UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Optimizing and Generalizing Quantum Fidelities

by Afham

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of A/Prof Christopher Ferrie and A/Prof Simon Devitt

University of Technology Sydney
Faculty of Engineering and Information Technology

June 2025

Certificate of Original Authorship

I, Afham, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: June 2, 2025

Acknowledgments

This thesis is the labor of a *distribution* over the love-hate spectrum, weighted significantly toward the *love* end; for that, I have countless people to thank.

I have thoroughly enjoyed my PhD journey, and my deepest gratitude goes to my advisor, Chris Ferrie. His unwavering support and belief in me have been invaluable and have allowed me to grow as a researcher. When I began my PhD, I lacked the rigor and knowledge to support my mathematical ambitions fully. Chris provided me with guidance, time, resources, and introductions to exceptional colleagues and mentors, all of which have been crucial to my development.

I am deeply indebted to Richard Kueng, who has not only co-authored the first article of my PhD but also remained an invaluable mentor throughout this period.

Certain forms of appreciation cannot be fully expressed in words, yet one must try. I thank Afrad Basheer, my friend, whom I've known since we were both ten years old. Not only have we known each other for 18 years now, but we have also had the fortune of pursuing PhDs in the same field, in the same department, at the same university.

I extend my gratitude to Marco Tomamichel for hosting me in his group for multiple research visits. This thesis would be far less rich without the discussions I've had with him and his students. In particular, I thank Roberto Rubboli for countless insightful conversations and for pointing me to important resources.

I am also grateful to my co-supervisor, Simon Devitt, and my committee members Sanjiang Li and Yuval Sanders, for their support and guidance. I have had the good fortune of meeting many excellent researchers, all of whom have left a lasting impact on me. I especially thank Nilanjana Datta, Mark Wilde, Francesco Buscemi, and Yuan Feng for their discussions and advice.

During this period, I have made many friends who have enriched my PhD experience. I thank Mauro Morales, Ian George (who, most importantly, introduced me to climbing), Upendra Kapshikar, and Rishabh Batra.

Academia has a strange propensity to consume one's life. I thank my parents, Ashraf and Rahana, and my sisters, Ishal and Nihika, for their unwavering support and love. Finally, I thank my *non-academic* friends—Ashish, Asif, Mannathu, Lipika, Joseph, Neha, and Dany—for keeping me sane.

I thank Sydney Quantum Academy for the financial support they have provided.

This is a thesis by compilation. Parts of this thesis are under peer-review for publication or being prepared for the same as indicated in subsequent chapters. The content have been edited for the purpose of a coherent presentation.

Declaration of Publications

I, **Afham**, declare that the following papers included in this thesis are my own work and that I have made significant contributions to each of them. Below, I provide details on the authorship, publication status, and contribution breakdown for each paper.

Title: *Quantum mean states are nicer than you think: fast algorithms to compute states maximizing average fidelity.*

Authors: Afham, Richard Kueng, and Chris Ferrie.

arXiv ID: 2206.08183

Publication outlet: Unpublished

Status: In preparation for publication

Contribution: CF and RK conceived the project. Afham did the majority of theoretical research and all of the numerical simulations. CF and RK conceived the applications. All authors wrote and approved the manuscript.

Title: *Riemannian-geometric generalizations of quantum fidelities and Bures-Wasserstein distance.*

Authors: Afham and Chris Ferrie.

arXiv ID: 2410.04937

Publication outlet: Journal of Mathematical Physics

Status: Under review

Contribution: Afham and CF conceived the project. Afham did the majority of the theoretical research and all of the numerical simulations. All authors wrote and approved the manuscript.

Signatures

Afham: Production Note: Signature removed prior to publication. Date: 6 February 2025

Chris Ferrie: Production Note: Signature removed prior to publication. Date: 6 Feb 2025

Richard Kueng: Production Note: Signature removed prior to publication. Date: February 6th, 2025

Contents

List of Figures

List of Tables

Abstract

Quantum fidelities are cornerstone metrics in quantum sciences, widely used to quantify the similarity of quantum states. Despite its prominence and age, several of its properties remain unexplored. This thesis advances our understanding of fidelities and its generalizations by addressing open problems and introducing novel frameworks rooted in quantum information theory and Riemannian geometry.

This thesis focuses mainly on three of problems. In the first part, we resolve the problem of maximizing average fidelity over finite ensembles of quantum states. By constructing a semidefinite program to compute the maximum average fidelity and deriving scalable fixed-point algorithms, we demonstrate significant improvements in computational runtime. We also derive novel bounds and expressions for near-optimal states, which are exact in special cases, such as when the ensemble consists of commuting states. These results provide new tools for applications, including Bayesian quantum tomography, where they address outstanding challenges.

In the second part, we extend the concept of fidelity by introducing a family of generalized fidelities based on the Riemannian geometry of the Bures–Wasserstein manifold. This framework unifies and generalizes several existing quantum fidelities, including Uhlmann-, Holevo-, and Matsumoto-fidelity, and preserves their celebrated properties. Through a rigorous mathematical treatment, we establish invariance and covariance properties, derive an Uhlmann-like theorem, and discuss some possible generalizations of quantum Rényi divergences.

In the third part, we study the problem of projecting positive matrices to certain convex and compact sets with respect to Bures distance (or equivalently Uhlmann fidelity). These convex and compact sets are defined by quantum channels, and for certain channels, most importantly partial trace, we derive a closed-form for the projection.

Using the closed form for partial-trace projection we demonstrate various applications for our results including quantum process tomography and random state generation. Moreover, these results also endow the *pretty good measurement* and Petz recovery map with novel geometric and operational interpretations.

