^{R^n} = \frac{1}{P_{X^n}(B)} \sum_{m,l} \frac{1}{|\mathcal{M}|} \operatorname{Tr}_{A^n} \left\{ (id \otimes \Upsilon_l^{(m)})[\psi_{RA}^{\rho}]^{\otimes n}) \right\} \mathcal{D}_n(B|l,m), \tag{3.2.1}$$

$$P_{X^n}(B) = \sum_{m,l} \frac{1}{|\mathcal{M}|} \operatorname{Tr}\left\{ \Upsilon_l^{(m)} \rho^{\otimes n} \right\} \mathcal{D}_n(B|l,m). \tag{3.2.2}$$

We define the average n-letter distortion for the source coding system with encoder /decoder pair $\mathcal{E}_n$ and $\mathcal{D}_n$, and set of distortion observable operators $\Delta(x), x \in \mathcal{X}$ and continuous memoryless source state $\rho^{\otimes n}$ as

$$d_n(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{X_i} \left[\operatorname{Tr}\left\{ \hat{\rho}_{X_i}^{R_i} \Delta(X_i) \right\} \right], \tag{3.2.3}$$

where $\hat{\rho}_{x_i}^{R_i} := \mathbb{E}_{x^{n\setminus[i]}} \left[\operatorname{Tr}_{n\setminus[i]} \left\{ \hat{\rho}_{X^n}^{R^n} \right\} \right]$ is the i-th local state of the PMR density operator in (3.2.1) defined for $x^n \in \mathcal{X}^n$ as described by Theorem 3.1. Consequently, the following definition of achievability is used throughout this paper.

Definition 3.4. *Given a probability measure* μ_x *on* $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ *and a distortion level* D*, and assuming a product input state of* $\rho^{\otimes n}$ *of continuous Hilbert space, a rate pair* (R, R_c) *is defined as achievable if, for any sufficiently large* n *and any positive value* $\epsilon > 0$*, there exists an* (n, R, R_c) *coding scheme comprising of a measurement encoder* $\mathcal{E}_n$ *and a decoder*

$\mathcal{D}_n$ as described in Definition 3.3 that satisfy the following conditions as defined by the corresponding distortion measure of (3.2.3)

$$X^n \sim \mu_X^n, \quad d_n(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) \leq D + \epsilon. \tag{3.2.4}$$

Then the closure of the achievable rate region for the continuous quantum system is expressed with single-letter characterizations by the following theorem:

Theorem 3.2. *Given a pair (μ_X, D) and having a product input state $\rho^{\otimes n}$ of continuous infinite-dimensional Hilbert space with limited von Neumann entropy, a rate pair (R, R_c) is inside the achievable rate region in accordance with the definition 3.4 if and only if there exists an intermediate state W with a corresponding measurement POVM $M = \{M^A(B), B \in \mathcal{B}_\mathcal{W}\}$ where $\mathcal{B}_\mathcal{W}$ is the σ-algebra of the Borel sets of $\mathcal{W}$, and randomized post-processing transformation $P_{X|W}$ which satisfies the rate inequalities*

$$R \geq I_g(W; R), \tag{3.2.5}$$

$$R + R_c \geq I(W; X), \tag{3.2.6}$$

where W, constructs a quantum Markov chain $R - W - X$ which generates the ensemble $E_w := \{\rho_w, w \in \mathcal{W}\}$ with the PMR states and the intermediate space, and generates the ensemble $E_x := \{\rho_x, x \in \mathcal{X}\}$ with the PMR states and the output space, from the set

$$\mathcal{M}_c(\mathcal{D}) = \left\{ (E_w, E_x) \middle| \begin{array}{l} \int_w P_{X|W}(A|w) \operatorname{Tr}\{M(dw)\rho\} = \mu_X(A) \quad \textit{for } A \in \mathcal{B}(\mathcal{X}) \\ \int_{x \in \mathbb{R}} \operatorname{Tr}_R \left[\rho_x \, \Delta_R(x)\right] \mu_X(dx) \leq D \end{array} \right\}. \tag{3.2.7}$$

Note that the cardinality bound does not exist in the continuous case as we have $\mathcal{W} \equiv \mathcal{X} \equiv \mathbb{R}$.

3.3 Proof of Achievability in the Continuous System

Given the pair (μ_X, D), assume there exists a continuous intermediate state W forming a quantum Markov chain $R - W - X$, with a corresponding continuous POVM $M = \{M(B), B \in \mathcal{B}(\mathcal{X})\}$ with outcomes in $\mathcal{W}$ space, defined as a set of Hermitian operators in $\mathcal{H}_A$ satisfying conditions of Theorem 3.2. Moreover, a corresponding classical post-processing channel $P_{X|W} : \mathcal{W} \times \sigma(\mathcal{X}) \to \mathbb{R}$ which is a mapping such that for every $w \in \mathcal{W}$, $P_{X|W}(.|w)$ is a probability measure on $\mathcal{B}(\mathcal{X})$ and for every $B \in \mathcal{B}(\mathcal{X})$, $P_{X|W}(B|.)$ is a Borel-measurable function. Then according to theorem 3.1, for any proper measurement POVM M, there exists a family of post-measurement density operators ρ_w along with a probability measure μ_ρ^M (or equivalently, ρ_x along with μ_X if we consider the overall POVM $\Lambda(B) \equiv \int_{\mathcal{W}} M(dw)P(B|w),\ B \in \mathcal{B}(\mathcal{X})$), such that

$$\int_w P_{X|W}(A|w)\mu_\rho^M(dw) = \mu_X(A), \quad \text{for all } A \subseteq \mathcal{B}(\mathcal{X}), \tag{3.3.1}$$

$$d(R, X) := \int_x \mathrm{Tr}\{\rho_x \Delta_R(x)\}\mu_X(dx) \leq D. \tag{3.3.2}$$

The continuous quantum system can be represented by $\{|n\rangle\}_{n=0}^{\infty}$, the number operators of the Fock basis, which is a countable infinite dimensional Hilbert space. We plan to use the source coding theorem of the discrete system from the previous section. To use Theorem 2.1, we first follow a similar approach as [23, 36, 44] to perform a clipping projection, that truncates the state into the finite-dimensional space of the first $k_1 + 1$ Fock states via an energy test:

$$C_{k_1} \equiv \left\{ \Pi_{k_1} := \sum_{n=0}^{k_1} |n\rangle\langle n| \,, \quad I - \Pi_{k_1} \right\}. \tag{3.3.3}$$

Therefore, for any small $\epsilon_c > 0$, there exists a large enough k_1 such that the probability of the state projecting to the first subspace is ϵ_c-close to unity, $\mathrm{Tr}\{\rho\Pi_{k_1}\} \geq 1 - \epsilon_c$. For a detailed description of the spectral decomposition of continuous-variable quantum systems, refer to [32, 50].

3.3.1 Information Processing Task

In the classical systems, to generate a discrete coding scheme from the continuous distributions, one simply extends the Markov chain $R-W-X$ to discrete variables by quantizing the input and output variables as $R_{k_1} - R - W - X - X_{k_2}$ where $k_1, k_2 \in \mathbb{N}$ are the quantization parameters forming 2^{k_i} levels within the $[-k_i, k_i]$ region [45].

In our QC system, because the output is classical, we can extend the original Markov chain by quantizing output X with Q_{K_2} where $K_2 = (k_2, k_2')$ is the pair of clipping region parameters creating the cut-off range $[-k_2, k_2]$ and k_2' is the precision parameter making $2^{k_2'}$ levels forming the quantized output space $\mathcal{X}_{K_2}$, such that

$$R \xrightarrow{M_w} W \xrightarrow{P_{X|W}} X \xrightarrow{Q_{k_2}} X_{K_2}. \tag{3.3.4}$$

However, for the source state, performing the inverse of measurement is not straightforward, as the quantum state collapses after measurement. Therefore, one cannot

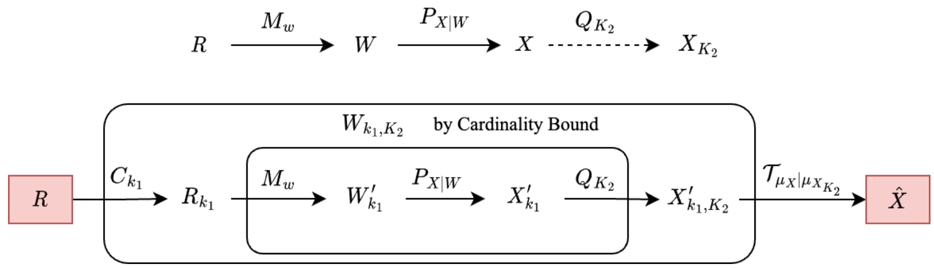

Figure 3.1: Markov Chain of the Alternative Approach: The upper diagram shows the original Markov chain provided by the single-letter intermediate state W. The lower diagram shows the single-letter Markov chain of the alternative approach in which the clipped source state is directly fed into the same continuous measurement M_w and the discrete output is obtained by quantizing X'_{k_1}. Finally, the optimal transport transforms the discrete output back to the continuous output $\hat{X}$.

simply apply the inverse of clipping projection to the projected state R_{k_1} to retrieve the state in R. This means we cannot extend the original Markov chain for the source state in this case. Instead, in an alternative approach, we prepare a separate Markov chain by clipping the quantum source state with clipping projection in (3.3.3) and then directly feeding that clipped state R_{k_1} into the same continuous measurement POVM M_w, as shown in figure 3.1. This is specifically possible as the clipped input state lies in a subspace of the same quantum Hilbert space.

$$R \xrightarrow{C_{k_1}} R_{k_1} \xrightarrow{M_w} W'_{k_1} \xrightarrow{P_{X|W}} X'_{k_1} \xrightarrow{Q_{k_2}} X'_{k_1,K_2}. \tag{3.3.5}$$

3.3.2 Proof of Rate Inequalities

In this approach, the data processing inequality does not directly apply to the system. Therefore we provide the following lemma to show the rate inequalities still hold after clipping.

Lemma 3.3. *Suppose having a Quantum Markov chain of the form $R-W-X$ satisfying the conditions in Theorem 3.2. Then by using the alternative clipping method as described by* (3.3.5)*, the clipped states still satisfy the following rate inequalities in the asymptotic regime*

$$\lim_{k_1\to\infty} I_g(R_{k_1}; W'_{k_1}) \leq I_g(R; W), \tag{3.3.6}$$

$$\lim_{k_1\to\infty} I(W'_{k_1}; X'_{k_1,K_2}) \leq I(W; X). \tag{3.3.7}$$

For the first rate inequality we directly use the following Proposition [Proposition 6 of [47] by Shirokov]:

Proposition 3.4. *Let $\{\rho^{(n)^A}\}$ be a sequence of states converging to a state $\rho^{(0)^A}$ (See Section 11.1 of [29]). Also, let $\{M_n\}$ be any arbitrary sequence of POVMs weakly converging to a POVM M_0, with the outcome space $\mathcal{W}$. If either $|\mathcal{W}| < +\infty$ or $\lim_{n\to\infty} H(\rho^{(n)^A}) = H(\rho^{(0)^A}) < +\infty$ then*

$$\lim_{n\to\infty} I_g(M_n, \rho^{(n)^A}) = I_g(M_0, \rho^{(0)^A}), \tag{3.3.8}$$

where $I_g(M, \rho)$ is the information gain as defined by Groenwold.

To use this proposition in our system, we build a refined POVM by combining the clipping projection and the M measurement POVM, in the following way. We first apply the clipping projection onto the input state ρ_A. If it is not in the projected subspace (with probability $\mathrm{Tr}\{(I-\Pi_{k_1})\rho_A\}$), we throw out the state and only notify the receiver by asserting an error bit A_{k_1}. Otherwise, if the input state is inside the clipping subspace (with probability $\mathrm{Tr}\{\Pi_{k_1}\rho_A\}$), the input is sent to the

M measurement POVM which produces the classical outcome. In this case, the post-collapsed state after clipping projection is

$$\hat{\rho}^{A'} = \frac{\Pi_{k_1}\rho_A\Pi_{k_1}}{\mathrm{Tr}\{\Pi_{k_1}\rho_A\}},$$

and the final a-posteriori average density operator for an event $B \in \mathcal{B}$ is obtained as

$$\hat{\rho}_B^{A''} = \frac{\sqrt{M(B)}\Pi_{k_1}\rho_A\Pi_{k_1}\sqrt{M(B)}}{\mathrm{Tr}\{M(B)\Pi_{k_1}\rho_A\Pi_{k_1}\}}.$$

Then the combination of the two above POVMs can be expressed in the following refined POVM

$$\hat{M}_{k_1} = \{\{M(B)\Pi_{k_1}, B \in \mathcal{B}\}, I - \Pi_{k_1}\}.$$

It can be shown that the above refined POVM is a valid POVM. We next provide the following lemma regarding the weak convergence of the refined POVM.

Lemma 3.5. *The sequence of $\hat{M}_{k_1}$ POVMs converge weakly to M continuous POVM.*

Proof. See Appendix C.2. □

Therefore, by applying the Proposition 3.4, it follows that

$$\lim_{k_1\to\infty} I(R_{k_1}, W'_{k_1}) = I(R; W). \tag{3.3.9}$$

Next, we prove the second rate-inequality (3.3.7). Recall that the clipping error indicator variable A_{k_1} is defined as the event that the state is not inside the clipping subspace. By applying the chain rule for mutual information we have

$$I(W, A_{k_1}; X_{K_2}) = I(W; X_{K_2}) + I(A_{k_1}; X_{K_2}|W) \leq H(A_{k_1}) + I(W; X_{K_2}),$$

which results in,

$$I(W, A_{k_1}; X_{K_2}) - H(A_{k_1}) \leq I(W; X_{K_2}) \leq I(W, A_{k_1}; X_{K_2}). \tag{3.3.10}$$

Therefore, as $H(A_{k_1})$ decays to zero when $k_1 \to \infty$, from squeeze theorem we have the following limit

$$\lim_{k_1 \to \infty} I(W, A_{k_1}; X_{K_2}) = I(W; X_{K_2}). \tag{3.3.11}$$

Further, using another chain rule we get

$$\begin{aligned} I(W, A_{k_1}; X_{K_2}) &= I(A_{k_1}; X_{K_2}) + I(W; X_{K_2}|A_{k_1}) \\ &= I(A_{k_1}; X_{K_2}) + P(A_{k_1} = 0).I(W; X_{K_2}|A_{k_1} = 0) \\ &\qquad + P(A_{k_1} = 1).I(W; X_{K_2}|A_{k_1} = 1). \end{aligned} \tag{3.3.12}$$

As the clipping region grows to infinity with $k_1 \to \infty$, the probability of clipping decays to zero $\lim_{k_1 \to \infty} P(A_{k_1} = 0) = 1$. Therefore, because A_{k_1} is a simple Bernoulli random variable with an arbitrarily small probability, then $I(A_{k_1}; X_{K_2}) \leq H(A_{k_1}) \to 0$.

Also for the third term, we can find the following asymptotic bound

$$\lim_{k_1\to\infty} P(A_{k_1}=1).I(W;X_{K_2}|A_{k_1}=1) \leq \lim_{k_1\to\infty} P(A_{k_1}=1).H(X_{K_2}) = 0, \tag{3.3.13}$$

where we appealed to the fact that quantized output with limited alphabet has limited entropy.

Finally, consider that $I(W;X_{K_2}|A_{k_1}=0) = I(W'_{k_1}, X'_{k_1,K_2})$ holds by definition because when the input state is inside the clipping subspace, the system behaves as if no clipping was performed. Combining all together, for any fixed K_2 we have

$$\lim_{k_1\to\infty} I(W, A_{k_1}; X_{K_2}) = \lim_{k_1\to\infty} I(W'_{k_1}, X'_{k_1,K_2}). \tag{3.3.14}$$

Then (3.3.11) and (3.3.14) together show that for any fixed K_2 we have

$$\lim_{k_1\to\infty} I(W'_{k_1}, X'_{k_1,K_2}) = I(W, X_{K_2}) \leq I(W;X), \tag{3.3.15}$$

where the inequality follows directly from the Data Processing Inequality and completes the proof. In addition, note that as $k_2, k'_2 \to \infty$, X_{K_2} converges weakly to X. Therefore, using lower semi-continuity of mutual information [42, 43], combined with the above data-processing inequality we further have

$$\lim_{k_2,k'_2\to\infty} \lim_{k_1\to\infty} I(W'_{k_1}, X'_{k_1,K_2}) = \lim_{k_2,k'_2\to\infty} I(W, X_{K_2}) = I(W;X). \tag{3.3.16}$$

3.3.3 Source-Coding Protocol for Continuous States

Having a quantum source generating a sequence of n independent continuous states as $\rho^{\otimes n}$, we apply a coding protocol on the source states, described in this section. We first separate the input states into proper and improper states by applying the clipping POVM Π_{k_1} which generates a sequence of error bits $A_{k_1}^n \equiv \{A_{i,k_1}\}_{i=1}^n$ defined as

$$A_{i,k_1} := \begin{cases} 0 & \text{if } \rho_i \in \Pi_{k_1} \\ 1 & \text{O.W.} \end{cases} . \tag{3.3.17}$$

Thus, according to WLLN, for any fixed $\epsilon_{cl} > 0$ and $k_1 \in \mathbb{N}$, there exists a value $N_0(\epsilon_{cl}, k_1)$ large enough such that for any $n \geq N_0(\epsilon_{cl}, k_1)$, the number of proper states (states in the clipping subspace)

$$T := \sum_{i=1}^{n} A_{i,k_1}, \tag{3.3.18}$$

is within the range $T \in [n(1 - P_{k_1} \pm \epsilon_{cl})]$ with probability no less than $1 - \epsilon_{cl}$. Then for any sequence with $T < t_{\text{min}}$ where $t_{\text{min}} := n(1 - P_{k_1} - \epsilon_{cl})$, we do not perform source coding, and instead assert a *source coding error* event E_{ce}. Upon receiving the coding error event, Bob will locally generate a sequence of random outcomes $\hat{X}_{\text{local}}^n$ with the desired μ_X^n output distribution. This ensures that in every sequence for which the coding is performed, there are $T > t_{\text{min}}$ independent source states which are inside the clipping region, for which we perform the coding scheme. For the rest of the $n - T$ states, we do not perform the coding, instead, we throw away the source state and send the error-index to the receiver.

Next, the error bits sequence is coded into indices of size $\binom{n}{t_{\text{min}}} + 1$ where the extra index is the event of a source coding error E_{ce}. Note here that the required classical rate, in this case, will be $R + \log_2 \left(\binom{n}{t_{\text{min}}} + 1 \right)$. The following limit

$$\lim_{\epsilon_{cl} \to 0} \lim_{k_1 \to \infty} \lim_{n \to \infty} R + \frac{1}{n} \log_2 \left(\binom{n}{n(1 - P_{k_1} - \epsilon_{cl})} + 1 \right) = R, \tag{3.3.19}$$

ensures that the extra error handling rate can be made arbitrarily small.