The existence of closed-form for specific channels is related to the saturation of the *data processing inequality* (DPI) for fidelity. Thus our results also provide explicit examples for the saturation of DPI for fidelity.

Together, these contributions provide a comprehensive study of quantum fidelities bridging foundational theory and practical applications.

Chapter 1

Introduction

1.1 Overview

Quantum computing is a fundamentally different paradigm of computing that leverages the quantum properties of the physical world. By harnessing the exponential difficulty in the classical simulation of quantum mechanical objects, we aim to construct algorithms whose performance surpasses that of algorithms run on *classical* computers.

This view was succinctly, albeit slightly crudely, described by Richard Feynman when he said *"Nature is not classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it does not look so easy"* [Fey82]. Thus the field of quantum computing was born, with the ultimate aim of creating *quantum computers* that can harness the quantum mechanical powers of nature to solve problems that are intractable for classical computers. The poster child for such a *super-classical* algorithm that runs on a quantum computer is Shor's factoring algorithm [Sho94].

Unlike classical computers which use classical resources like bits and classical gates, quantum computers use quantum resources *qubits* (quantum bits) and quantum gates. This crossover to the quantum side comes with its mathematical challenges. States of quantum bits are described by complex vectors and matrices, quantum gates, which manipulate the states of these quantum bits are described by complex matrices, and the outcomes of quantum computations are probabilistically determined via quantum *measurements*, which are modeled using special kinds of matrices. Thus the study of quantum information necessitates delving deeper into our collective mathematical toolbox which has led to the formation of a rich tapestry of mathematical results.

More formally, the states of quantum systems are described by *density ma-*

trices, which can be seen as the non-commutative (matrix) generalizations of *probability vectors* which describe states of classical systems in *classical* information theory. A fundamental problem in information theory involves comparing two states. Since (finite-dimensional) classical states can be described by (probability) vectors, this problem can be translated to the comparison of vectors.

Such comparisons are formalized by functions that take in two probability vectors and return a (typically nonnegative) real number. This class of functions will be collectively called *comparators* or *figures of merit*. Standard examples of comparators include distance functions such as Euclidean distance, Hellinger distance [Hel09], and total variation distance [LPW09], as well as divergences such as Kullback-Leibler divergence [KL51], Rényi divergences [Rén61], and f-divergences [Mor63; AS66; Csi67], to name a few. Each of the above comparators comes with its information-theoretic and/or geometric interpretation, and therefore the choice of the comparator depends on the context of the problem.

As one might expect, defining analogous figures of merit for quantum states—which are described by density matrices—requires the construction of more intricate functions, which are *quantizations* of these *classical* comparators. Quantization refers to the requirement that if the two (input) density matrices commute, then the output of the quantum comparator must equal the output of the corresponding classical comparator with inputs being the vector of eigenvalues of the commuting states. More formally, if $f_{\mathrm{C}} : \mathbb{R}_+^d \times \mathbb{R}_+^d \to \mathbb{R}_+$ is a classical comparator, a necessary condition for $f_{\mathrm{Q}} : \mathbb{H}_d \times \mathbb{H}_d \to \mathbb{R}_+$ to be a *quantization* of f_{C} is

$$[P, Q] = 0 \implies f_{\mathrm{Q}}(P, Q) = f_{\mathrm{C}}(\lambda(P), \lambda(Q)), \tag{1.1}$$

where $\lambda(P)$ denotes the vector of eigenvalues of P.

An implication of the non-commutative nature of quantum information is that quantizations of classical comparators are often not unique—there exist multiple, often infinite, quantizations of the same classical comparators. Perhaps the most famous such instance involves Rényi divergences, which has multiple families of quantum generalizations [WWY14; Mül+13; AD13; Tom15; Dup+14], each with interesting mathematical properties and operational interpretations [LMW13].

A popular comparator for classical distributions is the *classical fidelity* or *Bhattacharyya coefficient* [Bha43]. For positive vectors $p, q \in \mathbb{R}_+^d$, it is defined as

$$\mathrm{F}_{\mathrm{Cl}}(p, q) := \left\langle p^{\frac{1}{2}}, q^{\frac{1}{2}} \right\rangle = \sum_{i=1}^{d} \sqrt{p_i q_i}, \tag{1.2}$$

where square roots of nonnegative vectors are defined elementwise: $\left[p^{\frac{1}{2}}\right]_i :=$

$[p]_i^{\frac{1}{2}}$. Classical fidelity is not a distance measure, but a *similarity measure*. For probability vectors, classical fidelity satisfies the following bounds:

$$p, q \in \Delta_d \quad \Rightarrow \quad 0 \leq \mathrm{F_{Cl}}(p, q) \leq 1, \tag{1.3}$$

with the lower bound being saturated if the two vectors are orthogonal ($p \perp q$) and the upper bounds being saturated if the two vectors are equal. For arbitrary positive vectors, the upper bound is to be replaced with $\sqrt{\sum_{i=1}^d p_i \sum_{i=1}^d q_i}$.

Though classical fidelity is a similarity measure, it features prominently in the definition of a bona fide distance called the *Hellinger distance*:

$$\mathrm{h}(p, q) := \left\| p^{\frac{1}{2}} - q^{\frac{1}{2}} \right\|_2 = \sqrt{\sum_{i=1}^d (p_i + q_i) - 2\sqrt{p_i q_i}}. \tag{1.4}$$

Often we will be interested not in distances, but in *squared distances* as they show nicer mathematical properties. Some examples include the fact that Pythagorean theorem involves squared Euclidean distance and the Euclidean mean of a distribution minimizes the average *squared distance* over the distribution and not the average distance. Another manifestation of this statement is that divergences are typically analogous to squared distances, with squared Euclidean distance being a (Bregman) divergence.

Thus we will be typically interested in the squared Hellinger distance: $\mathrm{H}(p, q) \equiv \mathrm{h}(p, q)^2 = \left\| p^{\frac{1}{2}} - q^{\frac{1}{2}} \right\|_2^2$, for any pair $p, q \in \mathbb{R}_+^d$. Observe that the squared Hellinger distance can be thought of as the difference between the (sum of) arithmetic and geometric means of two vectors. More formally, respectively define the arithmetic mean and geometric mean of two positive vectors $p, q \in \mathbb{R}_+^d$ as

$$\mathrm{A}(p, q) := \frac{p + q}{2} \quad \text{and} \quad \mathrm{G}(p, q) := p^{\frac{1}{2}} \odot q^{\frac{1}{2}}, \tag{1.5}$$

where $u \odot v$ denotes the Hadamard (entrywise) product (also known as Schur product) of two vectors: $[u \odot v]_i = [u]_i \cdot [v]_i$. This implies

$$\frac{1}{2}\,\mathrm{H}(p, q) := \sum_{i=1}^d [\mathrm{A}(p, q)]_i - [\mathrm{G}(p, q)]_i. \tag{1.6}$$

While generalizing from positive vectors to positive semidefinite matrices, the (sum of) arithmetic mean part generalizes in an unambiguous manner: $\mathrm{A}(P, Q) := \frac{1}{2}\,\mathrm{Tr}[P + Q]$. However, there is no unique way to generalize the geometric mean part. Or equivalently, there is no unique quantization of classical fidelity. This leads to a panoply of *quantum fidelities*, which we now discuss.

Perhaps the most straightforward way of quantizing classical fidelity leads to *Holevo fidelity* [Kho72; Wil18]:

$$\mathrm{F}^{\mathrm{H}}(P,Q) := \left\langle P^{\frac{1}{2}}, Q^{\frac{1}{2}} \right\rangle = \mathrm{Tr}\left[P^{\frac{1}{2}}Q^{\frac{1}{2}}\right]. \tag{1.7}$$

Indeed we simply replace the Euclidean inner product of the positive vectors $p^{\frac{1}{2}}, q^{\frac{1}{2}}$ with the Hilbert–Schmidt inner product of the positive semidefinite matrices $P^{\frac{1}{2}}, Q^{\frac{1}{2}}$. The Holevo fidelity has been studied under many names like *quantum affinity* [LZ04] and *pretty good fidelity* [IRS17]. An operational interpretation of Holevo fidelity is that it is the overlap between the *canonical purifications* of P and Q.

Another quantization is obtained via taking *geometric mean* part literally, which leads to the *Matsumoto fidelity* [Mat10]:

$$\mathrm{F}^{\mathrm{M}}(P,Q) := \mathrm{Tr}[P\#Q] := \mathrm{Tr}\left[P^{\frac{1}{2}}\sqrt{P^{-\frac{1}{2}}QP^{-\frac{1}{2}}}P^{\frac{1}{2}}\right], \tag{1.8}$$

where $P\#Q := P^{\frac{1}{2}}\sqrt{P^{-\frac{1}{2}}QP^{-\frac{1}{2}}}P^{\frac{1}{2}}$ denotes the *matrix geometric mean* of two positive definite matrices [KA80]. See Section 2.1.5 for further properties of the matrix geometric mean. We refer to [Mat10; Mat14; CS20] for operational and geometric interpretations of the Matsumoto fidelity.

However the most popular quantum fidelity is the *Uhlmann fidelity* [Uhl76; Joz94]:[1]

$$\mathrm{F}^{\mathrm{U}}(P,Q) := \left\|P^{\frac{1}{2}}Q^{\frac{1}{2}}\right\|_1 = \mathrm{Tr}\left[\sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}}\right]. \tag{1.9}$$

The ubiquity and usefulness of Uhlmann fidelity is exemplified by the fact that it is often referred to as simply *fidelity*. Operationally it is the largest absolute overlap attainable over all purifications of P and Q. It also has geometric relevance as it features in the definition of the Bures distance, which is the natural distance of the *Bures manifold* of positive definite matrices, which is a Riemannian manifold. We refer to Bengtsson and Życzkowski [BŻ17] for further geometric interpretations of Uhlmann fidelity.

Note that all the above three fidelities are valid quantizations of classical fidelity: if $[P,Q] = 0$, then they all reduce to the classical fidelity between the spectra of P and Q. One can construct a *Hellinger-like* quantity from each of these fidelities:

$$\mathrm{H}(P,Q) := \mathrm{Tr}[P+Q] - 2\,\mathrm{F}(P,Q), \tag{1.10}$$

where $\mathrm{F}(\cdot,\cdot)$ denotes either of the three fidelities. For Holevo fidelity and Uhlmann

[1] Uhlmann originally defined the square of this quantity as *transition probability*.

fidelity, the corresponding Hellinger quantities define squares of bona fide distances called matrix Hellinger distance and Bures distance respectively. However, for Matsumoto fidelity, the corresponding Hellinger quantity is not a squared distance, but a divergence as it does not satisfy triangle inequality [BGJ19].