Then at Bob's side, the classical sequence $X'^{t}_{k_1,K_2}$ is constructed using the discrete coding scheme, and is fed to a memoryless optimal transport block $\mathcal{T}_{\mu|\mu_{K_2}}$ to generate the final continuous sequence $\hat{X}^t_{k_1,K_2}$. Finally, the sequence is padded with the $n - T$ locally generated independent values at the error positions to create the final $\hat{X}^n_{k_1,K_2}$. Bob then uses this sequence to prepare his final quantum states. Figure 3.2 shows the block diagram of this coding protocol.

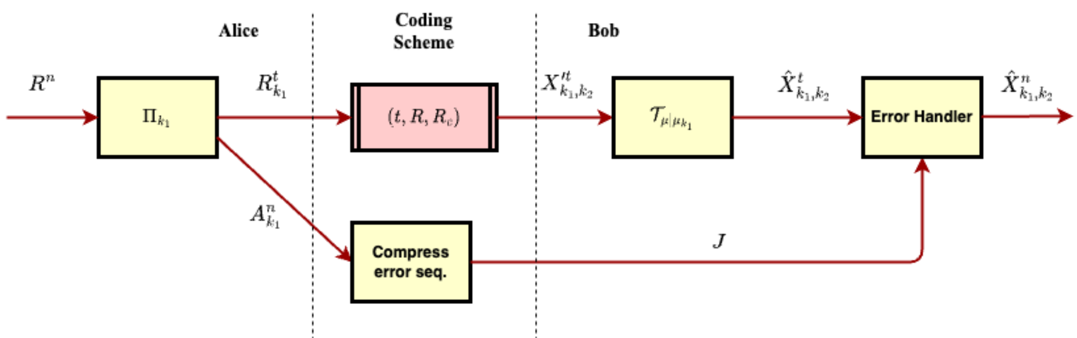

Figure 3.2: Continuous Coding Protocol

3.3.4 Proof of Distortion Constraint

The end-to-end average distortion for the above system is written as

$$\begin{aligned} d_n(R^n, \hat{X}^n_{k_1,K_2}) &= d_n\left(R^n, \hat{X}^n_{k_1,K_2}\middle|E_{ce}\right)P(E_{ce}) + d_n\left(R^n, \hat{X}^n_{k_1,K_2}\middle|\neg E_{ce}\right)(1-P(E_{ce})) \\ &\leq d_n\left(R^n, \hat{X}^n_{\text{local}}|E_{ce}\right)\epsilon_{cl} + d_n\left(R^n, \hat{X}^n_{k_1,K_2}\middle|\neg E_{ce}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} d\left(R_i, \hat{X}_{i,\text{local}}|E_{ce}\right)\epsilon_{cl} + d_n\left(R^n, \hat{X}^n_{k_1,K_2}\middle|\neg E_{ce}\right), \end{aligned} \tag{3.3.20}$$

where $\hat{X}^n_{\text{local}}$ is generated locally at Bob's side according to the fixed IID output distribution μ_X in the event of coding error E_{ce}. Therefore, in the first term above, for each i-th sample of the system, the uniform integrability of the distortion observable implies that it can be made arbitrarily small by selecting the proper value of ϵ_{cl}. As for the second term, we use the following lemma to provide a single-letter upper bound:

Lemma 3.6. *The end-to-end average n-letter distortion of the continuous system conditioned on the event of no coding error is asymptotically upper-bounded by the following single-letter distortion for any fixed value of $k_1, k_2, k_2' > 0$ as $n, t_{min} \to \infty$:*

$$\begin{aligned} \lim_{\substack{n\to\infty,\\ t_{min}\to\infty}} d_n(R^n, \hat{X}^n_{k_1,K_2}|\neg E_{ce}) \leq\ & d(R_{k_1}, X'_{k_1,K_2}) \\ &+ (P_{k_1} + \epsilon_{cl})d\left(R, \hat{X}_{local}|A_{k_1} = 1\right) \\ &+ \int_{\mathcal{X}} \mathrm{Tr}\left\{\bar{\bar{\rho}}^R_x\left(\Delta(x) - \Delta(Q_{K_2}(x))\right)\right\}\mu_X(dx), \end{aligned} \tag{3.3.21}$$

where $\bar{\bar{\rho}}_x^R$ is the asymptotic average PMR state, given by

$$\bar{\bar{\rho}}_x^R := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \rho_x^{R_i}.$$

Proof. See Appendix C.3 □

We then take the limit of this single-letter distortion as $k_1, k_2 \to \infty$. Note that as $\lim_{k_1\to\infty} P(A_{k_1} = 1) = 0$, the second term decays to zero by assuming that $\epsilon_{cl} \leq P_{k_1}$, as a result of uniform integrability of the distortion observable. The third term also decays to zero as follows:

$$\begin{aligned}
&\lim_{k_2\to\infty} \lim_{k_2'\to\infty} \int_{\mathcal{X}} \mathrm{Tr}\Big\{\bar{\bar{\rho}}_x^R \Big(\Delta(x) - \Delta(Q_{K_2}(x))\Big)\Big\}\mu_X(dx) \\
&\leq \lim_{k_2\to\infty} \lim_{k_2'\to\infty} \int_{-2^{k_2}}^{2^{k_2}} \|\Delta(x) - \Delta(Q_{K_2}(x))\|_1 \mu_X(dx) \\
&\quad + \lim_{k_2\to\infty} \int_{\mathcal{X}\setminus[-2^{k_2}, 2^{k_2}]} \mathrm{Tr}\Big\{\bar{\bar{\rho}}_x^R \Big(\Delta(x) - \Delta(Q_{K_2}(x))\Big)\Big\}\mu_X(dx) = 0,
\end{aligned} \tag{3.3.22}$$

where we split the integral into the cut-off range and out-of-range intervals and used the Holder's inequality. Then the first term above converges to zero due to the continuity of $\Delta(x)$ operator and the second term converges to zero by the definition of uniform integrability.

Next, consider that the following inequality holds by definition for the single-letter discrete distortion:

$$\begin{aligned}
d(R, X_{K_2}) &= d(R, X_{K_2}|A_{k_1} = 0)\,\mathrm{Pr}(A_{k_1} = 0) + d(R, X_{K_2}|A_{k_1} = 1)\,\mathrm{Pr}(A_{k_1} = 1) \\
&= d(R_{k_1}, X'_{k_1,K_2})\,\mathrm{Pr}(A_{k_1} = 0) + d(R, X_{K_2}|A_{k_1} = 1)\,\mathrm{Pr}(A_{k_1} = 1).
\end{aligned} \tag{3.3.23}$$

Then, by having the probability of clipping approach zero $\lim_{k_1\to\infty}\Pr(A_{k_1}=1)=0$, the second term above goes to zero as a direct result of uniform integrability, and we have the following asymptotic limit:

$$\lim_{k_1\to\infty} d(R_{k_1}, X'_{k_1,K_2}) = d(R, X_{K_2}) := \int_{\mathcal{X}} \mathrm{Tr}\big\{\sqrt{\rho}\,\Lambda_X(dz)\sqrt{\rho}\,\Delta\big(Q_{K_2}(z)\big)\big\}. \tag{3.3.24}$$

Then as $k_2, k'_2 \to \infty$, we can upper-bound this RHS distortion value by

$$\begin{aligned}\lim_{k_2,k'_2\to\infty} d(R, X_{K_2}) &\le D + \lim_{k_2,k'_2\to\infty} \left(d(R, X_{K_2}) - d(R, X)\right)\\ &= D + \lim_{k_2,k'_2\to\infty} \int_{\mathcal{X}} \mathrm{Tr}\Big\{\sqrt{\rho}\,\Lambda_X(dz)\sqrt{\rho}\,\Big(\Delta\big(Q_{K_2}(z)\big) - \Delta(z)\Big)\Big\} = D,\end{aligned} \tag{3.3.25}$$

where the last equality follows similarly from continuity and uniform integrability of the distortion observable operator $\Delta(x)$ as a function of $x \in \mathcal{X}$. Combining the above bounds to the single-letter expression in (3.3.21) shows

$$\lim_{\substack{n\to\infty,\\ t_{\min}\to\infty}} d_n(R^n, \hat{X}^n_{k_1,K_2}) \le D, \tag{3.3.26}$$

which completes the proof of achievability.

Chapter 4

Evaluation of the Qubit-Binary System

In this chapter, we study the example of Qubit Systems. Having the Qubit source state ρ and a Bernoulli output distribution Q_X, we aim to find the output-constrained rate-distortion function $R(D; \infty, \rho||Q_X)$ for the case of unlimited common randomness. By inverting this function we then achieve the rate-limited optimal transport cost $D(R; \infty, \rho||Q_X)$ function. We then provide the numerical results of a few example Qubit systems and plot the rate-distortion function.

4.1 Qubit System with Unlimited Common Randomness

For the case of the qubit QC system, we employ entanglement fidelity as the distortion measure, which can be written as

$$\begin{aligned}\mathrm{Tr}\{\Delta_{RX}\tau_{RX}\} &= \mathrm{Tr}\left\{\left(I - |\psi^{RA}\rangle\langle\psi^{RA}|\right)\left(\sum_x \mathrm{Tr}_A\left\{(\mathrm{id}\otimes M_x^A)\psi^{RA}\right\}\otimes|x\rangle\langle x|\right)\right\} \\ &= 1 - \langle\psi^{RA}|\left(\sum_x \sqrt{\rho}M_x\sqrt{\rho}\otimes|x\rangle\langle x|\right)|\psi^{RA}\rangle . \end{aligned} \tag{4.1.1}$$

Using the following spectral decomposition of ρ on the eigenbasis $|\varphi_t\rangle_{t=1}^d$,

$$\rho_A = \sum_{t=1}^{d} P_T(t)\, |\varphi_t\rangle\langle\varphi_t|_A ,$$

and by substituting the canonical purification of the above decomposition into (4.1.1), it simplifies to

$$\mathrm{Tr}\{\Delta_{RX}\tau_{RX}\} = 1 - \sum_x \langle x|\,\rho M_x^{\mathcal{T}_\varphi}\rho\,|x\rangle , \tag{4.1.2}$$

where $M_x^{\mathcal{T}_\varphi}$ is the transpose of M_x with respect to the $\{\varphi_t\}$ basis, defined as

$$M_x^{\mathcal{T}_\varphi} = \sum_{t,s} \langle\varphi_t|M_x|\varphi_s\rangle\, |\varphi_t\rangle\langle\varphi_s| .$$

In the presence of an unlimited amount of common randomness, the only effective rate becomes $I(W;R)$, which is lower-bounded by $I(X;R)$ because of the data processing inequality and the Markov chain $R - W - X$ [45]. Thus $W = X$ minimizes the mutual information, which means no local randomness is required at decoder. Therefore, using the main theorem, for a qubit system with input state ρ and Bernoulli(q_1) output distribution, the output-constrained rate-distortion function is obtained by

$$R(D;\infty,\rho||\text{Bern}(q_1)) = \min_{M_x^A} I(X;R)_\tau,$$
$$\text{subject to} \quad \begin{aligned} &\text{Tr}\big\{M_x^A\rho\big\} = q_x, \qquad \forall x \in \mathcal{X}, \\ &\textstyle\sum_x \text{Tr}_A\left\{(\text{id}_R \otimes M_x^A)\psi^{RA}\right\} = \rho, \\ &1 - \textstyle\sum_x \langle x|\,\rho M_x^{\mathcal{T}_\varphi}\rho\,|x\rangle \le D. \end{aligned} \tag{4.1.3}$$

where the quantum mutual information is with respect to the composite state $\tau_{RX} = \sum_x \sqrt{\rho}M_x\sqrt{\rho} \otimes |x\rangle\langle x|^X$.

4.1.1 Rate-Limited Optimal Transport for Qubit Measurement System

By addressing the above optimization problem, we obtain Theorem 4.1, which yields a transcendental system of equations that determines the output-constrained rate-distortion function for this system.

Theorem 4.1. *For the case of qubit input state ρ with matrix representation*

$$\rho = \begin{bmatrix} \rho_1 & \rho_2 \\ \rho_2^* & 1-\rho_1 \end{bmatrix},$$

and fixed output with Bernoulli$(1-q_0)$ *distribution, with the presence of an unlimited amount of common randomness, and using entanglement fidelity distortion measure, the output-constrained rate-distortion function and the corresponding optimal POVMs* M_0, M_1 *are provided as follows.*

For any distortion level D *above the threshold* $D \geq D_{R_0}$ *where*

$$D_{R_0} := 1 - q_0 \left\langle 0\middle|\rho^2|0\right\rangle - (1-q_0)\left\langle 1\middle|\rho^2|1\right\rangle. \tag{4.1.4}$$

the output state can be generated independently, thus $R(D;\infty,\rho||\mathit{Bern}(1-q_0)) = 0$ *with* $M_{0,R_0} = q_0 I$.

Otherwise if $D_{OT} \leq D < D_{R_0}$*, (where* D_{OT} *is the optimal transport cost) then,*

$$R(D;\infty,\rho||\mathit{Bern}(1-q_0)) = H(\rho) - q_0 H\left(\frac{N_{opt}}{q_0}\right) - (1-q_0)H\left(\frac{\rho - N_{opt}}{1-q_0}\right), \tag{4.1.5}$$

$$M_0 = \sqrt{\rho}^{-1} N_{opt} \sqrt{\rho}^{-1}, \quad M_1 = I - M_0. \tag{4.1.6}$$

The optimal parameter N_{opt}/q_0 *is the optimal PMR state conditioned on outcome 0, expressed by*

$$N_{opt} := \begin{bmatrix} n & s\rho_2/|\rho_2| \\ s\rho_2^*/|\rho_2| & q_0 - n \end{bmatrix},$$

whose variables n, s *are obtained by solving the transcendental system of equations*

$$\begin{cases} \frac{-as+b(n-q_0/2)}{E_1} \ln \frac{q_0/2+E_1}{q_0/2-E_1} + \frac{-a(s-|\rho_2|)+b(n-\rho_1+\frac{1-q_0}{2})}{E_2} \ln \frac{\frac{1-q_0}{2}+E_2}{\frac{1-q_0}{2}-E_2} & = 0 \\ an + bs + c & = 0 \end{cases}, \tag{4.1.7}$$

where

$$E_1(n,s) := \sqrt{\left(n - \frac{q_0}{2}\right)^2 + s^2}, \tag{4.1.8}$$

$$E_2(n,s) := \sqrt{\left(n - \rho_1 + \frac{1-q_0}{2}\right)^2 + (s - |\rho_2|)^2}. \tag{4.1.9}$$

The parameters a, b, c are fixed parameters of system based on the input and output states ρ, q_0 defined as

$$a := 1 - \frac{4|\rho_2|^2}{1+2k}, \qquad b := \frac{2|\rho_2|(2\rho_1 - 1)}{1+2k},$$

$$c := q_0\left(\rho_1 - 1 + \frac{2|\rho_2|^2}{1+2k}\right) + \langle 1|\rho^2|1\rangle - 1 + D, \qquad k := \sqrt{\det\{\rho\}}.$$

Proof. See Appendix A. □

The following two special input states result in interesting N_{opt} optimal matrices:

- Pure input state: For the case of pure input state, the rate-distortion curve reduces to a single point where the rate is $R = 0$, the optimal $N_{opt} = q_0\rho$ and $D = D_{R_0}$. This is because the pure input state has no correlation with the reference, so the receiver can simply use local randomness in its decoder.

- Diagonal (among canonical eigen-basis) quantum input state: In this case, the optimal operator N_{opt} will also be diagonal. One can simply find that in this case when $D \leq D_{R_0}$, the optimal operator is given by

$$N^{cl}_{opt} = \begin{bmatrix} 1 - D + (1-\rho_1)(q_0 + \rho_1 - 1) & 0 \\ 0 & D + \rho_1(q_0 + \rho_1 - 2) \end{bmatrix}.$$

4.1.2 Optimal Transport for Qubit Measurement System

The optimal transport scheme provides the minimum achievable distortion when the rate of information is not limited. The following theorem provides this value for the problem of the qubit measurement system.

Theorem 4.2. *For the case of qubit input state ρ and fixed output distribution Bern$(1-q_0)$, with the presence of the unlimited amount of common randomness, the optimal transport with respect to the entanglement fidelity distortion measure is obtained under different conditions of parameters as follows. Defining the parameter $Q := \frac{\rho_1 - 1/2}{\sqrt{1-4|\rho_2|^2}}$, we have the minimum transportation cost $D_{OT} := D(R = \infty, R_c = \infty, \rho||Bern(1-q_0))$ and the optimal operator $N_{OT} := \begin{bmatrix} n_{OT} & s_{OT}\rho_2/|\rho_2| \\ s_{OT}\rho_2^*/|\rho_2| & q_0 - n_{OT} \end{bmatrix}$ given by:*

1. *if* $\quad Q \leq \frac{\det(\rho)}{1-q_0} - \frac{1}{2}$ *then*

$$D_{OT} = q_0(1-\rho_1) + \det(\rho) + \frac{1-q_0}{2}\left(1 - \sqrt{1-4|\rho_2|^2}\right), \tag{4.1.10}$$
$$s_{OT} = \frac{b}{\sqrt{1-4|\rho_2|^2}}\frac{1-q_0}{2} + |\rho_2|,$$
$$n_{OT} = \left(\frac{a}{\sqrt{1-4|\rho_2|^2}} - 1\right)\frac{1-q_0}{2} + \rho_1.$$

2. *Else if* $Q \geq \frac{1}{2} - \frac{\det(\rho)}{q_0}$ *then*

$$D_{OT} = (1-q_0)\rho_1 + \det(\rho) + \frac{q_0}{2}\left(1 - \sqrt{1-4|\rho_2|^2}\right), \tag{4.1.11}$$
$$s_{OT} = \frac{b}{\sqrt{1-4|\rho_2|^2}}\frac{q_0}{2},$$
$$n_{OT} = \left(\frac{a}{\sqrt{1-4|\rho_2|^2}} + 1\right)\frac{q_0}{2}.$$

3. *Else if* $\frac{\det(\rho)}{1-q_0} - \frac{1}{2} \leq Q \leq \frac{1}{2} - \frac{\det(\rho)}{q_0}$ *then,*

$$D_{OT} = 1 - q_0\left(\rho_1 - 1 + \frac{2|\rho_2|^2}{1+2k}\right) - \langle 1|\rho^2|1\rangle - an_{OT} - bs_{OT}, \tag{4.1.12}$$
$$s_{OT} = \frac{(q_0 - 2\det(\rho))\,|\rho_2| + sgn\{a-b\}(1-2\rho_1)\sqrt{\Delta'}}{4|\rho_2|^2 + (1-2\rho_1)^2},$$
$$n_{OT} = \frac{2q_0|\rho_2|^2 - (1-2\rho_1)(\rho_1 q_0 - \det(\rho)) + sgn\{a-b\}2|\rho_2|\sqrt{\Delta'}}{4|\rho_2|^2 + (1-2\rho_1)^2}.$$

where $\Delta' := (q_0(1-q_0) - \det(\rho))\det(\rho)$.