This thesis focuses on three problems related to quantum fidelities. We first look at the problem of finding the state that maximizes average Uhlmann fidelity over probability distribution supported on a finite set of quantum states. Formally, let $\mathcal{R} := \{\rho_1, \ldots, \rho_n\}$ be an ensemble of quantum states and let $w \in \Delta_n$ be probability vectors encoding associated weights. We are interested in the problem

$$\sigma_\sharp := \operatorname*{argmax}_{\sigma \in \mathbb{D}_d} \sum_{i=1}^{n} w_i \, \mathrm{F}^{\mathrm{U}}(\rho_i, \sigma). \tag{1.11}$$

The state $\sigma_\sharp$ is called *optimal average state* (with respect to Uhlmann fidelity) or *(Uhlmann) fidelity barycenter*. We phrase this problem as a semidefinite program (SDP) and show that it exhibits complementary slackness if all the states in the ensemble are full-rank.

From complementary slackness relations, we derive a *fixed-point equation* satisfied uniquely by the optimal state. This leads to two fixed-point algorithms that exhibit superior numerical performance—in terms of runtime—compared to solving the SDP using standard numerical solvers (see Figure 3.2). We then discuss how one of the fixed-point algorithms can be interpreted as a *Projected Riemannian Gradient Descent* on the *Bures–Wasserstein manifold*. The Bures–Wasserstein manifold (often shortened to Bures manifold) is the Riemannian manifold of positive definite matrices whose natural distance is the Bures distance. We also derive tight and easily computable upper- and lower-bounds for the average fidelity achieved by the optimal average state and discuss applications in quantum tomography and connection to other results in quantum information.

In the second problem, we zoom out a bit to consider all the three previously discussed fidelities. In particular, we define and study *generalized fidelity* (and an associated *generalized Bures distance*). These objects naturally arise from the Riemannian geometry of the Bures–Wasserstein manifold. Generalized fidelity is parametrized by a positive definite matrix we call the *base* of the generalized fidelity. More formally, we define the generalized fidelity between P and Q *at* a base R as

$$\mathrm{F}_R(P, Q) := \operatorname{Tr}\left[\sqrt{R^{\frac{1}{2}} P R^{\frac{1}{2}}} R^{-1} \sqrt{R^{\frac{1}{2}} Q R^{\frac{1}{2}}}\right]. \tag{1.12}$$

For specific values of the base, the generalized fidelity recovers all three of the

previously mentioned fidelities, which we collectively call the *named fidelities*. In particular, choosing $R \in P, \mathbb{I}, P^{-1}$ recovers the Uhlmann, Holevo, and Matsumoto fidelities, respectively. Remarkably, generalized fidelity displays various geometric properties such as invariance and covariance as the base moves along certain geodesic-related paths on the manifold of positive definite matrices. We also study other properties and characterizations of generalized fidelity including a block-matrix characterization (which is related to the SDP for average fidelity defined in the previous problem) and an *Uhlmann-like* theorem for generalized fidelity which characterizes generalized fidelity based on purifications of the states involved. We then conclude the work with open problems and potential applications.

The third and final work explores the *projection* with respect to the Bures distance (or equivalently Uhlmann fidelity). Projections are a crucial component in optimization problems [Bub15]. Formally, a projection of a point x to a set $\mathcal{C}$ is the closest point $y \in \mathcal{C}$ to x with respect to some distance or divergence:

$$\Pi_{\mathcal{C}}[x] := \operatorname*{argmin}_{y \in \mathcal{C}} \mathrm{D}(y, x), \tag{1.13}$$

where D represents a squared distance or a divergence. Typically projections are done with respect to the Euclidean distance [Bub15] or Bregman divergences like relative entropy (a.k.a. Kullback–Leibler divergence) [NY83; HSF24; YCL22; YFT19].

However, Euclidean distance is often not the ideal distance to consider in quantum information as the projection could map density matrices (or positive semidefinite matrices) out of the positive semidefinite cone. Moreover, it also does not satisfy the data processing inequality (DPI) [Oza00]. Thus it is important to find projections that are more suitable to the geometry of quantum states, and Bures distance is often an ideal metric for this purpose [BŻ17].

To elaborate, we consider the projection of an arbitrary positive definite matrix to the convex and compact set formed by the intersection of the positive semidefinite cone and an affine subspace of the vector space of Hermitian matrices—a spectrahedron [RG95; Vin14; Chi23]. The sets we consider are formed by positive semidefinite matrices which are mapped to a given (constraint) matrix under a given quantum channel (linear completely positive and trace-preserving (CPT) map).

Formally, let $P \in \mathbb{P}_{\mathcal{H}}$, $\Lambda \in \mathrm{CPT}(\mathcal{H}, \mathcal{K})$, and $C \in \mathbb{P}_{\mathcal{K}}$. We are interested in the

optimization problem

$$\Pi_{\Lambda,C}[P] := \underset{Q:\Lambda(Q)=C}{\operatorname{argmin}} \mathrm{B}(P,Q), \tag{1.14}$$

where $\mathrm{B}(P,Q) := \mathrm{Tr}[P+Q] - 2\,\mathrm{F}^{\mathrm{U}}(P,Q)$ is the squared Bures distance.

We present a closed form function we call the *Gamma map* which can be constructed for any *constraint pair* (Λ, C). For certain channels, the Gamma map serve as the closed-form for Bures projection. These channels include partial trace (including trace), pinching channel (including the completely dephasing map), and projective measurements with mutually orthogonal projectors. Whether the Gamma map *works* for a certain channel is closely related to the saturation of the DPI for fidelity [LRD17; CS22b]. Thus our results also provide non-trivial examples of saturation of DPI for fidelity.

For example, given an arbitrary bipartite positive definite matrix P, we find the projection of P to the set of (bipartite positive definite) matrices with a fixed marginal on either space. In particular, if this marginal is chosen to be the identity matrix, this gives a closed-form solution for the *Bures projection* to the set of Choi matrices of quantum channels, which finds applications in quantum process tomography.

Apart from such practical results, our closed-form projection also gives a novel geometric and operational interpretation to the *pretty good measurement* [Bel75; HW94] as the *closest* measurement to a given ensemble of positive matrices with respect to total squared Bures distance (or equivalently Uhlmann fidelity). The closed form also endows a geometric interpretation to the Petz recovery map [Pet86b; Pet88]. Considering the wide applicability of this projection closed-form, we expect it to be useful in further optimization problems in quantum sciences.

We conclude this section by noting that all three problems are intimately connected to the geometry of the Bures–Wasserstein manifold. Common threads connecting each of the above three chapters are seen throughout and are discussed in Chapter 6, thereby providing a unified ending to the results discussed in this thesis. Finally, we conclude in Chapter 7, where we discuss the results, open questions, and further research directions.

1.2 Articles related to this thesis

1. Afham, Kueng, and Ferrie [AKF22]. Quantum mean states are nicer than you think: fast algorithms to compute states maximizing average fidelity; arXiv:2206.08183. Discussed in Chapter 3.

2. Afham and Ferrie [AF24]. Riemannian-geometric generalizations of quantum fidelities and Bures–Wasserstein distance; arXiv:2410.04937 (submitted to Journal of Mathematical Physics). Discussed in Chapter 4.

1.3 Thesis organization

The contents of this thesis is organized as follows.

In **Chapter 2** we introduce and summarize the mathematical preliminaries required for rest of the thesis. We first review necessary concepts from linear algebra, convex optimization, quantum information, and finally Riemannian geometry.

In **Chapter 3** we look at the first set of results of this thesis. Here we deal with the problem of finding the states that maximizes average Uhlmann fidelity over a given distribution of density matrices. The results include semidefinite programs which find the solutions and fixed-point algorithms that converge to the optimal state. These fixed-point algorithms are seen to perform numerically better than the semidefinite programs (solved via standard numerical solvers). We discuss how one of the fixed-point iterative algorithms can be seen as a *projected Riemannian gradient descent* on the Bures manifold. We also provide tight upper bounds and lower bounds that are easier to compute and conclude with a discussion of applications including in Bayesian quantum tomography.

Next, in **Chapter 4** we introduce and study generalized fidelity and generalized Bures distance. These generalizations are based on the Riemannian geometry of the Bures–Wasserstein manifold. The generalization depends on a (third) positive definite matrix called the *base* and after reviewing basic properties of generalized fidelity and generalized Bures distance, we show remarkable geometric properties of these quantities. These include invariance and covariance properties as the base traverses along certain geodesic-related paths over the manifold of positive definite matrices. We then study various characterizations of generalized fidelity such as block-matrix characterization and purification-based characterization. We also introduce an analogous generalization to certain families of quantum Rényi divergences and finally conclude the chapter with a discussion of potential applications in (quantum and classical) machine learning and open problems.

In **Chapter 5** we study the problem of *projecting* an arbitrary positive semidefinite matrix to certain convex subsets of the positive semidefinite cone. We present a function we call the Γ map and provide sufficient conditions for it to serve as the closed-form for projection. We discuss the channels for which the closed-form holds and provide explicit formulae for these channels. We then dis-

cuss the applications which include quantum process tomography, random quantum state and ensemble generation, and projection for optimization problems. Using these results, we also give a geometric/operational interpretation to the *pretty good measurement* and the Petz recovery map.

In **Chapter 6** we show how the results discussed in the previous three chapters share interesting mathematical and geometric relations. Finally we conclude with **Chapter 7** with a discussion of the results in the thesis, open problems, and future directions.

Notations

Acronyms

BW	Bures–Wasserstein
CP	Completely positive
CPT	Completely positive and trace-preserving
CPU	Completely positive and unital
DPI	Data processing inequality
FP	Fixed-point
HS	Hilbert–Schmidt
FoM(s)	Figure(s) of merit
PGM	Pretty good measurement
POVM	Positive operator-valued measure
PSD	Positive semidefinite
QSD	Quantum state discrimination
RGD	Riemannian gradient descent
SDP	Semidefinite program