Proof. See Appendix B. □

Interestingly, when the input state is prepared with the diagonal density operator along the eigenbasis of the output state (i.e. $\rho_2 = 0$), the optimal matrices N_{opt} and $\rho - N_{opt}$ encompasses the classical binary optimal transport scheme. One can see that the first and second conditions reduce to $q_0 \geq \rho_1$ and $q_0 < \rho_1$ and the third condition is empty. However, the distortion value will be obviously different.

4.1.3 Minimum Required Rate for Optimal Transport

Finally, it is important to note that an unlimited communication rate is not required for optimal transport. The minimum required rate $R_{\text{min, OT}}$ for the optimal transport scheme can be obtained by substituting optimal values s_{OT} and n_{OT} in (4.1.5), which results

$$R_{\text{min, OT}} = H(\rho) - q_0 H_b\left(\frac{1}{2} - \frac{E_1(n_{OT}, s_{OT})}{q_0}\right) - (1-q_0)H_b\left(\frac{1}{2} - \frac{E_2(n_{OT}, s_{OT})}{1-q_0}\right). \tag{4.1.13}$$

where $H_b(.)$ is the binary entropy function, and $E_1(n_{OT}, s_{OT})$ and $E_2(n_{OT}, s_{OT})$ are functions defined in (4.1.8), (4.1.9).

4.2 Numerical Results

We used the CVX package [26, 25] to find numerical solutions for the examples of this convex optimization problem. Also [22] provides the CVX functions for the von-Neumann entropy functions. The output-constrained rate-distortion function is numerically evaluated for the following set of examples with fixed $q_0 = 1/2$ and

$\rho_1 = 1/2$ parameters and different off-diagonal values.

$$\text{Ex. 1:} \quad q_0 = \frac{1}{2} \quad \rho_a = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}, \qquad \text{Tr}\{\rho_a^2\} = 0.5.$$

$$\text{Ex. 2:} \quad q_0 = \frac{1}{2} \quad \rho_b = \begin{bmatrix} 1/2 & 0.1319 - 0.0361i \\ 0.1319 + 0.0361i & 1/2 \end{bmatrix}, \qquad \text{Tr}\{\rho_b^2\} = 0.5374.$$

$$\text{Ex. 3:} \quad q_0 = \frac{1}{2} \quad \rho_c = \begin{bmatrix} 1/2 & 0.0754 - 0.2307i \\ 0.0754 + 0.2307i & 1/2 \end{bmatrix}, \qquad \text{Tr}\{\rho_c^2\} = 0.6178.$$

$$\text{Ex. 4:} \quad q_0 = \frac{1}{2} \quad \rho_d = \begin{bmatrix} 1/2 & -0.1399 - 0.3872i \\ -0.1399 + 0.3872i & 1/2 \end{bmatrix}, \qquad \text{Tr}\{\rho_d^2\} = 0.8390.$$

These rate-distortion functions are plotted in Figure 4.1, which shows that starting from a maximally mixed state (Ex.1.), as the source state becomes purer, it requires less communication rate to maintain the same level of entanglement fidelity.

In the case of a pure source state, the rate-distortion function reduces to a single point at the no transmission rate. This is intuitively acceptable as the pure state is independent of the reference state. So the receiver can generate random outcomes independent of the source. However, the entanglement fidelity distortion will not be zero because the measurement collapses the state into deterministic outcomes and hence it will not fully recover the source state. On the contrary, the maximally mixed state has the maximum dependence on the reference state which requires the maximum rate of transmission to recover the state with the same level of distortion.

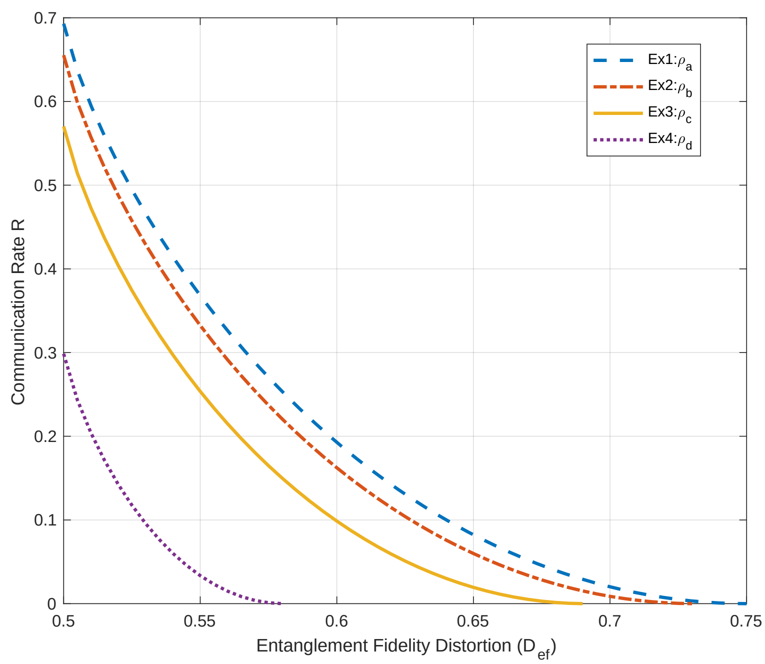

Figure 4.1: Output-Constrained rate-distortion function with unlimited common randomness for the examples with $\rho_1 = 0.5$

Chapter 5

Evaluation of the Quantum Gaussian System

In this chapter we evaluate the quantum-to-classical optimal transport for the case of Gaussian quantum systems with Gaussian source states and Gaussian output distribution. We further restrict our evaluation to unlimited common randomness ($R_c = \infty$). In the following section, we first introduce the principal definitions for the Gaussian quantum systems.

5.1 A Brief Overview of Quantum Gaussian Systems

Consider a continuous-variable quantum system with infinite-dimensional Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^s)$, which is the Hilbert space of s harmonic oscillators. The Hilbert space is defined based on the canonical observable operators $Q_1, P_1, Q_2, P_2, ..., Q_s, P_s$ with continuous eigenspectra, where Q_i, P_i are the position and momentum quadrature operators of the i-th harmonic oscillator. Thus, having the wave-function

$\psi \in \mathcal{L}^2(\mathbb{R}^s)$, the observables act as the eigenfunctions of the wave-function as follows [19]:

$$(Q_i\psi)(q) = q_i\psi(q), \quad (P_i\psi)(q) = -i\frac{\partial}{\partial q_i}\psi(q). \tag{5.1.1}$$

The canonical observables further satisfy the Canonical Commutation Relation (CCR) which is a result of the Heisenberg uncertainty principle. For more convenience, we combine the quadrature operators as elements of a $2s \times 2s$ vector operator:

$$R_1 = Q_1, \quad R_2 = P_1, \quad \cdots, \quad R_{2s-1} = Q_m, \quad R_{2s} = P_s, \tag{5.1.2}$$

which redefines the CCR as

$$[R_i, R_j] = i\Delta_{ij}I_{\mathcal{H}}, \quad i, j = 1, ..., 2m, \tag{5.1.3}$$

with Δ being non-degenerate skew-symmetric symplectic matrix defined as

$$\Delta = \bigoplus_{k=1}^{s} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{5.1.4}$$

5.1.1 Phase-Space Representation [29, 50]

Similar to the discrete case, a continuous quantum state is represented by its density operator $\rho \in \mathcal{D}(\mathcal{H})$. However, because of the continuity of the eigenfunctions, it is not possible to represent the density operator in matrix format. Instead, the density operator has an equivalent representation over a real symplectic space which is interpreted as the Wigner quasi-probability distribution. We first define the Weyl operator

$$W(z) = \exp\{iRz\}. \tag{5.1.5}$$

where $z := [q_1, p_1, ..., q_s, p_s]$ with $z \in \mathbb{R}^{2s}$ is a vector of eigenvalues of the corresponding quadrature operators. The CCR can be interpreted in the Weyl-Segal format as follows:

$$W(z)W(z') = \exp\left\{-\frac{i}{2}\Delta(z, z')\right\}W(z + z') \tag{5.1.6}$$

where

$$\Delta(z, z') = z^T \Delta z \tag{5.1.7}$$

is the canonical symplectic form and Δ is the symplectic matrix defined in (5.1.4). Then for a density operator ρ of an arbitrary quantum state, the Wigner characteristic function is defined as

$$\phi_\rho(z) = \mathrm{Tr}\{\rho W(z)\}. \tag{5.1.8}$$

One can further revert back to the density operator by performing this operator Fourier transform

$$\rho = \frac{1}{(2\pi)^s} \int \phi_\rho(z) W(-z) d^{2s} z. \tag{5.1.9}$$

However, by taking a Fourier transform of the characteristic function with respect to a variable $x \in \mathcal{X} \equiv \mathcal{Z} = \mathbb{R}^{2s}$, we obtain the Wigner quasi-probability distribution

$$\Phi_\rho(x) = \int_Z \phi_\rho(z) \exp\{-ix^T z\} \frac{d^{2s}z}{(2\pi)^{2s}}. \tag{5.1.10}$$

The Wigner quasi-probability distribution function is a real function normalized to 1, but not necessarily non-negative. For a detailed definition of the Wiegner function see [[50] section II.A] and [[46] Chapter 4].

The domain of these Wigner functions $\mathcal{Z} = \mathbb{R}^{2s}$ together with the symplectic matrix Δ form a *symplectic space* $\mathcal{K} := (\mathcal{Z}, \Delta)$ which is called the phase space. Hence, the Wigner characteristic function and Wigner distribution are the unique phase space representations of the quantum state ρ according to Stone-von Neumann's uniqueness theorem.

A transformation matrix T is called a symplectic transformation if it maps the symplectic space into itself; i.e. preserving the symplectic form

$$\Delta(Tz, Tz') = \Delta(z, z') \quad \text{for all } \; z, z' \in \mathcal{Z}. \tag{5.1.11}$$

5.1.2 Statistical Moments

Similar to the ordinary probability distribution, we can further define the statistical moments based on the Wigner functions, which provide important properties of the quantum state. It can be shown that the following relations hold between the moments of the Wigner functions and the quantities of the density operator in Hilbert space:

Mean (Displacement)

Having the first moment, also called the displacement by m and the mean value of x as

$$m := \mathrm{Tr}\{\rho R\}, \quad \bar{x} := \mathbb{E}_{\phi_\rho}\left[x\right], \tag{5.1.12}$$

then the following relation holds

$$\bar{x} = m. \tag{5.1.13}$$

A quantum state ρ with a zero mean can be moved to a mean value m with the displacement operator $D(m) := W(\Delta^{-1}m)$ by

$$D(m)\rho D(m)^\dagger. \tag{5.1.14}$$

Covariance Matrix

For any two operators X, Y one can write

$$XY = \frac{1}{2}\{X, Y\} + \frac{1}{2}[X, Y], \tag{5.1.15}$$

where $[X, Y]$ and $\{X, Y\}$ are the commutator and anti-commutator of X, Y respectively. Therefore, having a quantum state ρ, and the quadrature operators $R = \{R_1, ..., R_n\}$, we define two real matrices, a symmetric and an anti-symmetric one

$$\alpha = B_\rho(R) = \frac{1}{2}\left[\operatorname{Tr}\rho\Big\{(R_j - m_j), (R_k - m_k)\Big\}\right]_{j,k=1,\ldots,n}, \tag{5.1.16}$$

$$\Delta = C_\rho(R) = \frac{1}{2}\left[\operatorname{Tr}\rho\Big[(R_j - m_j), (R_k - m_k)\Big]\right]_{j,k=1,\ldots,n}, \tag{5.1.17}$$

where $m_k = Tr\rho R_k$. The matrices α and Δ are called the covariance matrix and the commutation matrix of the density operator, respectively [29]. Then we have the following Hermitian matrix

$$\alpha - \frac{i}{2}\Delta = \operatorname{Tr}\big\{(R-m)^T \rho\,(R-m)\big\} \tag{5.1.18}$$

where $.^T$ is a transposition with respect to the vector space of R. The above covariance matrix is related to the covariance matrix $V_x := \mathbb{E}_{\Phi_\rho}\left[(x-\bar{x})(x-\bar{x})^T\right]$ of the Wigner function as follows

$$V_x = \alpha. \tag{5.1.19}$$

Further, the following inequality holds (in the positive semi-definite sense) for the covariance matrix of the quantum state which is a result of the uncertainty principle:

$$\alpha \geq \pm\frac{i}{2}\Delta. \tag{5.1.20}$$

The Symplectic Eigenvalues

Having the bilinear form in (5.1.7) with covariance matrix α in the symplectic space (Z, Δ), there exists a symplectic transformation T in Z which diagonalizes α with the matrix [Lemma 12.12 [29]]

$$\tilde{\alpha} := T^T \alpha T = \text{diag} \begin{bmatrix} \gamma_j & 0 \\ 0 & \gamma_j \end{bmatrix}. \tag{5.1.21}$$

where $\gamma_j > 0$ are called the *symplectic eigenvalues* of the j-th mode. "The eigenvalues of the covariance matrix α don't have an intrinsic meaning as they describe a quadratic form rather than an operator and depend on the choice of basis in Z. However, the operator $\hat{\alpha} = \Delta^{-1}\alpha$ has a basis-free meaning"[31]. The eigenvalues of $\hat{\alpha}$ are complex values $\pm\gamma_j$. Diagonalization of this operator is the same as the *normal mode decomposition* of the phase space and leads to the above transformation in (5.1.21).

The symplectic eigenvalues γ_j can be simply obtained from

$$\left(\Delta^{-1}\alpha\right)^2 = \text{diag} \begin{bmatrix} -\gamma_j^2 & 0 \\ 0 & -\gamma_j^2 \end{bmatrix} \tag{5.1.22}$$

which gives

$$\text{abs}\left(\Delta^{-1}\alpha\right) = \sqrt{\left(\Delta^{-1}\alpha\right)^2} = \text{diag} \begin{bmatrix} \gamma_j & 0 \\ 0 & \gamma_j \end{bmatrix} = \hat{\alpha}. \tag{5.1.23}$$

Moreover, the matrix uncertainty relation of (5.1.20) can be described in the following format

$$\left(\Delta^{-1}\alpha\right)^2 \geq -\frac{1}{4}I \tag{5.1.24}$$

with equality when the system is in a pure state.

In the special case when the state is a product state of the modes, which is when α is a block-diagonal of s separate α_j 2-by-2 matrices, the symplectic eigenvalues are simply the determinant of each mode

$$\gamma_j = \det\{\alpha_j\}. \tag{5.1.25}$$

5.1.3 Quantum Gaussian States

Next, the quantum Gaussian state is defined on $\mathcal{H}$ as a state whose Wigner representation is Gaussian;

$$\phi_\rho(z) = \exp\left[-\frac{1}{2}z^T\alpha z + im^T z\right],$$
$$\Phi_\rho(x) = \frac{\exp\left[-\frac{1}{2}(z-m)^T\alpha^{-1}(z-m)\right]}{(2\pi)^s\sqrt{\det\alpha}}$$

Similar to classical Gaussian distributions, for a Gaussian quantum state, the displacement and covariance matrix are sufficient to fully represent the Gaussian state.

5.1.4 Continuous-Variable Measurement POVM

The measurement POVM M has a general form as defined in [29]. An important group of observables is the general form of the covariant Gaussian observable as provided in [30]:

$$\tilde{M}(d^{2s}z) = D(Kz)\rho_G D(Kz)^\dagger \frac{|\det K|^2 d^{2s}z}{\pi^s} \tag{5.1.26}$$

where ρ_G is a density operator that is the parameter of the Gaussian observable and is different from the input d.o. ρ_N. A special case of this group of measurements is the heterodyne measurement, for which the POVM is given by

$$M_*(d^{2s}z) = D(z)\rho_0 D(z)^\dagger \frac{d^{2s}z}{\pi^s} = |z\rangle\langle z| \frac{d^{2s}z}{\pi^s}. \tag{5.1.27}$$

For any measurement POVM M with the collection of $\{\rho_z\}_{z\in\mathcal{Z}}$ post-measured reference states, the information gain function is defined as

$$I_g(\rho_N; M) = H(\rho_N) - \int_{\mathcal{Z}} H(\rho_z)\mu_Z(dz). \tag{5.1.28}$$

The von-Neumann entropy of the general Gaussian state with covariance matrix α is given by [31],

$$H(\rho_N) = \frac{1}{2}\text{Sp}g\left(\text{abs}(\Delta^{-1}\alpha) - \frac{I}{2}\right). \tag{5.1.29}$$

where Sp $(.)$ is the matrix trace as opposed to the Tr the trace in Hilbert space and that

$$g(x) = \begin{cases} (x+1)\log(x+1) - x\log x, & x > 0 \\ 0, & x = 0 \end{cases} \tag{5.1.30}$$

is the Gordon function.

5.1.5 The Distortion observable

The Transpose Hilbert Space

In this chapter, we use the same distortion observable operator that was used by [19] with some modifications to make it suitable for the Measurement system. Note that the distortion observable operator is applied to a coupling of the source and destination states as defined by a composite state. In a prior work [24], defines the coupling $\Pi \in (\mathcal{H}_2 \otimes \mathcal{H}_1)$ for marginal states $\rho, \sigma \in \mathcal{H}$ such that

$$\mathrm{Tr}_{\mathcal{H}_2}\{\Pi\} = \rho, \quad \mathrm{Tr}_{\mathcal{H}_1}\{\Pi\} = \sigma, \tag{5.1.31}$$

where $\mathcal{H}_1, \mathcal{H}_2$ are the copies of $\mathcal{H}$. As [19] mentions, with this definition of coupling, there is no physical channel that can physically interpret the provided optimal transport plan. Instead, [19] introduces a coupling in the transpose Hilbert space. First, we define the transpose of an operator with respect to a basis as follows:

Definition 5.1. *Let $\{|i\rangle\}_{i=1}^{d}$ be a basis in the Hilbert space $\mathcal{H}$. Having an operator $A \in \mathcal{H}$,* *the operator $A^{\mathcal{T}} \in \mathcal{H}^*$ is defined as the tranpose of A where*

$$A^{\mathcal{T}} := \sum_{i}^{d} \sum_{j}^{d} \langle j|A|i\rangle \, |i\rangle\langle j|_{H^*} \, . \tag{5.1.32}$$

Next, we define the canonical purification as follows:

Definition 5.2. *The canonical purification of a state $\rho \in \mathcal{H}_A$ to the composite state of reference and local system is given by*

$$|\psi\rangle_{RA} = (I \otimes \sqrt{\rho})\, |\Gamma\rangle_{RA} \tag{5.1.33}$$

where $|\Gamma\rangle_{RA} := \sum_{i=1}^{d} |i\rangle_R \, |i\rangle_A$ is the maximally entangled Bell state.

We can further show that as [19] mentions,

Lemma 5.1. *The canonical purification of $|\psi\rangle_{RA}$ resides in the composite state of the transpose and local state; i.e.,*

$$|\psi\rangle_{RA} \in \mathcal{H}_A \otimes \mathcal{H}_A^*. \tag{5.1.34}$$

Proof. By expanding the density operator of the canonical purification, we have

$$|\psi_{RA}\rangle\langle\psi_{RA}| = (I_R \otimes \sqrt{\rho}_A)\, |\Gamma_{RA}\rangle\langle\Gamma_{RA}|\, (I \otimes \sqrt{\rho}_A) \tag{5.1.35}$$

$$= (I_R \otimes \sqrt{\rho}_A) \left(\sum_{i=1}^{d} |i\rangle\langle j|_R \otimes |i\rangle\langle j|_A \right) (I \otimes \sqrt{\rho}_A). \tag{5.1.36}$$

Then the reference state is obtained by tracing over the local state A,

$$\mathrm{Tr}_A \{|\psi_{RA}\rangle\langle\psi_{RA}|\} = \mathrm{Tr}_A \left\{ (I_R \otimes \sqrt{\rho}_A) \left(\sum_{i=1}^{d} |i\rangle\langle j|_R \otimes |i\rangle\langle j|_A \right) (I \otimes \sqrt{\rho}_A) \right\} \tag{5.1.37}$$

$$= \sum_{i=1}^{d} \langle j|\rho|i\rangle \, |i\rangle\langle j|_R \tag{5.1.38}$$

$$= (\rho^{\mathcal{T}})^R \tag{5.1.39}$$

This shows that the canonical purification $|\psi\rangle_{RA} \in (\mathcal{H} \otimes \mathcal{H}^*)$. □

Therefore, [19] defines the coupling for their optimal transport problem according to an optimal transportation channel Φ with the constraint that $\Phi(\rho) = \sigma$, as a composite state $\Pi_\Phi \in (\mathcal{H} \otimes \mathcal{H}^*)$ such that

$$\Pi_\Phi = (\Phi \otimes I^*_{\mathcal{H}})(|\psi_{RA}\rangle\langle\psi_{RA}|) \tag{5.1.40}$$

which ensures

$$\mathrm{Tr}_{\mathcal{H}} \{\Pi_\Phi\} = \rho^{\mathcal{T}}, \quad \mathrm{Tr}_{\mathcal{H}^*} \{\Pi_\Phi\} = \sigma. \tag{5.1.41}$$

In light of the above argument, it is important to define their distortion observable in the following form [19]:

$$C = \frac{1}{2s} \sum_{i=1}^{2s} (R_i^{\mathcal{T}} \otimes I_{\mathcal{H}} - I_{\mathcal{H}^*} \otimes R_i)^2. \tag{5.1.42}$$

This is particularly important because the reference state resides in the H^* Hilbert space, therefore, if we want to compare the outcomes of the states, we must use the $R_i^{\mathcal{T}}$ operator for the reference so that $\mathrm{Tr}\{R^{\mathcal{T}}\rho^{\mathcal{T}}\} = \mathrm{Tr}\{R\rho\}$.

The Quantum-to-Classical Distortion Observable

In this part, we introduce the quadratic operator distortion observable for the Gaussian QC systems. But first, we show that a similar lemma also holds for the measurement systems as well.

Lemma 5.2. *Let $\rho \in \mathcal{H}_A$ be the source state and $\mathcal{M} \equiv \{M\}_u \in \mathcal{U}$ be a measurement POVM where $\mathcal{U}$ is the space of the outcomes. Then, the conditional post-measured reference states and the unrevealed post-measured reference state are all in the transpose Hilbert space $\mathcal{H}^*$.*

Proof. Let $\Lambda = M_u^\dagger M_u$, and $\lambda_u = P(U = u) = \mathrm{Tr}\{M_u\rho\}$. Then the state of the system after measurement given that $\{X = u\}$ is

$$\left|\psi_u^{RA'}\right\rangle := \frac{(I \otimes M_u)\,|\phi_{RA}\rangle}{\sqrt{\lambda_u}}, \tag{5.1.43}$$

and its density operator is

$$\psi_u^{RA'} = \frac{(I \otimes M_u)\,|\phi_{RA}\rangle\!\langle\phi_{RA}|\,(I \otimes M_u^\dagger)}{\lambda_u}. \tag{5.1.44}$$

Then the state of the reference given that $\{U = u\}$ is

$$\mathrm{Tr}_{A'}\left\{\psi_u^{RA'}\right\} = \frac{1}{\lambda_u}\sum_{i,j}\left\langle j\middle|\sqrt{\rho}M_u^\dagger M_u\sqrt{\rho}\middle|i\right\rangle_{A'}|i\rangle\!\langle j|_R \tag{5.1.45}$$

$$= \frac{\left(\sqrt{\rho}\Lambda_u\sqrt{\rho}\right)^{\mathcal{T}}}{\lambda_u} := \left(\hat{\rho}_u^{\mathcal{T}}\right)^R \tag{5.1.46}$$

If the outcomes are not revealed, then the reference state is

$$\sum_{u} \lambda_u \frac{\left(\sqrt{\rho}\Lambda_u\sqrt{\rho}\right)^{\mathcal{T}}}{\lambda_u} = \left(\rho^{\mathcal{T}}\right)^R. \tag{5.1.47}$$

The proof simply extends to the continuous space POVMs as well. □

Thus, the post-measured composite state is defined as follows

$$\psi^{RZ} = \sum_{z\in\mathcal{Z}} \sqrt{\rho_A}^{\mathcal{T}} \Lambda_z^{\mathcal{T}} \sqrt{\rho_A}^{\mathcal{T}} \otimes |z\rangle\langle z| \tag{5.1.48}$$

$$= \sum_{z\in\mathcal{Z}} \rho_z^{\mathcal{T}} \otimes \pi(z)\, |z\rangle\langle z| \tag{5.1.49}$$

Then applying the distortion observable operator (5.1.42) to the post-measured composite state gives

$$d(\rho_A, \Lambda) = \mathrm{Tr}\{\psi^{RZ} C\}$$
$$= \frac{1}{2s}\,\mathrm{Tr}\left\{ \left(\sum_z (\rho_z^{\mathcal{T}})^R \otimes \pi(z)\,|z\rangle\langle z|\right) \left(R_i^{\mathcal{T}} \otimes I_{\mathcal{H}} - \bar{m}_i(z) + \bar{m}_i(z) - I_{\mathcal{H}^*} \otimes R_i\right)^2 \right\} \tag{5.1.50}$$
$$= \frac{1}{2s}\,\mathrm{Tr}\left\{ \left(\sum_z (\rho_z^{\mathcal{T}})^R \otimes \pi(z)\,|z\rangle\langle z|\right) \left[\left(R_i^{\mathcal{T}} \otimes I_{\mathcal{H}} - \bar{m}_i(z)\right)^2 + \left(I_{\mathcal{H}^*} \otimes R_i - \bar{m}_i(z)\right)^2 + 2\left(R_i^{\mathcal{T}} \otimes I_{\mathcal{H}} - \bar{m}_i(z)\right)\left(\bar{m}_i(z) - I_{\mathcal{H}^*} \otimes R_i\right)\right] \right\} \tag{5.1.51}$$
$$= \frac{1}{2s}\,\mathrm{Tr}\left\{ \left(\sum_z (\rho_z^{\mathcal{T}})^R \otimes \pi(z)\,|z\rangle\langle z|\right) \left[\left(R_i^{\mathcal{T}} \otimes I_{\mathcal{H}} - \bar{m}_i(z)\right)^2 + \left(I_{\mathcal{H}^*} \otimes R_i - \bar{m}_i(z)\right)^2\right] \right\} \tag{5.1.52}$$
$$= \frac{1}{2s}\sum_{i=1}^{2s}\sum_{z\in\mathcal{Z}} \pi(z)\Big[\mathrm{Tr}\{\rho_z^R (R_i - \bar{m}_i(z))^2\} + \langle z|(R_i - \bar{m}_i(z))^2|z\rangle\Big] \tag{5.1.53}$$
$$= \frac{1}{2s}\sum_{i=1}^{2s}\Big[\sum_{z\in\mathcal{Z}} \pi(z)\,\mathrm{Tr}\{\rho_z^R (R_i - \bar{m}_i(z))^2\} + \mathbb{E}_\pi\left[(Z_{R_i} - \bar{m}_i(Z))^2\right]\Big] \tag{5.1.54}$$
$$= \frac{1}{2s}\sum_{i=1}^{2s}\Big[\sum_{z\in\mathcal{Z}} \pi(z)[\Sigma(z)]_{ii} + \mathbb{E}_\pi\left[(Z_{R_i} - \bar{m}_i(Z))^2\right]\Big] \tag{5.1.55}$$

where $\bar{m}_i(z) = \mathrm{Tr}\{\rho_z^R R_i\} = \mathrm{Tr}\Big\{(\rho_z^{\mathcal{T}})^R R_i^{\mathcal{T}}\Big\}$ is the first moment of the quantum state and $\Sigma(z)$ is the covariance matrix of the state. One can further write the distortion constraint as

$$d(\rho_A, \Lambda) = \frac{1}{2s} \sum_{z \in \mathcal{Z}} \pi_Z(z) \cdot \mathrm{Sp}\{\Sigma(z)\} + \frac{1}{2s} \sum_{z \in \mathcal{Z}} \pi_Z(z) \cdot \|z - \bar{m}(z)\|^2 \tag{5.1.56}$$

The above formulation describes the distortion function for a point-to-point system, where the measurement outcome space is the final output space $\mathcal{Z}$.

5.2 OC Rate-Distortion Function of QC Gaussian Systems with Unlimited Common Randomness

Recall that the single-letter conditions of the rate pair being inside the rate region according to the coding theorem 3.2 are

$$R \geq I_g(R; W), \tag{5.2.1}$$

$$R + R_c \geq I(W; X), \tag{5.2.2}$$

$$\int_w P_{X|W}(A|w) \,\mathrm{Tr}\{M(dw)\rho\} = \mu_X(A), \quad \text{for } A \in \mathcal{B}(\mathcal{X}), \tag{5.2.3}$$

$$\int_{x \in \mathbb{R}} \mathrm{Tr}_R \left[\rho_x \, \Delta_R(x)\right] \mu_X(dx) \leq D. \tag{5.2.4}$$

Note that when evaluating this system and finding the rate-distortion curve, due to the ensemble-observable duality [30], we can either work with the measurement itself or the post-measured reference ensembles. When we choose measurement POVM $M(dw)$ as the variable of optimization, the input marginal constraint becomes a trivial result of $\int_{\mathcal{W}} M(dw) = I$. Because multiplying both sides by $\sqrt{\rho}$ from left and right will give the input marginal constraint $\int_{\mathcal{W}} \sqrt{\rho} M(dw) \sqrt{\rho} = \rho$.

On the other hand, when choosing the measurement outcome ensemble $\{\rho_w, \pi_W\}$ as the variables of optimization, the input constraint becomes non-trivial, so we have both input and output constraints as

$$\int_{\mathcal{W}} \rho_w \pi_W(w) = \rho \tag{5.2.5}$$

$$\int_{\mathcal{W}} P_{X|W}(x|w)\pi_W(w) = \mu_X(x). \tag{5.2.6}$$

Furthermore, when having unlimited amount of common randomness ($R_c = \infty$), the single-letter constraints reduce to

$$R \geq I_g(R; W) = I_g(R; X) \tag{5.2.7}$$

$$\int_{x\in\mathbb{R}} \mathrm{Tr}_R \left[\rho_x \, \Delta_R(x)\right] \mu_X(dx) \leq D, \tag{5.2.8}$$

because with no loss of generality, one can assume $X = W$ as a result of the data processing inequality.

5.2.1 Optimality of Gaussian Quantum Measurement

We next provide the following theorem about the optimality of the Gaussian measurements in the case of unlimited common randomness.

Theorem 5.3. *Suppose having a rate-limited Gaussian QC optimal transport Measurement system with unlimited common randomness with a Gaussian source state ρ with mean vector m_ρ and covariance matrix Σ_ρ and a Gaussian destination distribution $\pi_Z \equiv \mathcal{N}(\mu_Z, \Sigma_Z)$. Having a distortion observable of the form* (5.1.42)*, for any feasible distortion threshold D, the optimal measurement that minimizes the transmission rate (Groenwold's* *information quantity) is a Gaussian measurement of the form*

$$D(Kz)\rho_{N_G}D(Kz)^\dagger \tag{5.2.9}$$

where ρ_N is a zero-mean Gaussian quantum state with covariance matrix Σ_N determining the noise of the measurement, and K is a matrix transformation of the form

$$K = \Sigma_Z^{-1/2}\left(\Sigma_Z^{1/2}(\Sigma_\rho - \Sigma_N)\Sigma_Z^{1/2}\right)^{1/2}\Sigma_Z^{-1/2}. \tag{5.2.10}$$

Proof. With no loss of generality, we first assume that the quantum source state and the classical output distribution are both transported to the origin having zero means. Assume having an arbitrary set of PMR states $\{\hat{\rho}_z\}_{z\in\mathcal{Z}}$ with mean values $\bar{m}(z)$ and covariance matrices $\hat{\Sigma}(z)$, which are the equivalent of the classical backward channels $P_{X|Y}(.|y)$ for different realizations of $y \in \mathcal{Y}$. In this setting $\bar{m}(z) = \mathrm{Tr}\{\rho_z R\}$ is equivalent to the classical MMSE estimator $\tilde{X} = \mathbb{E}\left[X|Y\right]$. Therefore, the law of total variance in lemma 5.4 holds for the covariance of the source:

$$\Sigma_\rho = \int \hat{\Sigma}(z)\pi_Z(dz) + \tilde{\Sigma}. \tag{5.2.11}$$

We further define a centralized version of the PMR states

$$\hat{\rho}_{c,z} := D^\dagger(\bar{m}(z))\ \hat{\rho}_z\ D(\bar{m}(z)) \tag{5.2.12}$$

which shifts all of the PMR states to origin. Thus, we introduce the following upperbound for the conditional entropy of PMR states

$$H(R|Z) = \int H(\hat{\rho}_z)\mu_Z(dz) \tag{5.2.13}$$

$$= \int H\left(\hat{\rho}_{c,z}\right)\mu_Z(dz) \tag{5.2.14}$$

$$\leq H\left(\hat{\rho}_N := \int \hat{\rho}_{c,z}\ \mu_Z(dz)\right) \leq H(\tilde{\rho}_{N_G}) \tag{5.2.15}$$

where the first inequality follows from the concavity of von-Neumann entropy. We defined $\hat{\rho}_N$ to be the average state with zero mean and the covariance matrix $\hat{\Sigma}_N$, where

$$\hat{\Sigma}_N = \mathrm{Tr}\big\{\hat{\rho}_N(R-\bar{m}_N)(R-\bar{m}_N)^T\big\} = \mathrm{Tr}\big\{\hat{\rho}_N RR^T\big\} \tag{5.2.16}$$

$$= \mathrm{Tr}\bigg\{\left(\int D^\dagger(\bar{m}(z))\ \hat{\rho}_z\ D(\bar{m}(z))\mu_Z(dz)\right) RR^T\bigg\} \tag{5.2.17}$$

$$= \int \hat{\Sigma}(z)\mu_Z(dz). \tag{5.2.18}$$

Using Lemma 5.4 and the above equality we find the covariance of estimator $\tilde{\Sigma} = \Sigma_\rho - \hat{\Sigma}_N$.

Also in (5.2.15), in the last inequality, $\tilde{\rho}_{N_G}$ is introduced as a Gaussian quantum state with the same covariance matrix $\hat{\Sigma}_N$ which appeals to the Quantum Gaussian entropy maximization theorem [32]. Further, the distortion function for the arbitrary PMR ensemble is

$$d(\hat{\rho}_z, \pi_Z) = \frac{1}{2s}\int \mathrm{Sp}\left(\hat{\Sigma}(z)\right)\mu_Z(dz) + \frac{1}{2s}\int \|z-\bar{m}(z)\|^2\mu_Z(dz) \tag{5.2.19}$$

$$= \frac{1}{2s}\,\mathrm{Sp}\left(\hat{\Sigma}_N\right) + \frac{1}{2s}\int \|z-\bar{m}(z)\|^2\mu_Z(dz). \tag{5.2.20}$$

We next form a different set of PMR ensembles equivalent to the quantum Gaussian measurement of the form

$$\tilde{\rho}_z = D(Kz)\tilde{\rho}_{N_G} D(Kz)^\dagger. \tag{5.2.21}$$

Note that by using $\tilde{\rho}_z$, because the covariance matrix $\hat{\Sigma}_N$ is fixed, the first term of distortion does not change.

We further limit the selection of the K matrix such that it satisfies the covariance matrix of estimator $\tilde{\Sigma} = K\Sigma_Z K^T$. This, in turn, makes sure that the marginal constraint is satisfied $\rho = \int \tilde{\rho}_z \mu_Z(dz)$ because the source quantum state is assumed to be Gaussian. The second term of distortion is only a function of $\bar{m}(z)$. So by preserving the covariance matrix $\tilde{\Sigma} = \Sigma_\rho - \hat{\Sigma}_N$, we select K to be the optimal transportation from Gaussian distribution with Σ_Z to Gaussian distribution with $\tilde{\Sigma}$; i.e. we choose

$$K := \Sigma_Z^{-1/2} \left(\Sigma_Z^{1/2} (\Sigma_\rho - \Sigma_N) \Sigma_Z^{1/2} \right)^{1/2} \Sigma_Z^{-1/2}. \tag{5.2.22}$$

This will ensure that the following relation holds:

$$\mathbb{E}_Z \left[\|Z - \bar{m}(Z)\|^2 \right] \geq W_2^2(\mathcal{N}(\Sigma_X - \Sigma_N), \mathcal{N}(\Sigma_Y)) = \mathbb{E}_Z \left[\|Z - KZ\|^2 \right]. \tag{5.2.23}$$

where the inequality appeals to lemma 5.5.