Sets and Vector Spaces

$\mathbb{N}, \mathbb{R}, \mathbb{C}$	Natural, real, and complex numbers.
$\mathbb{R}^d, \mathbb{C}^d$	d-dimensional real and complex vectors.
$\mathcal{A}, \mathcal{B}, \mathcal{C}$	Denote sets.
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	Denote complex vector spaces ($\mathbb{C}^d$ for some $d \in \mathbb{N}$).
$[n]$	The index set $\{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$.
$\mathbb{M}_d, \mathbb{M}_{\mathcal{X}}$	$d \times d$ (square) matrices (acting on the space $\mathcal{X}$).
$\mathbb{H}_d, \mathbb{H}_{\mathcal{X}}$	$d \times d$ Hermitian matrices (acting on the space $\mathcal{X}$).
$\mathbb{P}_d, \mathbb{P}_{\mathcal{X}}$	$d \times d$ PSD matrices (acting on the space $\mathcal{X}$).
$\mathbb{P}_d^+, \mathbb{P}_{\mathcal{X}}^+$	$d \times d$ positive definite matrices (acting on the space $\mathcal{X}$).
$\mathbb{U}_d, \mathbb{U}_{\mathcal{X}}$	$d \times d$ unitary matrices (acting on the space $\mathcal{X}$).
$\mathbb{D}_d, \mathbb{D}_{\mathcal{X}}$	$d \times d$ density matrices (acting on the space $\mathcal{X}$).
$\mathbb{M}_{\mathcal{X},\mathcal{Y}}$	Matrices corresponding to linear maps from $\mathcal{X}$ to $\mathcal{Y}$.
$\mathrm{LM}(\mathcal{X}, \mathcal{Y})$	linear maps from $\mathbb{M}_{\mathcal{X}}$ to $\mathbb{M}_{\mathcal{Y}}$.
$\mathrm{CP}(\mathcal{X}, \mathcal{Y})$	completely positive maps from $\mathbb{M}_{\mathcal{X}}$ to $\mathbb{M}_{\mathcal{Y}}$.
$\mathrm{CPT}(\mathcal{X}, \mathcal{Y})$	quantum channels (CPT maps) from $\mathbb{M}_{\mathcal{X}}$ to $\mathbb{M}_{\mathcal{Y}}$.
$\mathrm{CPU}(\mathcal{X}, \mathcal{Y})$	CP and unital (CPU maps) from $\mathbb{M}_{\mathcal{X}}$ to $\mathbb{M}_{\mathcal{Y}}$.
$J_{\mathrm{CP}}(\mathcal{X}, \mathcal{Y})$	Choi matrices of completely positive maps $\mathrm{CP}(\mathcal{X}, \mathcal{Y})$.
$J_{\mathrm{CPT}}(\mathcal{X}, \mathcal{Y})$	Choi matrices of $\mathrm{CPT}(\mathcal{X}, \mathcal{Y})$.
$J_{\mathrm{CPU}}(\mathcal{X}, \mathcal{Y})$	Choi matrices of $\mathrm{CPU}(\mathcal{X}, \mathcal{Y})$.
Δ_d	Probability vectors of length d.
$f^{-1}[b]$	$\{a : f(a) = b\}$. The preimage of a function $f : \mathcal{A} \to \mathcal{B}$ for any $b \in \mathcal{B}$.

If the symbol for any matrix class is used without a subscript indicating dimension or vector space, it denotes the set of *all* matrices of that type (the union over all dimensions d). For example, $\mathbb{M} \equiv \bigcup_{d \in \mathbb{N}} \mathbb{M}_d$.

Inner products, norms and distances

$\langle u, v\rangle, \langle u\|v\rangle$	$\sum_{i=1}^{d} \bar{u}_i v_i$. Standard inner product on $\mathbb{C}^d$
$\langle A, B\rangle$	$\mathrm{Tr}[A^* B]$. Hilbert–Schmidt inner product between $A, B \in \mathbb{M}_d$.
$\|u\|, \|u\|_2$	Euclidean norm of a vector u.
$\|A\|_1$	Trace norm of a matrix $A \in \mathbb{M}$.
$\|A\|_2$	$\sqrt{\langle A, A\rangle}$. Frobenius norm of a matrix $A \in \mathbb{M}$.
$\|A\|_\infty, \|A\|$	Spectral norm of a matrix $A \in \mathbb{M}$.
$\mathrm{F}(P, Q), \mathrm{F}^{\mathrm{U}}(P, Q)$	Uhlmann fidelity between $P, Q \in \mathbb{P}_d$.
$\mathrm{F}^{\mathrm{H}}(P, Q)$	Holevo fidelity between P and $Q \in \mathbb{P}_d$.
$\mathrm{F}^{\mathrm{M}}(P, Q)$	Matsumoto fidelity between P and $Q \in \mathbb{P}_d$.
$\mathrm{F}_R(P, Q)$	Generalized fidelity between P and $Q \in \mathbb{P}_d$ at $R \in \mathbb{P}_d^+$.
$\mathrm{B}(P, Q)$	Squared Bures distance between P and Q.
$\mathrm{B}_R(P, Q)$	Sq. generalized Bures distance between P and Q at R.
$\mathfrak{g}$	Symbol to denote a metric tensor.
$\mathfrak{g}_p(u, v), \langle u, v\rangle_p^{\mathfrak{g}}$	Inner product between tangent vectors $u, v \in \mathrm{T}_p\mathcal{M}$ at p on a Riemannian manifold $(\mathcal{M}, \mathfrak{g})$.
$\langle \cdot, \cdot\rangle_P^{\mathrm{Bu}}$	The Bures–Wasserstein metric tensor at $P \in \mathbb{P}_d^+$.
Π	Projection (typically with respect to Bures distance).
$\Pi_{\Lambda, C}$	Bures projection onto the set $\Lambda^{-1}[C]$.
$\hat{\Pi}_{\mathcal{X}, C}$	Bures projection onto the set of PSD matrices with $\mathcal{X}$-marginal being C.