The above steps show that for a fixed noise covariance $\hat{\Sigma}_N$, the PMR ensemble of the Gaussian form (5.2.21) does not reduce the conditional entropy while also preserving the distortion and marginal constraints. This proves that replacing

$\{\hat{\rho}_z, \pi_Z(dz)\}$ with the Gaussian measurement $\{\tilde{\rho}_z, \pi_Z(dz)\}$ causes no loss of optimality. □

Lemma 5.4. *For a source quantum state ρ and a measurement POVM with the corresponding PMR ensemble $\{\hat{\rho}_z^T, \pi_Z(dz)\}$, we have*

$$\Sigma_\rho = \int \hat{\Sigma}(z)\pi_Z(dz) + \tilde{\Sigma} \tag{5.2.24}$$

where

$$\Sigma_\rho = \mathrm{Tr}\{\rho(R-m)(R-m)^T\} \tag{5.2.25}$$

$$\hat{\Sigma}(z) = \mathrm{Tr}\{\rho_z(R-\bar{m}(z))(R-\bar{m}(z))^T\} \quad \textit{for all } z \in \mathcal{Z} \tag{5.2.26}$$

$$\tilde{\Sigma} := \mathit{Cov}(m(Z)) = \mathbb{E}_Z\left[m(Z)m(Z)^T\right] - \bar{m}\bar{m}^T \tag{5.2.27}$$

are the covariance matrix of the source, the conditional covariance of PMR state $\hat{\rho}_z$ and the covariance of the estimation $\bar{m}(Z)$ respectively.

Proof. Starting with second moment of the source state

$$\mathrm{Tr}\{\rho RR^T\} = \mathrm{Tr}\left\{\left(\int \hat{\rho}_z \pi_Z(dz)\right) RR^T\right\} = \int \mathrm{Tr}\{\rho_z RR^T\}\pi_Z(dz) \tag{5.2.28}$$

$$= \int \left(\hat{\Sigma}(z) + \bar{m}(z)\bar{m}(z)^T\right)\pi_Z(dz), \tag{5.2.29}$$

then the covariance of the source state is

$$\Sigma_\rho = \mathrm{Tr}\{\rho RR^T\} - \bar{m}\bar{m}^T = \int \left(\hat{\Sigma}(z)\right)\pi_Z(dz) + \mathbb{E}_Z\left[\bar{m}(z)\bar{m}(z)^T\right] - \bar{m}\bar{m}^T. \tag{5.2.30}$$

□

Lemma 5.5. *Assuming a classical system having a set of source distributions $\Omega_X(0, \Sigma_X)$ and a set of destination distributions $\Omega_Y(0, \Sigma_Y)$ with zero mean and fixed given covariance, the minimum MSE distance between the two sets is the Wasserstein distance between two Gaussian distributions with given moments. i.e.,*

$$\min_{\substack{X \sim \mathcal{X} \in \Omega_X \\ Y \sim \mathcal{Y} \in \Omega_Y}} \mathbb{E}\left[\|X - Y\|^2\right] = W_2^2(\mathcal{N}(0, \Sigma_X), \mathcal{N}(0, \Sigma_Y)). \tag{5.2.31}$$

Proof. The desired conclusion follows from the observation that $\mathbb{E}\left[\|X - Y\|^2\right]$ is preserved when (X, Y) is replaced by the jointly Gaussian pair (X_G, Y_G) with the same joint covariance matrix. □

5.2.2 The Quantum-Classical Gaussian Optimization problem

To obtain the OC rate-distortion function, by using the above Gaussian optimality theorem, it suffices to find the optimal noise covariance matrix Σ_N as follows:

$$\begin{aligned} R(D; R_c = \infty, \rho||\pi_Z) &= \min_{\Sigma_N} H_G(\Sigma_\rho) - H_G(\Sigma_N) \\ \text{subject to:} \quad & \frac{1}{2s}\,\text{Sp}\,(\Sigma_N) + \frac{1}{2s}\int \|z - Kz\|^2 \pi_Z(dz) + \frac{1}{2s}\|m_\rho - \mu_Z\|^2 \leq D, \end{aligned} \tag{5.2.32}$$

where K is the optimal transportation mapping to the classical output space and the distortion function given by (5.2.22) similarly follows:

$$\begin{aligned} d(\rho, M, h) &= \frac{1}{2s}\,\text{Sp}\,(\Sigma_N) + \frac{1}{2s}\mathbb{E}\left[\|(K - I)\,Z\|^2\right] + \frac{1}{2s}\|m_\rho - \mu_Z\|^2 \\ &= \frac{1}{2s}\text{Sp}\left(\Sigma_\rho + \Sigma_Z - 2\left(\Sigma_Z^{1/2s}(\Sigma_\rho - \Sigma_N)\Sigma_Z^{1/2}\right)^{1/2}\right) + \frac{1}{2s}\|m_\rho - \mu_Z\|^2 \leq D. \end{aligned}$$

The above distortion inequality can be further simplified to

$$\frac{1}{2s}\,\mathrm{Sp}\left((\Sigma_Z^{1/2}(\Sigma_\rho - \Sigma_N)\Sigma_Z^{1/2})^{1/2}\right) \geq \frac{1}{2}\left(D_{\max} - D\right) \tag{5.2.33}$$

where

$$D_{\max} := \frac{1}{2s}\,\mathrm{Sp}\left(\Sigma_\rho + \Sigma_Z\right) + \frac{1}{2s}\|m_\rho - \mu_Z\|^2 \tag{5.2.34}$$

is the zero-crossing point at which the required transmission rate to achieve the distortion $D \geq D_{\max}$ is zero. This means that in this range the acceptable distortion level is high enough to allow Bob to generate the output independent of the source according to the given distribution. Thus the optimization problem in (5.2.32) reduces to

$$R(D; R_c = \infty, \rho||\pi_Z) := \min_N \frac{1}{2}\mathrm{Sp}\ g\left(\left|\Delta^{-1}\Sigma_\rho\right| - \frac{I}{2}\right) - \frac{1}{2}\mathrm{Sp}\ g\left(\left|\Delta^{-1}\Sigma_N\right| - \frac{I}{2}\right)$$

$$\text{subject to: } \frac{1}{s}\,\mathrm{Sp}\left((\Sigma_Z^{1/2}(\Sigma_\rho - \Sigma_N)\Sigma_Z^{1/2})^{1/2}\right) \geq D_{\max} - D \tag{5.2.35}$$

$$\Sigma_N \leq \Sigma_\rho, \quad \left|\Delta^{-1}\Sigma_N\right| \geq \frac{I}{2} \tag{5.2.36}$$

where $|A| := \sqrt{AA^T}$ is the absolute value of the matrix A, and $g(.)$ is the Gordon function. Next, we solve this optimization problem for the following cases.

5.2.3 The isotropic Gaussian source and destination

Assume having a Gaussian source state ρ with an isotropic covariance matrix $\Sigma_\rho = \sigma_\rho^2 I_{2s\times 2s}$ and a Gaussian destination distribution π_Z with covariance matrix $\Sigma_Z = \sigma_Z^2 I_{2s\times 2s}$. Let the Gaussian measurement have a noise covariance of

$\Sigma_N = n\sigma_\rho^2 I_{2s\times 2s}$, where n is the parameter of optimization. Thus the optimal noise-to-power ratio is obtained for this case:

$$n^* = \begin{cases} \text{N/A} & D < D_{\text{min}}, \\ 1 - \left(\frac{D_{\text{max}} - D}{2\sigma_\rho \sigma_Z}\right)^2 \quad \text{for} & D_{\text{min}} \le D \le D_{\text{max}}, \\ 1 & D_{\text{max}} < D. \end{cases} \tag{5.2.37}$$

where $D_{\text{max}} = (m_\rho - m_Z)^2 + (\sigma_\rho^2 + \sigma_Z^2)$. Moreover, the feasibility conditions on $n^* \ge \frac{1}{2\sigma_\rho^2}$ will give the optimal transport cost $D_{\text{min}} = D_{\text{max}} - 2\sigma_Z\sqrt{(\sigma_\rho^2 - \frac{1}{2})}$. Then the OC rate-distortion function is obtained by

$$R\Big(D; \mathcal{QN}(m_\rho, \sigma_\rho^2)||\mathcal{N}(m_Z, \sigma_Z^2)\Big) = g\Big(\sigma_\rho^2 - 1/2\Big) - g\Big(n^*\sigma_\rho^2 - 1/2\Big). \tag{5.2.38}$$

5.2.4 The one-mode Gaussian case

For a one-mode Gaussian system $s = 1$, recall from (5.1.22) and (5.1.25) that $|\Delta^{-1}\Sigma_\rho| = \sqrt{\det\{\Sigma_\rho\}}$. In this case, the problem reduces to

$$R(D) := \min_N \; g\left(\sqrt{\det\{\Sigma_\rho\}} - 1/2\right) - g\left(\sqrt{\det\{\Sigma_N\}} - 1/2\right), \tag{5.2.39}$$

$$\text{subject to } \operatorname{Sp}\left((\Sigma_Z^{1/2}(\Sigma_\rho - \Sigma_N)\Sigma_Z^{1/2})^{1/2}\right) \ge D_{\text{max}} - D. \tag{5.2.40}$$

We further change the variable of optimization to

$$X := \left(\Sigma_Z^{1/2}(\Sigma_\rho - \Sigma_N)\Sigma_Z^{1/2}\right)^{1/2}. \tag{5.2.41}$$

Thus, $\Sigma_N = \Sigma_\rho - \Sigma_Z^{-1/2} X^2 \Sigma_Z^{-1/2}$. Then the above problem is equivalent to

$$\min_X - g\left(\sqrt{\frac{\det\{\Sigma - X^2\}}{\det\{\Sigma_Z\}}} - \frac{1}{2}\right), \tag{5.2.42}$$

$$0 \preccurlyeq X^2, \quad \det\bigl(\Sigma - X^2\bigr) \geq \frac{1}{4}\det(\Sigma_Z), \tag{5.2.43}$$

$$c \leq \mathrm{Sp}\,(X)\,, \tag{5.2.44}$$

where $\Sigma := \Sigma_Z^{1/2}\Sigma_\rho\Sigma_Z^{1/2}$ and $c = D_{\max} - D$. The above optimization problem is convex and can be solved using KKT conditions. Note that inequalities of (5.2.43) are feasibility constraints, therefore, the Lagrangian function is of the form:

$$L(X) = -g\left(\sqrt{\frac{\det\{\Sigma - X^2\}}{\det\{\Sigma_Z\}}} - \frac{1}{2}\right) - \pi\mathrm{Sp}\,(X)\,. \tag{5.2.45}$$

We develop the KKT conditions,

$$\frac{1}{2}\phi(X)\cdot\log\frac{\phi(X)+1/2}{\phi(X)-1/2}\left[(\Sigma - X^2)^{-1}X + X(\Sigma - X^2)^{-1}\right] = \pi I \tag{5.2.46}$$

$$\pi(\mathrm{Sp}\,(X) - c) = 0, \quad \pi \geq 0 \tag{5.2.47}$$

where $\phi(X) := \sqrt{\frac{\det\{\Sigma - X^2\}}{\det\{\Sigma_Z\}}}$. Assuming the spectral decomposition of matrix $\Sigma = U\mathrm{diag}(\sigma_i^2)_{i=1:2}U^T$, we only consider the solutions of the form $X = U\mathrm{diag}(x_i)U^T$ which have the same diagonalization as Σ. This reduces the KKT conditions to the following system of equations:

$$\phi(X) \cdot \log \frac{\phi(X)+1/2}{\phi(X)-1/2} \frac{x_i}{\sigma_i^2 - x_i^2} = \pi, \quad \text{for } i = 1,2 \tag{5.2.48}$$

$$\sum_i x_i = c, \quad \pi \geq 0. \tag{5.2.49}$$

The term $\phi(X)$ does not depend on the index i, and therefore, it is just a scaling of π which can be absorbed in

$$\tilde{\pi} = \pi \left(\phi(X) \log \frac{\phi(X)-1/2}{\phi(X)+1/2} \right)^{-1} \tag{5.2.50}$$

This gives the final form of the solution,

$$\frac{x_i}{\sigma_i^2 - x_i^2} = \tilde{\pi}, \quad \text{for } i = 1,2 \tag{5.2.51}$$

$$\sum_i x_i = c, \quad \tilde{\pi} \geq 0. \tag{5.2.52}$$

It suffices to show that for any given $c_{\text{min}} < c < c_{\text{max}}$, there exists a solution for the above equations. Solving (5.2.48) for x_i gives

$$x_i = -\frac{1}{2\tilde{\pi}} + \sqrt{(\frac{1}{2\tilde{\pi}})^2 + \sigma_i^2}, \quad \text{for } i = 1,2, \tag{5.2.53}$$

which is always non-negative and hence always valid. Then by defining $y := 1/(2\tilde{\pi})$, we rewrite the second equation as a function of y,

$$f(y) := -ny + \sum_{i=1}^{n} \sqrt{y^2 + \sigma_i^2} = c, \tag{5.2.54}$$

where in this case $n = 2$. Recall that $c = D_{\text{max}} - D$ and $D \in (D_{\text{min}}, D_{\text{max}})$. Therefore, $c_{\text{min}} = 0$. The value of c_{max} corresponds to D_{min} which is the Wasserstein distance of the QC system, implied by the feasibility constraint (5.2.43):

$$\det\{\Sigma - X^2\} \geq \frac{1}{4}\det\{\Sigma_Z\}. \tag{5.2.55}$$

We next provide a loose upperbound $\hat{c}_{\text{max}} >> c_{\text{max}}$ by relaxing the constraint to $\det\{\Sigma - X^2\} \geq 0$. Therefore, $\hat{c}_{\text{max}} = \text{Sp}\left(\Sigma^{1/2}\right) = \sum_i \sigma_i$.

we have that $c_{\text{min}} = 0$ results in $y = \infty$ and $c_{\text{max}} = \sum_i \sigma_i$ results in $y = 0$, thus both have valid solutions. Further, $f(y)$ function is a monotonically decreasing function of y, therefore, there is a one-to-one correspondence between values of $c \in (c_{\text{min}}, \hat{c}_{\text{max}})$ and the values of $y \in (0, \infty)$. Thus, for any feasible value of $c \in (c_{\text{min}}, \hat{c}_{\text{max}})$, there exists a solution of the form $X = U\text{diag}(x_i)_{i=1,2}U^T$. This approach is similar to the solution in the classical case proved in [8].

5.2.5 The multi-mode Gaussian case with independent modes

We can easily extend the previous one-mode Gaussian system to s-mode independent Gaussian system, where source state is an s-mode Gaussian state with a block diagonal matrix $\Sigma_\rho = \oplus_{i=1}^s \Sigma_{\rho_i}$, where Σ_{ρ_i} is the covariance matrix of i-th mode. Similarly, we assume the output distribution to be a multivariate Gaussian system with s independent 2-dimensional Gaussian distributions comprising the covariance $\Sigma_Z = \oplus_{i=1}^s \Sigma_{Z_i}$. Therefore, in this case the optimization problem in (5.2.39) is generalized as follows

$$R(D) := \min_{\substack{\Sigma_{N_i},\\ i=1,\cdots,s}} \sum_{i=1}^{s} \left[g\left(\sqrt{\det\{\Sigma_{\rho_i}\}} - 1/2\right) - g\left(\sqrt{\det\{\Sigma_{N_i}\}} - 1/2\right)\right] \tag{5.2.56}$$

$$\text{subject to } \sum_{i=1}^{s} \frac{1}{s}\mathrm{Sp}\left((\Sigma_{Z_i}^{1/2}(\Sigma_{\rho_i} - \Sigma_{N_i})\Sigma_{Z_i}^{1/2})^{1/2}\right) \geq D_{\max} - D. \tag{5.2.57}$$

Again, similar to (5.2.41) we change the variables of optimization to

$$X_i := \left(\Sigma_{Z_i}^{1/2}(\Sigma_{\rho_i} - \Sigma_{N_i})\Sigma_{Z_i}^{1/2}\right)^{1/2} \quad \text{for } i = 1, \cdots, s. \tag{5.2.58}$$

This results in the following equivalent optimization problem

$$\min_{\substack{X_i,\\ i=1,\ldots,s}} -\sum_{i=1}^{s} g\left(\sqrt{\frac{\det\{\Sigma_i - X_i^2\}}{\det\{\Sigma_{Z_i}\}}} - \frac{1}{2}\right), \tag{5.2.59}$$

$$0 \preccurlyeq X_i^2, \quad \det\big(\Sigma_i - X_i^2\big) \geq \frac{1}{4}\det(\Sigma_{Z_i}), \tag{5.2.60}$$

$$c \leq \sum_i \mathrm{Sp}\left(X_i\right), \tag{5.2.61}$$

where $\Sigma_i := \Sigma_{Z_i}^{1/2}\Sigma_{\rho_i}\Sigma_{Z_i}^{1/2}$ and $c = D_{\max} - D$. The above problem is solved with a similar approach as one-mode case by considering X_i to be same diagonalizable as Σ_i, which results in an extension of the one-mode solution (5.2.51) as follows:

$$\phi(X_i) \cdot \log \frac{\phi(X_i) + 1/2}{\phi(X_i) - 1/2} \frac{x_{ij}}{\sigma_{ij}^2 - x_{ij}^2} = \pi, \quad \text{for } i = 1, \cdots, s, \quad j = 1, 2 \tag{5.2.62}$$

$$\sum_{i=1}^{s}\sum_{j=1}^{2} x_{ij} = c. \tag{5.2.63}$$

where for the i-th mode, $X_i = U_i \mathrm{diag}(x_{ij})_{j=1,2} U_i^T$ and $\Sigma_i = U_i \mathrm{diag}(\sigma_{ij}^2)_{j=1,2} U_i^T$.

5.2.6 QC Wasserstein distance (D_{min})

The other extreme point in the rate-limited Wasserstein distance curve is the Wasserstein distance itself (D_{min}), which is the point at which no constraints are applied to the rate of transmission. In the classical settings, the 2nd-order Wasserstein distance between Gaussian distributions occurs at $R = \infty$. However, in what follows we show that this is not the case in QC systems. The following optimization problem obtains the Wasserstein distance,

$$W_2^2(\rho||\pi_Z) = \min_X D_{\text{max}} - \text{Sp}\,(X) \tag{5.2.64}$$

$$\text{subject to:} \quad \Delta^{-1}\Sigma_N \geq \frac{1}{4}I \quad \text{or} \quad \det\{\Sigma - X^2\} \geq \frac{1}{4}\det\{\Sigma_Z\}. \tag{5.2.65}$$

This simplifies in the one-mode case to

$$\max_{x_i} \sum_i x_i \tag{5.2.66}$$

$$\text{subject to} \prod_{i=1}^{2}(\sigma_i^2 - x_i^2) \geq \frac{1}{4}\det\{\Sigma_Z\}. \tag{5.2.67}$$

This again is a convex function that has the following solution:

$$\begin{cases} \frac{1}{2\lambda} = x_2(x_1^2 - \sigma_1^2) = x_1(x_2^2 - \sigma_2^2) \\ (x_1^2 - \sigma_1^2)(x_2^2 - \sigma_2^2) = \frac{1}{4}\det\{\Sigma_Z\} \end{cases}. \tag{5.2.68}$$

And the minimum required rate for this QC Wasserstein distance is

$$R_{W_2} = g\left(\sqrt{\det\{\Sigma_\rho\}} - 1/2\right) \tag{5.2.69}$$

which is the von-Neumann entropy of the quantum source state.

5.2.7 Numerical Example of Quantum-Classical Gaussian system

We consider an example of the above system when the input state is a Gaussian quantum state and output system is a classical Gaussian distribution with covariance matrices

$$\Sigma_\rho = \begin{bmatrix} 1.0783 & -0.8976 \\ -0.8976 & 1.3155 \end{bmatrix}, \Sigma_Z = \begin{bmatrix} 1.7471 & -1.2224 \\ -1.2224 & 0.8583 \end{bmatrix}. \tag{5.2.70}$$

One can find the Symplectic eigenvalues of the input covariance matrix as $\alpha_s = 0.7828$ and the eigenvalues of the output matrix as $\lambda_{1,2} = 0.002, 2.6034$. The displacement is assumed to be zero for both source and destination. For this system, the suboptimal OC rate-distortion function generated by calculating the noise covariance from the above expressions is given in Figure 5.1. The interesting observation is that in contrast to the classical Gaussian optimal transport for which the Wasserstein distance is achieved when the information rate is infinite $(R = \infty)$, in the QC setting, due to the Heisenberg Uncertainty principle, the maximum rate cannot be infinite as the accessible information of measurement is limited according to Theorem 2 of [30]. The simulations are performed using Strawberry fields [35] and Walrus [28] packages and Matlab.

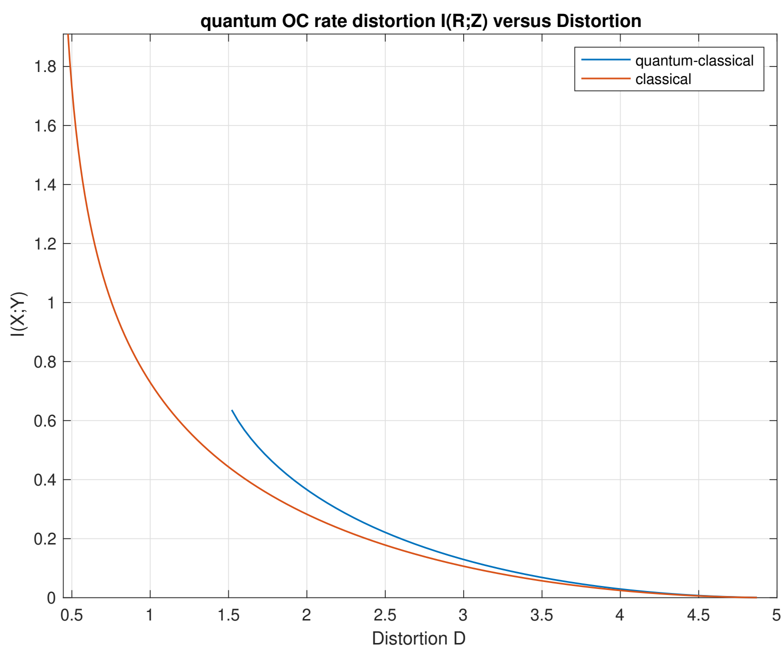

Figure 5.1: Output Constrained Rate-Distortion function of the Gaussian system (5.2.70). The inverse of this plot is the rate-limited QC Wasserstein distance of 2nd order.

Chapter 6

Discussion and Conclusion

6.1 Discussion

We would also like to mention some key differences between our source coding theorem and other works. Specifically, in comparison to [17], our system has the additional constraint that the output must follow a predetermined distribution in the exact i.i.d. format. This provides a multi-letter protocol that governs the optimal transportation of a quantum source state to a target distribution, through a rate-limited classical channel with limited common randomness. In contrast to the conventional rate-distortion theorem for which the common randomness provides no performance improvements, in this problem, the common randomness can help reduce the communication rate by providing the extra randomness required to ensure the output has the desired i.i.d. distribution.

The proof of the discrete coding theorem builds on analytical tools such as Winter's measurement compression theorem [53], the non-feedback measurement simulation [38] and the batch post-processing of [49]. In contrast to their work where a quantum-classical channel is being faithfully simulated in a nearly perfect sense, in this work, we ensure the output is following the desired distribution in the perfect i.i.d. format while maintaining the distortion threshold. Moreover, the analysis of the continuous-variable quantum systems is also one of the key contributions of this paper as the proofs of the discrete theorems do not directly apply to the continuous case as was discussed in Section 3.

Consider an example where we have a sequence of product quantum Gaussian states. Suppose we want to store these states' information in a classical memory system for later use. Our goal is to prepare quantum states that are as similar as possible to the original source states, with the constraint that these states also need to be product Gaussian states. In this scenario, our theorem can help by first estimating the amount of quantum data lost due to the entanglement-breaking channel in the form of the minimum distortion from the source. It also can calculate the required storage space in the classical memory system.

6.2 Conclusion

We introduced the output-constrained lossy source coding with limited classical common randomness for the quantum-to-classical systems. We further used this source coding scheme to establish the concept of rate-limited quantum-to-classical optimal transport. The theorem provides a computable single-letter characterization of the achievable rate region (R, R_c) to fulfill the distortion level in accordance with a generally defined form of distortion observable.

Next, the example of qubit systems with unlimited common randomness was analytically evaluated. The analytical expression for the rate-limited QC optimal transport was provided in the form of a transcendental equation. Furthermore, the minimum achievable transportation cost in case of an unlimited communication rate and common randomness (the lowest possible distortion in the OC rate-distortion curve) was provided using analytical expressions, which show that in the case of the source qubit state having a diagonal density operator along the canonical eigenbasis and using the entanglement fidelity distortion measure, the optimal transport scheme recover the classical optimal transport scheme for the minimum transportation cost.

Moreover, we extended this theorem to the continuous-variable quantum systems with the help of an alternative continuous coding protocol. Finally, using the continuous coding theorem we evaluated the rate-limited Wasserstein distance of 2nd order for the Gaussian quantum systems under the presence of the unlimited amount of common randomness. The proof is built upon a Gaussian measurement optimality theorem which states that for a measurement system with Gaussian quantum source state and Gaussian output distribution, Gaussian measurement POVM minimizes the Groenwold's information while maintaining the distortion constraint.

The QC rate-limited 2-Wasserstein distance plot shows that, unlike the classical optimal transport for which the rate can grow to infinity, in the QC system, the rate has a finite limit due to Heisenberg's uncertainty principle.

In future works, we aim at finding the rate-limited optimal transport for the variants of this system, for example, the QC measurement optimal transport with quantum side information. Moreover, the optimal transport coding theorems can also be provided for the case of fully quantum channels which would generalize the Quantum Wasserstein distance in [19]. In addition to this, the case of one-shot transmission is of particular interest which will provide more practical applications of the subject.

Appendix A

Proof of Theorem 4.1

The optimization problem in (4.1.3) is equivalent to the following entropy maximization problem

$$
\begin{aligned}
\max_{M_0,M_1} q_0 H(\rho_0) + q_1 H(\rho_1) \\
\text{s.t.} \quad q_0 + q_1 &= 1 \\
\operatorname{Tr}\{M_x \rho\} &= q_x, \;\; x \in \{0,1\} \\
\rho_x &= \sqrt{\rho} M_x \sqrt{\rho}/q_x, \;\; x \in \{0,1\} \\
\rho_0 q_0 + \rho_1 q_1 &= \rho \\
\langle 0| \rho M_0 \rho |0\rangle + \langle 1| \rho M_1 \rho |1\rangle &\geq 1 - D \\
M_0 + M_1 &= I \\
M_0, M_1 &\geq 0
\end{aligned}
$$

where $H(\rho) = -\operatorname{Tr}\{\rho \ln \rho\}$ is the von-Neumann entropy function. Note that although the distortion formula in (4.1.2) has transposed POVM operator $M_0^{\mathcal{T}_\varphi}$, we can ignore this transpose in the distortion constraint of the above optimization problem. The reason is that the eigenvalues of ρ_x are preserved under the transposition with respect to any basis. This implies that the entropy function also does not change under transposition,

$$H(\rho_x) = H(\sqrt{\rho} M_0^{\mathcal{T}} \sqrt{\rho}).$$

Therefore, we may remove the transpose from all the terms in the optimization problem. Then, by defining $N = \sqrt{\rho} M_0 \sqrt{\rho}$ the conditional post-measurement reference state given outcome zero, as the variable of optimization, the optimization problem reduces to the following standard form

$$\min_{N} \quad \mathrm{Tr}\{N \ln(N/q_0)\} + \mathrm{Tr}\left\{(\rho - N) \ln\left(\frac{\rho - N}{1 - q_0}\right)\right\}, \tag{A.0.1}$$

$$\text{s.t.,} \quad \langle 0| \sqrt{\rho} N \sqrt{\rho} |0\rangle + \langle 1| \sqrt{\rho}(\rho - N)\sqrt{\rho} |1\rangle \geq 1 - D, \tag{A.0.2}$$

$$\mathrm{Tr}\{N\} = q_0, \tag{A.0.3}$$

$$0 \preceq N \preceq \rho, \tag{A.0.4}$$

where ρ is the input state and $0 \leq q_0 \leq 1$ is the zero-output probability, which are the given parameters of the problem. This is a convex optimization problem, as the objective function comprises negative entropy functions that are concave, and the constraint is linear. The distortion constraint (A.0.2) is further simplified to

$$\mathrm{Tr}\{NG\} \geq 1 - D - \left\langle 1|\rho^2|1 \right\rangle, \tag{A.0.5}$$

where

$$G := \sqrt{\rho} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \sqrt{\rho}. \tag{A.0.6}$$

A.1 Zero-Crossing Point

To find the zero crossing point, we first ignore the distortion constraint and find the optimal operator when $D \to \infty$. Thus, simply taking the derivative of the objective function and equating it to zero gives

$$\begin{aligned} \frac{df(N)}{dN} &= \frac{d}{dN}\left[\mathrm{Tr}\left\{ N \ln \frac{N}{q_0} \right\} + \mathrm{Tr}\left\{ (\rho - N) \ln \frac{\rho - N}{1 - q_0} \right\} \right] \\ &= \left(I + \ln \frac{N}{q_0} \right)^T - \left(I + \ln \frac{\rho - N}{1 - q_0} \right)^T := 0 \\ \implies \ln \frac{N}{q_0} &= \ln \frac{\rho - N}{1 - q_0} \\ \implies N &= q_0 \rho \end{aligned}$$

This implies the measurement POVMs $M_0 = q_0 I,\ M_1 = (1 - q_0) I$ in the infinite-distortion case. Substituting these values into the mutual information expression gives $I(R; X) = 0$ which means the communication rate will be zero and the input and output will be independent, which is intuitively acceptable. This is also observed from the M_0, M_1 which both apply the same identity operator on the source state regardless of the outcome. Substituting this $N = q_0\rho$ in the distortion constraint gives the zero-crossing point

$$D_{R_0} := 1 - q_0 \left\langle 0 \middle| \rho^2 \middle| 0 \right\rangle - (1 - q_0) \left\langle 1 \middle| \rho^2 \middle| 1 \right\rangle .$$

Thus, for all values of entanglement fidelity distortion in the range $D \geq D_{R_0}$, the output and reference state are independent so the rate is zero.

A.2 Non-Zero Rate Region

Next, in the non-zero region $D_{OT} \leq D \leq D_{R_0}$, the distortion constraint is active. Recall that the convex problem (A.0.1) is on the domain of positive semi-definite Hermitian matrices $N \in \mathcal{S}_+^n$. Also the trace of N_{opt} is constrained to be $\mathrm{Tr}\{N\} = q_0$. So the states are shown in the expanded matrix form

$$\rho = \begin{bmatrix} \rho_1 & \rho_2 \\ \rho_2^* & 1 - \rho_1 \end{bmatrix}, \quad N = \begin{bmatrix} n & n_{\text{off}} \\ n_{\text{off}}^* & q_0 - n \end{bmatrix},$$

where $n, \rho_1 \in \mathbb{R}^+$ and $n_{\text{off}}, \rho_2 \in \mathbb{C}$. Further, the PSD constraints of (A.0.4) reduce to $\det(N) \geq 0$ and $\det(\rho - N) \geq 0$. Thus we have

$$|n_{\text{off}}|^2 \leq n(q_0 - n), \tag{A.2.1}$$

$$|\rho_2 - n_{\text{off}}|^2 \leq (\rho_1 - n)(1 - \rho_1 - q_0 + n). \tag{A.2.2}$$

Next, examining the objective function (A.0.1) and constraints shows a symmetry among the value of n_{off}. Note that the first term in (A.0.1) is a function of eigenvalues of matrix N which only depends on $|n_{\text{off}}|$. Similarly, the second term is only a function of $|\rho_2 - n_{\text{off}}|$. These two expressions can be illustrated as two circles in the complex plane of n_{off}. These circles create a symmetry w.r.t. the direction of vector ρ_2 in this complex plane. Moreover, we can rewrite the LHS of distortion constraint (A.0.5) as

$$\begin{aligned}\text{Tr}\{NG\} &= \text{Tr}\left\{\begin{bmatrix} n & n_{\text{off}} \\ n_{\text{off}}^* & q_0 - n \end{bmatrix}\begin{bmatrix} g_1 & g_2\rho_2/|\rho_2| \\ g_2\rho_2^*/|\rho_2| & g_3 \end{bmatrix}\right\} \\ &= (g_1 - g_3)n + \frac{2g_2}{|\rho_2|}\text{Re}\{n_{\text{off}}\,\rho_2^*\} + q_0 g_3.\end{aligned}$$

By expanding the term $\text{Re}\{n_{\text{off}}\,\rho_2^*\}$ in above expression, the distortion function is

$$\text{Re}\{n_{\text{off}}\}\,\text{Re}\{\rho_2\} + \text{Im}\{n_{\text{off}}\}\,\text{Im}\{\rho_2\} + f(n, q_0, \rho) \geq 0,$$

where $f(n, q_0, \rho)$ is a function comprised of the remaining terms. This is a half-plane in the complex plane of n_{off}, with its slope orthogonal to the slope of symmetry line ρ_2. Therefore, the objective function and all the constraints are symmetric w.r.t. the line ρ_2 in the complex plane. In view of the fact that the problem is a convex program, the solution must occur on the line of symmetry. This means the optimal solution has the shape

$$N_{opt} = \begin{bmatrix} n & s\rho_2/|\rho_2| \\ s\rho_2^*/|\rho_2| & q_0 - n \end{bmatrix}, \tag{A.2.3}$$

where $n, s \in \mathbb{R}$ are the variables of optimization.

The active distortion constraint is also expanded in the form

$$\begin{aligned} &\mathrm{Tr}\{NG\} = 1 - D - \langle 1|\rho^2|1\rangle \\ \Longrightarrow &\Big(g_1 - g_3\Big)n + \Big(2g_2\Big)s + \Big(q_0 g_3 + \langle 1|\rho^2|1\rangle - 1 + D\Big) = 0, \end{aligned} \tag{A.2.4}$$

where G is defined in (A.0.6) and its elements are obtained by

$$G := \begin{bmatrix} g_1 & \frac{g_2}{|\rho_2|}\rho_2 \\ \frac{g_2}{|\rho_2|}\rho_2^* & g_3 \end{bmatrix} = \frac{1}{1+2k}\begin{bmatrix} (\rho_1 + k)^2 - |\rho_2|^2 & (2\rho_1 - 1)\rho_2 \\ b(2\rho_1 - 1)\rho_2^* & |\rho_2|^2 - (1 - \rho_1 + k)^2 \end{bmatrix},$$

with $k := \sqrt{\det\{\rho\}}$. Therefore, expanding the objective function yields the following scalar convex optimization problem

$$\begin{aligned} &\min_{n,s} \sum_{i=1,2} \lambda_{N_i}(n,s)\ln(\lambda_{N_i}(n,s)/q_0) + \sum_{i=1,2} \lambda_{d_i}(n,s)\ln(\lambda_{d_i}(n,s)/(1-q_0)) \\ &\text{s.t. } \Big(g_1 - g_3\Big)n + \Big(2g_2\Big)s + \Big(q_0 g_3 + \langle 1|\rho^2|1\rangle - 1 + D\Big) = 0 \end{aligned}$$

where $n, s \in \mathbb{R}$. The terms λ_{N_i} and λ_{d_i} are the eigenvalues of N and $\rho - N$ respectively, obtained by

$$\begin{aligned} \lambda_{N_1,N_2} &= \frac{q_0}{2} \pm E_1(n,s), \\ \lambda_{d_1,d_2} &= \frac{1-q_0}{2} \pm E_2(n,s), \end{aligned}$$

where $E_1(n, s)$ and $E_2(n, s)$ are as defined in (4.1.8), (4.1.9).