Miscellaneous

$A^\intercal, \bar{A}$	The transpose and conjugate of a matrix $A \in \mathbb{M}$.
A^*	$\bar{A}^\intercal$. The adjoint (conjugate transpose) of $A \in \mathbb{M}$.
$\intercal$	Transpose ($\intercal(A) \equiv A^\intercal$).
$P^{\frac{1}{2}}, \sqrt{P}$	Equivalent notations for the PSD square root of $P \in \mathbb{P}$.
$\lvert A\rvert$	$\sqrt{A^*A}$. Modulus of a matrix $A \in \mathbb{M}$.
$\mathrm{Pol}\,(A)$	$A\lvert A\rvert^{-1}$. Unitary factor of the polar decomposition of invertible $A \in \mathbb{M}$.
P, Q, R	Variables to denote positive semidefinite matrices.
ρ, σ	Variables to denote density matrices.
Φ, Ψ	Variables to denote linear maps between matrices.
Φ^*	The adjoint of a linear map Φ.
$\tilde{\Phi}_\rho$	The Petz recovery map of the channel-state pair (Φ, ρ).
$\Phi^\intercal$	$\intercal \circ \Phi \circ \intercal$. The *transpose twirl* of a linear map Φ.
$\lvert\psi\rangle, \lvert\phi\rangle$	Variables to denotes statevectors.
$\mathbb{I}_d, \mathbb{I}_\mathcal{X}$	The $d \times d$ Identity matrix (acting on the space $\mathcal{X}$)
$\lvert\omega\rangle$	$\sum_{i=1}^d \lvert i\rangle\lvert i\rangle$. Unnormalized canonical Bell state.
Ω	$\lvert\omega\rangle\langle\omega\rvert$. Density matrix of the unnormalized canonical Bell state.
$\mathcal{M}$	Symbol to denote a manifold.
$(\mathcal{M}, \mathfrak{g})$	Riemannian manifold $\mathcal{M}$ and metric tensor $\mathfrak{g}$.
$\mathrm{T}_x\mathcal{M}$	The tangent space at $x \in \mathcal{M}$.
$P \star Q$	$Q^{\frac{1}{2}}PQ^{\frac{1}{2}}$. Star product for $P, Q \in \mathbb{P}_d$.
$\frac{P}{Q}$	$Q^{-\frac{1}{2}}PQ^{-\frac{1}{2}}$. Symmetrized division for $P \in \mathbb{P}_d$ and $Q \in \mathbb{P}_d^+$.
$[A, B]$	$AB - BA$. The commutator of $A, B \in \mathbb{M}_d$.
$\mathrm{Tr}, \mathrm{Tr}_\mathcal{X}$	The trace and partial trace (over $\mathcal{X}$) maps.
$\mathrm{Tr}_{\mathcal{Z}\setminus\mathcal{X}}$	The partial trace over every subsystem of $\mathcal{Z}$ except $\mathcal{X}$.
$[Z]_\mathcal{X}, Z_\mathcal{X}$	The $\mathcal{X}$-marginal of a square matrix Z.
$\mathrm{Id}_\mathcal{X}$	The identity superoperator acting on $\mathbb{M}_\mathcal{X}$.
$\mathcal{L}_P(V)$	Solution to matrix Lyapunov equation satisfying $P\mathcal{L}_P(V) + \mathcal{L}_P(V)P = V$.
$\mathcal{L}_P^{-1}(Q)$	$QP + PQ$. Inverse to the matrix Lyapunov operator.

Chapter 2

Mathematical Preliminaries

In this chapter, we provide a brief discussion of the mathematical prerequisites needed for this thesis. For a detailed understanding of the mathematics discussed in most of this section, see Watrous [Wat18, Mathematical Preliminaries].

2.1 Linear algebra

The mathematical description of quantum systems involves vectors, matrices, and other tools from linear algebra and matrix analysis. In this section, we give a summary of relevant concepts from linear algebra. Throughout this thesis, we will reserve $d \in \mathbb{N}$ to denote the dimension (length) of a vector or the length of each axis of a square matrix.

2.1.1 Some classes of matrices

We now look at some classes of matrices that we will use frequently.

- **Complex square matrices.** The set of all $d \times d$ complex square matrices is denoted by $\mathbb{M}_d$. The set of all complex square matrices is denoted by $\mathbb{M}$.
- **Hermitian matrices.** $H \in \mathbb{M}$ is Hermitian if $H = H^*$. The set of $d \times d$ Hermitian matrices is denoted by $\mathbb{H}_d$.
- **Unitary matrices.** $U \in \mathbb{M}$ is *unitary* if its adjoint (conjugate-transpose) equals its inverse: $U^* = U^{-1}$. The set of $d \times d$ unitary matrices is denoted by $\mathbb{U}_d$.
- **Positive semidefinite matrices.** A matrix P is *positive semidefinite* (a.k.a. positive) if it is Hermitian and has every eigenvalue to be nonnegative. The set of $d \times d$ positive semidefinite matrices is denoted by $\mathbb{P}_d$. An alternate notation for positive semidefinite matrices we will use is $P \geq 0$.

- **Positive definite matrices.** A matrix P is *positive definite* if it is Hermitian and has every eigenvalue to be strictly positive. The set of $d \times d$ positive definite matrices is denoted by $\mathbb{P}_d^+$. An alternate notation for positive definite matrices we will use is $P > 0$.

- **Density matrices.** A matrix ρ is a *density matrix* if it is a positive semidefinite matrix of unit trace: $\rho \geq 0$ and $\mathrm{Tr}[\rho] = 1$. The set of $d \times d$ density matrices is denoted by $\mathbb{D}_d$.

If two complex spaces $\mathcal{X}$ and $\mathcal{Y}$ are of the same dimension, that is $\mathcal{X} = \mathbb{C}^d$ and $\mathcal{Y} = \mathbb{C}^d$, then they are isomorphic and we will denote this equivalence as $\mathcal{X} \cong \mathcal{Y}$. For a vector space $\mathcal{X} = \mathbb{C}^d$, we denote $\mathbb{M}_\mathcal{X}$ to denote the set of all square matrices acting on $\mathcal{X}$. Indeed, by the fact that $\mathcal{X} = \mathbb{C}^d$, the sets $\mathbb{M}_\mathcal{X}$ and $\mathbb{M}_d$ are isomorphic and thus will be treated the same. Essentially we refer to it as $\mathbb{M}_\mathcal{X}$ ($\mathbb{M}_d$) if labeling the space as $\mathcal{X} \cong \mathbb{C}^d$ is (not) important.