A.2.1 The Transcendental Equation of Optimal Solution

We solve by substituting the linear constraint into the objective function, then taking the derivative w.r.t. n and equating it to zero. The final solution becomes in the form of the following transcendental system of equations of n, s which provides the optimal values n_{opt} and s_{opt}:

$$\frac{-as + b(n - q_0/2)}{E_1} \ln \frac{q_0/2 + E_1}{q_0/2 - E_1} + \frac{-a(s - |\rho_2|) + b(n - \rho_1 + \frac{1-q_0}{2})}{E_2} \ln \frac{\frac{1-q_0}{2} + E_2}{\frac{1-q_0}{2} - E_2} = 0$$

$$an + bs + c = 0$$

where $a := g_1 - g_3$, $b := 2g_2$ and $c := q_0 g_3 + \langle 1|\rho^2|1\rangle - 1 + D$ are the parameters of the linear distortion constraint and are given by

$$a = 1 - \frac{4|\rho_2|^2}{1+2k},$$
$$b = \frac{2|\rho_2|(2\rho_1 - 1)}{1+2k},$$
$$g_3 = \rho_1 - 1 + \frac{2|\rho_2|^2}{1+2k}.$$

Based on the optimal values n_{opt}, s_{opt}, the optimal POVM operators are

$$M_{0,\text{opt}} = \sqrt{\rho}^{-1} N_{\text{opt}} \sqrt{\rho}^{-1}, \quad M_{1,\text{opt}} = I - M_0,$$
$$N_{\text{opt}} = \begin{bmatrix} n_{opt} & s_{opt} . \, \rho_2/|\rho_2| \\ s_{opt} . \, \rho_2^*/|\rho_2| & q_0 - n_{opt} \end{bmatrix}.$$

Appendix B

Proof of Theorem 4.2

To find the optimal transportation cost D_{OT} of the Qubit system, we assume having an unlimited available rate and find the minimum possible distortion. In this case, the problem reduces to

$$\begin{aligned} \min_{M_0, M_1} D_{ef} := 1 - \sum_x \langle x | \rho M_x \rho | x \rangle \, , \\ \text{s.t.,} \quad M_0, M_1 \geq 0, \\ M_0 + M_1 = I, \\ \mathrm{Tr}\{M_0 \rho\} = q_0. \end{aligned}$$

Note that we used the same argument as in Appendix A to remove the transpose operator from the formulations. Then again using the change of variable $N := \sqrt{\rho} M_0 \sqrt{\rho}$ reduces the problem to the following semi-definite programming, whose result is the operator for optimal transport N_{OT}

$$\begin{aligned} \min_{N} 1 - \langle 1|\rho^2|1\rangle - \mathrm{Tr}\{NG\}, \\ \text{s.t.,} \quad 0 \preceq N \preceq \rho, \\ \mathrm{Tr}\{N\} = q_0. \end{aligned}$$

Similar to the rate-distortion problem in the previous section, the symmetry implies N_{OT} to be in the form of (A.2.3). With this assumption, the problem reduces to the scalar optimization below:

$$\begin{aligned} &\min_{n,s} f_0(n,s) := -(g_1 - g_3)n - (2g_2)s + 1 - q_0 g_3 - \langle 1|\rho^2|1\rangle, \\ &\text{s.t.} \;\; f_1(n,s) := \left(n - \frac{q_0}{2}\right)^2 + s^2 < \left(\frac{q_0}{2}\right)^2, \\ &\quad\;\; f_2(n,s) := \left(n - \left(\rho_1 - \frac{1-q_0}{2}\right)\right)^2 + (s - |\rho_2|)^2 < \left(\frac{1-q_0}{2}\right)^2. \end{aligned}$$

This is quadratic-constrained linear programming which is a convex problem. By forming the Lagrangian function

$$\mathcal{L}(n,s,\lambda_1,\lambda_2) = f_0(n,s) + \lambda_1 f_1(n,s) + \lambda_2 f_2(n,s), \qquad \lambda_{1,2} \geq 0$$

taking the partial derivatives with respect to n, s and equating to zero we obtain the optimal variables as a function of λ_i Lagrange multipliers,

$$\begin{aligned} n_{opt}(\lambda_1,\lambda_2) &= \frac{a + q_0\lambda_1 + 2(\rho_1 - \frac{1-q_0}{2})\lambda_2}{2(\lambda_1+\lambda_2)}, \\ s_{opt}(\lambda_1,\lambda_2) &= \frac{b + 2\lambda_2|\rho_2|}{2(\lambda_1+\lambda_2)}. \end{aligned}$$

Then we substitute above expressions in the following complementary slackness conditions

$$\lambda_1 f_1\Big(n_{opt}(\lambda_1,\lambda_2),\ s_{opt}(\lambda_1,\lambda_2)\Big) = 0,$$
$$\lambda_2 f_2\Big(n_{opt}(\lambda_1,\lambda_2),\ s_{opt}(\lambda_1,\lambda_2)\Big) = 0,$$
$$\lambda_{1,2} \geq 0,$$

which result in the following scenarios.

1. Minimum happens at f_2 circle ($\lambda_1 = 0,\ \lambda_2 \neq 0$):

   Because $\lambda_2 \neq 0$, then $f_2(n_{opt}, s_{opt}) = 0$. This results in

   $$\lambda_2 = \frac{\sqrt{a^2+b^2}}{1-q_0}.$$

   The corresponding optimal values for variables are obtained as

   $$\begin{aligned} s_{opt} &= \frac{b(1-q_0)}{2\sqrt{a^2+b^2}} + |\rho_2| = \frac{b}{\sqrt{1-4|\rho_2|^2}}\frac{1-q_0}{2} + |\rho_2|, \\ n_{opt} &= \frac{a(1-q_0)}{2\sqrt{a^2+b^2}} + \rho_1 - \frac{1-q_0}{2} = \left(\frac{a}{\sqrt{1-4|\rho_2|^2}} - 1\right)\frac{1-q_0}{2} + \rho_1, \\ D_{OT} &= 1 - q_0 g_3 - \langle 1|\rho^2|1\rangle - \frac{1-q_0}{2}\sqrt{a^2+b^2} - a(\rho_1 - \frac{1-q_0}{2}) - b|\rho_2| \\ &= q_0(1-\rho_1) + \det(\rho) + \frac{1-q_0}{2}\left(1 - \sqrt{1-4|\rho_2|^2}\right). \end{aligned}$$

   Note that $a^2+b^2 = 1-4|\rho_2|^2$. Also, substituting the above values into the feasibility condition for $f_2(n_{opt}, s_{opt}) \leq 0$ provides the conditions required for this scenario:

$$\frac{1-q_0}{2} + \frac{1-q_0}{\sqrt{a^2+b^2}} \left(a(\rho_1 - \frac{1}{2}) + b|\rho_2| \right) \le \det(\rho).$$

2. Minimum happens at f_1 circle ($\lambda_1 \neq 0, \lambda_2 = 0$):

   Because $\lambda_1 \neq 0$, then $f_1(n_{opt}, s_{opt}) = 0$. This results in

$$\lambda_1 = \frac{\sqrt{a^2+b^2}}{q_0}.$$

   The corresponding optimal values for variables are obtained as

$$\begin{aligned} s_{opt} &= \frac{bq_0}{2\sqrt{a^2+b^2}} = \frac{b}{\sqrt{1-4|\rho_2|^2}} \frac{q_0}{2}, \\ n_{opt} &= \frac{aq_0}{2\sqrt{a^2+b^2}} + \frac{q_0}{2} = \left(\frac{a}{\sqrt{1-4|\rho_2|^2}} + 1 \right) \frac{q_0}{2}, \\ D_{OT} &= 1 - q_0 g_3 - \langle 1|\rho^2|1 \rangle - \frac{q_0}{2}\sqrt{a^2+b^2} - \frac{aq_0}{2} \\ &= (1-q_0)\rho_1 + \det(\rho) + \frac{q_0}{2} \left(1 - \sqrt{1-4|\rho_2|^2} \right). \end{aligned}$$

   Also, substituting the above values into the feasibility condition for $f_2(n_{opt}, s_{opt}) \le 0$ provides the conditions required for this scenario:

$$\frac{q_0}{2} - \frac{q_0}{\sqrt{a^2+b^2}} \left(a(\rho_1 - \frac{1}{2}) + b|\rho_2| \right) \le \det(\rho).$$

3. Minimum happens at intersection of circles ($\lambda_1 \neq 0, \lambda_2 \neq 0$):

   If neither of the conditions of the above scenarios is satisfied, then the result happens at the intersection of two circles

$$\left(n - \frac{q_0}{2}\right)^2 + s^2 = \left(\frac{q_0}{2}\right)^2,$$
$$\left(n - \left(\rho_1 - \frac{1-q_0}{2}\right)\right)^2 + (s - |\rho_2|)^2 = \left(\frac{1-q_0}{2}\right)^2,$$

which is obtained by the following expressions under separate conditions:

- If $a \geq b$,

$$s_{opt} = \frac{2B + q_0 A + A\sqrt{\Delta}}{2(1+A^2)},$$
$$n_{opt} = \frac{q_0 - 2AB + \sqrt{\Delta}}{2(1+A^2)}.$$

- If $a < b$,

$$s_{opt} = \frac{2B + q_0 A - A\sqrt{\Delta}}{2(1+A^2)},$$
$$n_{opt} = \frac{q_0 - 2AB - \sqrt{\Delta}}{2(1+A^2)}.$$

where

$$A := \frac{1 - 2\rho_1}{2|\rho_2|},$$
$$B := \frac{\rho_1 q_0 - \det(\rho)}{2|\rho_2|},$$
$$\Delta := q_0^2 - 4B^2 - 4q_0 AB.$$

Then the optimal transport value can be obtained by substituting the above results in the expression below:

$$D_{OT} = 1 - q_0 g_3 - \left\langle 1 | \rho^2 \middle| 1 \right\rangle - a n_{opt} - b s_{opt}.$$

Appendix C

Proof of Useful Lemmas

C.1 Proof of Lemma 2.2

The proof appeals to the time-sharing method similar to the proof of Lemma 1 of [45]. We know that by definition 2.5, the rate-distortion function $R(D; R_c, \rho||Q_X)$ is the infimum of the achievable rates for fixed (R_c, D). Also by using the time-sharing argument, we know that $\mathcal{R}(D, \rho||Q_X)$ is a connected interval. Therefore, having two distortion threshold values D_1 and D_2 and a value of $\alpha \in (0, 1)$, by definition, for any $\delta > 0$ there exists an achievable rate R_i in the range

$$R(D_i; R_c, \rho||Q_X) \leq R_i \leq R(D_i; R_c, \rho||Q_X) + \delta, \quad i = 1, 2.$$

In other words, there exist coding schemes (n, R_1, R_c) and (n, R_2, R_c) with fixed Q_X^n output distribution satisfying

$$\mathrm{Tr}\big\{\Delta^{(n)}(\mathcal{D}_{(i)}^n \circ \mathcal{M}_{(i)}^n)(\phi_\rho^{RA})^{\otimes n}\big\} \leq D_i + \delta, \quad i = 1, 2$$

where $\mathcal{M}^n_{(i)}$ and $\mathcal{D}^n_{(i)}$ are the measurement and decoder of each source code respectively. Next, we define a mega-block of source code with size nN where n is the size of each inner block and N is the number of inner blocks. Then we set up the system such that the first $k_N \in \mathbb{N}$ blocks use the first coding scheme, and the rest use the second coding scheme as

$$\mathcal{M}^{nN} := \Bigg(\underbrace{\mathcal{M}^n_{(1)}, ..., \mathcal{M}^n_{(1)}}_{k_N\text{-times}} \underbrace{\mathcal{M}^n_{(2)}, ..., \mathcal{M}^n_{(2)}}_{N-k_N\text{-times}} \Bigg), \quad \mathcal{D}^{nN} := \Bigg(\underbrace{\mathcal{D}^n_{(1)}, ..., \mathcal{D}^n_{(1)}}_{k_N\text{-times}} \underbrace{\mathcal{D}^n_{(2)}, ..., \mathcal{D}^n_{(2)}}_{N-k_N\text{-times}} \Bigg).$$

Also, set k_N such that $\lim_{M\to\infty} \frac{k_M}{M} = \alpha$. The new source coding has a rate of

$$\begin{aligned} R_{\text{new}} &= \frac{k_N}{N} R_1 + \frac{N-k_N}{N} R_2 \\ &\le \frac{k_N}{N} R(D_1; R_c, \rho||Q_X) + \frac{N-k_N}{N} R(D_2; R_c, \rho||Q_X) + \delta, \end{aligned}$$

with distortion level

$$D_{\text{new}} = \text{Tr}\Big\{ \big(\Delta^{(n)}\big)^{\otimes N} (\mathcal{D}^{nN} \circ \mathcal{M}^{nN})(\phi^{RA}_\rho)^{\otimes nN} \Big\} \le \frac{K_N}{N} D_1 + \frac{N-k_N}{N} D_2 + \delta.$$

Therefore, by assigning proper values to N and δ, we ensure that

$$\begin{aligned} R_{\text{new}} &\le \alpha R(D_1; R_c, \rho||Q_X) + (1-\alpha) R(D_2; R_c, \rho||Q_X) + \epsilon, \\ D_{\text{new}} &\le \alpha D_1 + (1-\alpha) D_2 + \epsilon, \end{aligned}$$

for the fixed ϵ of the definition 2.3. This new coding scheme is achievable as it is the time-sharing between two achievable source codes. Then, according to the definition of the achievable rate region $\mathcal{R}(D, \rho||Q_X)$, the minimum rate of new coding is bounded by

$$R(D_{\text{new}}; R_c, \rho||Q_X) \leq R_{\text{new}}.$$

Substituting the upperbounds into the above inequality gives

$$R\bigg(\alpha D_1 + (1-\alpha)D_2; R_c, \rho||Q_X,\bigg) \leq \alpha R(D_1; R_c, \rho||Q_X) + (1-\alpha)R(D_2; R_c, \rho||Q_X) + \epsilon.$$

Since ϵ can be arbitrarily small, then this inequality converges to the exact convex inequality. Thus, the proof holds.

C.2 Proof of Lemma 3.5

By definition of the weak convergence of operators in [[29] section 11.1], it suffices to show that for each subset $B \in \mathcal{B}$ and any two arbitrary states $\phi, \psi \in \mathcal{H}$, the inner-product converges as

$$\lim_{k_1 \to \infty} \left\langle \psi \middle| \hat{M}_{k_1}(B) \middle| \phi \right\rangle = \langle \phi | M(B) | \psi \rangle .$$

Starting with $I - \Pi_{k_1}$, we show that this operator vanishes to zero when $k_1 \to \infty$, for any $\psi, \phi \in \mathcal{H}$

$$\begin{aligned} \lim_{k_1 \to \infty} \Big| \langle \phi | (I - \Pi_{k_1}) | \psi \rangle \Big| &\leq \lim_{k_1 \to \infty} \| (I - \Pi_{k_1}) \ket{\psi} \|_2 . \| (I - \Pi_{k_1}) \ket{\phi} \|_2 \\ &= \lim_{k_1 \to \infty} \sqrt{\langle \phi | (I - \Pi_{k_1}) | \phi \rangle} \cdot \sqrt{\langle \psi | (I - \Pi_{k_1}) | \psi \rangle}, \end{aligned}$$

where the inequality appeals to Cauchy-Schwartz's inequality for Hilbert space. By using the definition of the projector operator and Bessel's inequality and the fact that $|n\rangle$ are orthonormal set, we have

$$\langle\phi|\Pi_{k_1}|\phi\rangle = \sum_{n=1}^{k_1} |\langle n|\phi\rangle|^2 \leq \|\phi\|^2.$$

The above inequality changes to equality when the orthonormal set is a complete orthonormal basis (Parseval's identity), which is when $k_1 \to \infty$. Substituting this into the inner-product expression proves that

$$\lim_{k_1\to\infty} \Big| \langle\phi|(I - \Pi_{k_1})|\psi\rangle \Big| = 0.$$

Moreover, for the other operators, we have

$$\begin{aligned}\lim_{k_1\to\infty} \langle\phi|M(B)\Pi_{k_1}|\psi\rangle &= \lim_{k_1\to\infty} \langle\phi'|\Pi_{k_1}|\psi\rangle \\ &= \langle\phi'|\psi\rangle \\ &= \langle\phi|M(B)|\psi\rangle\,,\end{aligned}$$

where $|\phi'\rangle = M(B)\,|\phi\rangle$, and the argument is similar to previous operator. These together prove that the sequence of POVMs $\hat{M}_{k_1}$ weakly converge to the M POVM.

C.3 Proof of Lemma 3.6

We condition the average n-letter distortion on the sequence of clipping errors $A_{k_1}^n$ as follows:

$$
\begin{aligned}
d_n(R^n, \hat{X}_{k_1,K_2}^n | \neg E_{ce}) &= \mathbb{E}_{A_{k_1}^n | \neg E_{ce}} \left[d_n(R^n, \hat{X}_{k_1,K_2}^n | A_{k_1}^n) \right] \\
&= \mathbb{E}_{A_{k_1}^n | \neg E_{ce}} \left[\frac{1}{n} \sum_{i=1}^{n} d(R_i, \hat{X}_{i,k_1,K_2} | A_{k_1}^n) \right] \\
&= \mathbb{E}_{A_{k_1}^n | \neg E_{ce}} \left[\frac{1}{n} \sum_{i:A_{i,k_1}=1} d(R_i, \hat{X}_{i,k_1,K_2} | A_{k_1}^n) + \frac{1}{n} \sum_{i:A_{i,k_1}=0} d(R_i, \hat{X}_{i,k_1,K_2} | A_{k_1}^n) \right] \\
&= \mathbb{E}_{T | \neg E_{ce}} \left[\frac{n-T}{n} d(R, \hat{X}_{\text{local}} | A_{k_1} = 1) + \frac{T}{n} d_T(R_{k_1}^T, \hat{X}_{k_1,K_2}^T) \right] \\
&\leq (P_{k_1} + \epsilon_{cl}) d\left(R, \hat{X}_{\text{local}} | A_{k_1} = 1\right) + \mathbb{E}_{T | \neg E_{ce}} \left[d_T(R_{k_1}^T, \hat{X}_{k_1,K_2}^T) \right]
\end{aligned} \tag{C.3.1}
$$

where in the last equality, we generate a local random value $\hat{X}_{\text{local}}$ at Bob's side for any sample with an asserted error bit $A_{i,k_1} = 1$. Then for any fixed $T = t$, we expand the second term inside the expectation, by adding and removing an intermediate term

$$
\begin{aligned}
&\lim_{\substack{n\to\infty, \\ t_{\min}\to\infty}} d_t(R_{k_1}^t, \hat{X}_{k_1,K_2}^t) \\
&= \lim_{\substack{n\to\infty, \\ t_{\min}\to\infty}} d_t(R_{k_1}^t, {X'}_{k_1,K_2}^t) + \left(d_t(R_{k_1}^t, \hat{X}_{k_1,K_2}^t) - d_t(R_{k_1}^t, {X'}_{k_1,K_2}^t) \right) \\
&\leq d(R_{k_1}, X'_{k_1,K_2}) + \lim_{\substack{n\to\infty, \\ t_{\min}\to\infty}} \left(d_t(R_{k_1}^t, \hat{X}_{k_1,K_2}^t) - d_t(R_{k_1}^t, {X'}_{k_1,K_2}^t) \right).
\end{aligned} \tag{C.3.2}
$$

The first term in inequality follows for sufficiently large t from the discrete rate-distortion theorem. Note that we have coupled the parameters n, k_1, ϵ_{cl} in a way that the value of t_{min} is sufficiently large.

The second part of (C.3.2) is the distortion caused by the optimal transport block in the receiver. However, unlike the classical case, we cannot use triangle inequality in this system. Therefore, we expand the distortions for each one and find the difference. Thus, assume that $\{\Lambda_{x^t}^{\mathcal{X}_{K_2}^t}\}_{x^t \in \mathcal{X}_{K_2}^t}$ is the t-collective measurement of the t-letter discrete rate-distortion coding which according to discrete coding theorem generates $\tilde{X}^t$ with perfect i.i.d. pmf $\mu_{X_{K_2}}$. We also define the continuous measurement POVM $\hat{\Lambda} \equiv \{\hat{\Lambda}(B), B \in \mathcal{B}(\mathcal{X}^t)\}$ which combines the discrete measurement coding with the output stage optimal transport block. For any event $A \subseteq \mathcal{B}(\mathcal{X}^t)$ we define

$$\hat{\Lambda}(A) := \sum_{x^t \in \mathcal{X}_{K_2}^t} \Lambda_{x^t}^{\mathcal{X}_{K_2}^t} \pi_{OT}^t\Big(A\Big|x^t\Big),$$

where $\pi_{OT}(.|x)$ for $x \in \mathcal{X}_{K_2}$ is the optimal transport channel from the discrete space to the continuous space. Specifically, as the discrete coding produces i.i.d. discrete pmf $\mu_{X_{K_2}}$ and the final output is required to have i.i.d. continuous distribution μ_X, then the optimal transport is a simple dequantization channel given for any event $A \subseteq \mathcal{B}(\mathcal{X}^t)$ by

$$\pi_{OT}^t\Big(A\Big|x^t\Big) = \frac{\mu_X^t(A \cap RQ_{K_2}(x^t))}{\mu_X^t\left(\mathcal{RQ}_{K_2}(x^t)\right)}, \tag{C.3.3}$$

where $\mathcal{RQ}_{K_2}(x^t)$ is the quantization region of $x^t \in \mathcal{X}_{K_2}^t$. Next, the i-th letter distortion is expanded as follows:

$$\begin{aligned} d(R_{i,k_1}, X'_{i,k_1,K_2}) &= \sum_{x^t \in \mathcal{X}_{K_2}^t} \operatorname{Tr}\Big\{ \operatorname{Tr}_{[t]\setminus i} \Big\{ \omega_{k_1} \Lambda_{x^t}^{\mathcal{X}_{K_2}^t} \omega_{k_1} \Big\} \Delta(x_i) \Big\} \\ &= \sum_{x_i \in \mathcal{X}_{K_2}} \operatorname{Tr}\left\{ \operatorname{Tr}_{[t]\setminus i} \left\{ \omega_{k_1} \left(\sum_{x^{[t]\setminus i} \in \mathcal{X}_{K_2}^{t-1}} \Lambda_{x^t}^{\mathcal{X}_{K_2}^t} \right) \omega_{k_1} \right\} \Delta(x_i) \right\} \\ &= \sum_{x \in \mathcal{X}_{K_2}} \operatorname{Tr}\big\{ \hat{\rho}_x^{R_i} \Delta(x) \big\} \mu_{X_{K_2}}(x), \end{aligned} \tag{C.3.4}$$

where $\omega_{k_1} := \sqrt{\rho_{k_1}^{\otimes t}}$ and we defined

$$\hat{\rho}_x^{R_i} := \frac{1}{\mu_{X_{K_2}}(x)} \operatorname{Tr}_{[t]\setminus i} \left\{ \omega_{k_1} \left(\sum_{x^{[t]\setminus i} \in \mathcal{X}_{K_2}^{t-1}} \Lambda_{x^t}^{\mathcal{X}_{K_2}^t} \right) \omega_{k_1} \right\},$$

as the operator representing the unnormalized post-measured reference state of the i-th system after discrete measurement POVM. Also,

$$
\begin{aligned}
&d(R_{i,k_1}, \hat{X}_{i,k_1,K_2}) \\
&= \int_{\mathcal{X}^t} \mathrm{Tr}\Big\{ \mathrm{Tr}_{[t]\setminus i} \Big\{ \omega_{k_1} \hat{\Lambda}(dz^t) \omega_{k_1} \Big\} \Delta(z_i) \Big\} \\
&= \int_{\mathcal{X}^t} \mathrm{Tr}\left\{ \mathrm{Tr}_{[t]\setminus i} \left\{ \omega_{k_1} \left(\sum_{x^t \in \mathcal{X}^t_{K_2}} \Lambda_{x^t}^{\mathcal{X}^t_{K_2}} \pi^t_{OT}\Big(dz^t \Big| x^t\Big) \right) \omega_{k_1} \right\} \Delta(z_i) \right\} \\
&= \sum_{x^t \in \mathcal{X}^t_{K_2}} \int_{\mathcal{RQ}_{K_2}(x^t)} \mathrm{Tr}\Big\{ \mathrm{Tr}_{[t]\setminus i} \Big\{ \omega_{k_1} \Lambda_{x^t}^{\mathcal{X}^t_{K_2}} \pi^t_{OT}\Big(dz^t \Big| x^t\Big) \omega_{k_1} \Big\} \Delta(z_i) \Big\} \\
&= \sum_{x_i \in \mathcal{X}_{K_2}} \int_{\mathcal{RQ}_{K_2}(x_i)} \mathrm{Tr}\Bigg\{ \mathrm{Tr}_{[t]\setminus i} \Bigg\{ \omega_{k_1} \Bigg(\sum_{x^{[t]\setminus i}} \Lambda_{x^t}^{\mathcal{X}^t_{K_2}} \Bigg(\int_{\mathcal{RQ}_{K_2}(x^{[t]\setminus i})} \pi^{t-1}_{OT}\Big(dz^{[t]\setminus i} | x^{[t]\setminus i}\Big) \Bigg) \Bigg) \omega_{k_1} \Bigg\} \\
&\qquad\qquad \cdot \Delta(z_i) \pi_{OT}(dz_i | x_i) \Bigg\} \\
&= \sum_{x_i \in \mathcal{X}_{K_2}} \int_{\mathcal{RQ}_{K_2}(x_i)} \mathrm{Tr}\Bigg\{ \mathrm{Tr}_{[t]\setminus i} \Bigg\{ \omega_{k_1} \Bigg(\sum_{x^{[t]\setminus i}} \Lambda_{x^t}^{\mathcal{X}^t_{K_2}} \Bigg) \omega_{k_1} \Bigg\} \Delta(z_i) \pi_{OT}(dz_i | x_i) \Bigg\} \\
&= \sum_{x \in \mathcal{X}_{K_2}} \mathrm{Tr}\Bigg\{ \hat{\rho}_x^{R_i} \Bigg(\int_{\mathcal{RQ}_{K_2}(x)} \Delta(z) \pi_{OT}(dz | x) \Bigg) \Bigg\} \mu_{X_{K_2}}(x),
\end{aligned}
\tag{C.3.5}
$$

Substituting expressions (C.3.4) and (C.3.5), we find the distortion of the optimal transport block as

$$\begin{aligned}
&d_t(R^t_{k_1}, \hat{X}^t_{k_1,K_2}) - d_t(R^t_{k_1}, {X'}^t_{k_1,K_2}) \\
&= \frac{1}{t}\sum_{i=1}^{t}\sum_{x\in\mathcal{X}_{K_2}} \mathrm{Tr}\Bigg\{\hat{\rho}^{R_i}_x\Bigg(\int_{\mathcal{R}\mathcal{Q}_{K_2}(x)} \Delta(z)\pi_{OT}(dz|x) - \Delta(x)\Bigg)\Bigg\}\mu_{X_{K_2}}(x) \\
&= \sum_{x\in\mathcal{X}_{K_2}} \mathrm{Tr}\Bigg\{\bar{\rho}^R_{t,x}\Bigg(\int_{\mathcal{R}\mathcal{Q}_{K_2}(x)} \Delta(z)\pi_{OT}(dz|x) - \Delta(x)\Bigg)\Bigg\}\mu_{X_{K_2}}(x) \\
&= \sum_{x\in\mathcal{X}_{K_2}} \mathrm{Tr}\Bigg\{\bar{\rho}^R_{t,x}\Bigg(\int_{\mathcal{R}\mathcal{Q}_{K_2}(x)} \Big(\Delta(z) - \Delta(x)\Big)\pi_{OT}(dz|x)\Bigg)\Bigg\}\mu_{X_{K_2}}(x) \\
&= \sum_{x\in\mathcal{X}_{K_2}} \int_{\mathcal{R}\mathcal{Q}_{K_2}(x)} \mathrm{Tr}\Big\{\bar{\rho}^R_{t,x}\Big(\Delta(z) - \Delta(x)\Big)\Big\}\pi_{OT}(dz|x)\mu_{X_{K_2}}(x) \\
&= \int_{\mathcal{X}} \mathrm{Tr}\Big\{\bar{\rho}^R_{t,Q_{K_2}(x)}\Big(\Delta(x) - \Delta(Q_{K_2}(x))\Big)\Big\}\mu_X(dx),
\end{aligned} \tag{C.3.6}$$

where $\bar{\rho}^R_{t,x}$ is the t-letter average PMR state defined as

$$\bar{\rho}^R_{t,x} := \frac{1}{t}\sum_{i=1}^{t}\hat{\rho}^{R_i}_x.$$

Substituting (C.3.6) and (C.3.2) into (C.3.1) provides the following bound

$$\begin{aligned}
d_n(R^n, \hat{X}^n_{k_1,K_2}|\neg E_{ce}) &= (P_{k_1} + \epsilon_{cl})d\Big(R, \hat{X}_{\text{local}}|A_{k_1} = 1\Big) + d(R_{k_1} + X'_{k_1,K_2}) \\
&+ \int_{\mathcal{X}} \mathrm{Tr}\Big\{\lim_{t_{\min}\to\infty}\mathbb{E}_{T|\neg E_{ce}}\Big[\bar{\rho}^R_{Q_{K_2}(x)}\Big]\Big(\Delta(x) - \Delta(Q_{K_2}(x))\Big)\Big\}\mu_X(dx) \\
&= (P_{k_1} + \epsilon_{cl})d\Big(R, \hat{X}_{\text{local}}|A_{k_1} = 1\Big) + d(R_{k_1} + X'_{k_1,K_2}) \\
&+ \int_{\mathcal{X}} \mathrm{Tr}\Big\{\bar{\bar{\rho}}^R_x\Big(\Delta(x) - \Delta(Q_{K_2}(x))\Big)\Big\}\mu_X(dx),
\end{aligned}$$

where the asymptotic post-measured average reference state is given by

$$\bar{\bar{\rho}}_x^R := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \rho_x^{R_i}.$$

Bibliography

[1] T. A. Atif and S. S. Pradhan. Distributed quantum faithful simulation and function computation using algebraic structured measurements. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 124–129. IEEE, 2021.

[2] T. A. Atif, M. Heidari, and S. S. Pradhan. Faithful simulation of distributed quantum measurements with applications in distributed rate-distortion theory. *IEEE Transactions on Information Theory*, 68(2):1085–1118, 2021.

[3] Y. Bai, X. Wu, and A. Özgür. Information constrained optimal transport: From talagrand, to marton, to cover. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2210–2215, 2020. doi: 10.1109/ISIT44484.2020.9174478.

[4] H. Barnum. Quantum rate-distortion coding. *Physical Review A*, 62(4):042309, Sept. 2000. ISSN 1050-2947, 1094-1622. doi: 10.1103/PhysRevA.62.042309. URL `https://link.aps.org/doi/10.1103/PhysRevA.62.042309`.

[5] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter. The quantum reverse shannon theorem and resource tradeoffs for simulating quantum

channels. *IEEE Transactions on Information Theory*, 60(5):2926–2959, 2014. doi: 10.1109/TIT.2014.2309968.

[6] Y. Blau and T. Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6228–6237, 2018.

[7] Y. Blau and T. Michaeli. Rethinking lossy compression: The rate-distortion-perception tradeoff. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 675–685. PMLR, 09–15 Jun 2019. URL `https://proceedings.mlr.press/v97/blau19a.html`.

[8] J. Chen. Private communication. *Manuscript in Preparation*.

[9] J. Chen, L. Yu, J. Wang, W. Shi, Y. Ge, and W. Tong. On the rate-distortion-perception function. *IEEE Journal on Selected Areas in Information Theory*, pages 1–1, 2022. doi: 10.1109/JSAIT.2022.3231820.

[10] S. Cole, M. Eckstein, S. Friedland, and K. Życzkowski. Quantum optimal transport. *arXiv preprint arXiv:2105.06922*, 2021.

[11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1 edition, Sept. 2005. ISBN 9780471241959 9780471748823. doi: 10.1002/047174882X. URL `https://onlinelibrary.wiley.com/doi/book/10.1002/047174882X`.

[12] P. Cuff. Distributed channel synthesis. *IEEE Transactions on Information Theory*, 59(11):7071–7096, 2013.

[13] P. W. Cuff. *Communication in networks for coordinating behavior*. Stanford University, 2009.

[14] P. W. Cuff, H. H. Permuter, and T. M. Cover. Coordination capacity. *IEEE Transactions on Information Theory*, 56(9):4181–4206, 2010.

[15] N. Datta and C. Rouzé. Relating relative entropy, optimal transport and fisher information: a quantum hwi inequality. In *Annales Henri Poincaré*, volume 21, pages 2115–2150. Springer, 2020.

[16] N. Datta, M.-H. Hsieh, and M. M. Wilde. Quantum rate distortion, reverse shannon theorems, and source-channel separation. *IEEE Transactions on Information Theory*, 59(1):615–630, 2012.

[17] N. Datta, M.-H. Hsieh, M. M. Wilde, and A. Winter. Quantum-to-classical rate distortion coding. *Journal of Mathematical Physics*, 54(4):042201, 2013.

[18] G. De Palma and D. Trevisan. Quantum optimal transport with quantum channels. In *Annales Henri Poincaré*, volume 22, pages 3199–3234. Springer, 2021.

[19] G. De Palma and D. Trevisan. Quantum optimal transport with quantum channels. *Annales Henri Poincaré*, Mar 2021. ISSN 1424-0661. doi: 10.1007/s00023-021-01042-3. URL http://dx.doi.org/10.1007/s00023-021-01042-3.

[20] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd. The quantum wasserstein distance of order 1. *IEEE Transactions on Information Theory*, 67(10):6627–6643, 2021. doi: 10.1109/TIT.2021.3076442.

[21] A. El Gamal and Y.-H. Kim. *Network information theory*. Cambridge university press, 2011.

[22] H. Fawzi, J. Saunderson, and P. A. Parrilo. Semidefinite approximations of the matrix logarithm. *Foundations of Computational Mathematics*, 2018. Package cvxquad at `https://github.com/hfawzi/cvxquad`.

[23] S. Ghorai, P. Grangier, E. Diamanti, and A. Leverrier. Asymptotic security of continuous-variable quantum key distribution with a discrete modulation. *Physical Review X*, 9(2):021059, 2019.

[24] F. Golse, C. Mouhot, and T. Paul. On the mean field and classical limits of quantum mechanics. *Communications in Mathematical Physics*, 343:165–205, 2016.

[25] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[26] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, Mar. 2014.

[27] H. J. Groenewold. A problem of information gain by quantal measurements. *International Journal of Theoretical Physics*, 4(5):327–338, 1971.

[28] B. Gupt, J. Izaac, and N. Quesada. The walrus: a library for the calculation of hafnians, hermite polynomials and gaussian boson sampling. *Journal of*

Open Source Software, 4(44):1705, 2019. doi: 10.21105/joss.01705. URL `https://doi.org/10.21105/joss.01705`.

[29] A. S. Holevo. *Quantum Systems, Channels, Information*. De Gruyter, Berlin, Boston, 2019. ISBN 9783110642490. doi: doi:10.1515/9783110642490. URL `https://doi.org/10.1515/9783110642490`.

[30] A. S. Holevo. Gaussian maximizers for quantum gaussian observables and ensembles. *IEEE Transactions on Information Theory*, 66(9):5634–5641, 2020.

[31] A. S. Holevo and R. F. Werner. Evaluating capacities of bosonic gaussian channels. *Physical Review A*, 63(3):032312, 2001.

[32] A. S. Holevo, M. Sohma, and O. Hirota. Capacity of quantum gaussian channels. *Physical Review A*, 59(3):1820, 1999.

[33] K. Ikeda. Foundation of quantum optimal transport and applications. *Quantum Information Processing*, 19(1):25, 2020.

[34] L. V. Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

[35] N. Killoran, J. Izaac, N. Quesada, V. Bergholm, M. Amy, and C. Weedbrook. Strawberry fields: A software platform for photonic quantum computing. *Quantum*, 3:129, mar 2019. doi: 10.22331/q-2019-03-11-129. URL `https://doi.org/10.22331%2Fq-2019-03-11-129`.

[36] J. Lin, T. Upadhyaya, and N. Lütkenhaus. Asymptotic security analysis of discrete-modulated continuous-variable quantum key distribution. *Physical Review X*, 9(4):041064, 2019.

[37] H. Liu, G. Zhang, J. Chen, and A. Khisti. Cross-domain lossy compression as entropy constrained optimal transport. *IEEE Journal on Selected Areas in Information Theory*, 2022.

[38] M. M Wilde, P. Hayden, F. Buscemi, and M.-H. Hsieh. The information-theoretic costs of simulating quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 45(45):453001, Oct 2012. ISSN 1751-8121. doi: 10.1088/1751-8113/45/45/453001. URL `http://dx.doi.org/10.1088/1751-8113/45/45/453001`.

[39] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

[40] T. Ogawa and H. Nagaoka. Making good codes for classical-quantum channel coding via quantum hypothesis testing. *IEEE Transactions on Information Theory*, 53(6):2261–2266, 2007. doi: 10.1109/TIT.2007.896874.

[41] M. Ozawa. Conditional probability and a posteriori states in quantum mechanics. *Publications of the Research Institute for Mathematical Sciences*, 21(2): 279–295, 1985.

[42] M. S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day, 1964.

[43] E. Posner. Random coding strategies for minimum entropy. *IEEE Transactions on Information Theory*, 21(4):388–391, 1975.

[44] R. Renner and J. I. Cirac. de finetti representation theorem for infinite-dimensional quantum systems and applications to quantum cryptography. *Physical review letters*, 102(11):110504, 2009.

[45] N. Saldi, T. Linder, and S. Yüksel. Output constrained lossy source coding with limited common randomness. *IEEE Transactions on Information Theory*, 61(9):4984–4998, 2015. doi: 10.1109/TIT.2015.2450721.

[46] A. Serafini. *Quantum continuous variables: a primer of theoretical methods*. CRC Press, Taylor & Francis Group, CRC Press is an imprint of the Taylor & Francis Group, an informa business, Boca Raton, 2017. ISBN 9781482246346.

[47] M. E. Shirokov. On lower semicontinuity of the quantum conditional mutual information and its corollaries. *Proceedings of the Steklov Institute of Mathematics*, 313(1):203–227, 2021.

[48] C. Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[49] A. B. Wagner. The rate-distortion-perception tradeoff: The role of common randomness, 2022. URL `https://arxiv.org/abs/2202.04147`.

[50] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84(2):621–669, May 2012. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.84.621. URL `https://link.aps.org/doi/10.1103/RevModPhys.84.621`.

[51] M. M. Wilde. *Quantum information theory*. Cambridge University Press, 2013.

[52] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999. doi: 10.1109/18.796385.

[53] A. Winter. “extrinsic”and “intrinsic”data in quantum measurements: Asymptotic convex decomposition of positive operator valued measures. *Communications in mathematical physics*, 244(1):157–185, 2004.

[54] J. Yard, P. Hayden, and I. Devetak. Quantum broadcast channels. *IEEE Transactions on Information Theory*, 57(10):7147–7162, 2011. doi: 10.1109/TIT.2011.2165811.

[55] G. Zhang, J. Qian, J. Chen, and A. Khisti. Universal rate-distortion-perception representations for lossy compression. *Advances in Neural Information Processing Systems*, 34:11517–11529, 2021.