The **trace** of a matrix is defined as the sum of its diagonal entries. It is denoted as Tr and it is a linear map from square matrices to $\mathbb{C}$.

2.1.2 Inner products, norms, and distances

We now discuss the related concepts of inner products, norms, and distances defined on vector spaces. For the rest of this subsection, let $\mathcal{V}$ be vector space over the base field $\mathbb{C}$. This can be the vector space of *vectors* or matrices, as the analysis requires no such distinction.

Inner products

An (complex) inner product over $\mathcal{V}$ is a complex-valued binary operation $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{C}$ satisfying

1. Linearity in second argument: $\langle u, \alpha v_1 + v_2 \rangle = \alpha \langle u, v_1 \rangle + \langle u, v_2 \rangle$.

2. Conjugate symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle}$.

3. Positive definiteness: $\langle u, u \rangle \geq 0$ with equality if and only if $u = 0$.

Note that (1) and (2) implies conjugate linearity in the first argument: $\langle \alpha u_1 + u_2, v \rangle = \overline{\alpha} \langle u_1, v \rangle + \langle u_2, v \rangle$. Various inner products exist on the space of vectors and matrices. The dot product over vectors is perhaps the most common inner product, and over matrices, we have the Hilbert–Schmidt (a.k.a. Frobenius) inner product defined as

$$\langle A, B \rangle_{\mathrm{HS}} \equiv \langle A, B \rangle := \mathrm{Tr}[A^* B], \tag{2.1}$$

for matrices A and B of compatible shape. One can define a *real* inner product similarly.

Norms

A *norm* $\|\cdot\|$ on a vector space $\mathcal{V}$ is a function of the form $\|\cdot\| : \mathcal{V} \to \mathbb{R}_+$ satisfying the following properties [Wat18]. Here $A, B \in \mathcal{V}$ are chosen arbitrarily.

1. Positive definiteness: $\|A\| \geq 0$, with $\|A\| = 0$ if and only if $A = 0$.
2. Homogeneity: $\|\alpha A\| = |\alpha|\|A\|$ for all $\alpha \in \mathbb{C}$.
3. Triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$.

Every inner product on $\mathcal{V}$ induces a norm of the form $\|A\| = \sqrt{\langle A, A\rangle}$ for any $A \in \mathcal{V}$. However, the converse is not true—not every norm on $\mathcal{V}$ is induced by an inner product. Throughout the thesis, we will utilize various matrix norms, with an important family of norms being *Schatten-p norms*. The Schatten p-norm of a matrix A is defined as

$$\|A\|_p := \mathrm{Tr}\left[|A|^p\right]^{\frac{1}{p}}, \tag{2.2}$$

where $|A| = \sqrt{A^*A}$. The Schatten norm recovers three important norms of specific values of p.

1. **Trace norm** ($p = 1$): $\|A\|_1 := \mathrm{Tr}\left[|A|\right]$ is the sum of singular values of A.
2. **Hilbert–Schmidt norm** ($p = 2$): $\|A\|_2 := \sqrt{\mathrm{Tr}\left[|A|^2\right]} = \sqrt{\langle A, A\rangle}$ is the norm induced by the HS inner product.
3. **Spectral norm** ($p \to \infty$): $\|A\|_\infty = \max_{u \in \mathcal{X}:\|u\|_2=1} \|Au\|_2$ coincides with the largest singular value of A.

In the last line, we used $\|u\|$ to denote the Euclidean norm of vectors. We now discuss a variational expression for trace norm which turns out to be useful later on [Wat18; NC01].

Proposition 2.1.1 (Variational characterization of trace norm)**.** *For any $A \in \mathbb{M}_d$, it holds that*

$$\|A\|_1 = \max_{U:\|U\|_\infty \leq 1} |\langle U, A\rangle| = \max_{U:\|U\|_\infty \leq 1} \mathrm{Re}\langle U, A\rangle. \tag{2.3}$$

If A is full-rank, the optimal U is unique and equal to the unitary factor appearing in the polar decomposition of A. See Section 2.1.4 for proof and further details regarding polar decomposition.

Distances

A *distance function* or a *metric* on a set $\mathcal{X}$ is a function of the form $\mathrm{d} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ such that for all $x, y, z \in \mathcal{X}$ we have

1. **Non-negativity** : $\mathrm{d}(x, y) \geq 0$ with equality if and only if $x = y$.
2. **Symmetry** : $\mathrm{d}(x, y) = \mathrm{d}(y, x)$.
3. **Triangle inequality** : $\mathrm{d}(x, z) \leq \mathrm{d}(x, y) + \mathrm{d}(y, z)$.

Unlike for inner products and norms, we do not require the underlying set to be a vector space for distances.

Every norm $\|\cdot\|$ induces a distance by the relation $\mathrm{d}(x, y) := \|x - y\|$. *Trace distance* and *Hilbert–Schmidt distance*, the distances induced by trace norm and Hilbert–Schmidt norm respectively, are important in quantum information.

However, not every distance metric needs to be induced by a norm. One such distance metric, which is of great importance for this thesis, is the *Bures distance* which is defined on the set of positive semidefinite matrices as

$$\mathrm{b}(P, Q) := \sqrt{\mathrm{Tr}[P + Q] - 2\left\|P^{\frac{1}{2}}Q^{\frac{1}{2}}\right\|_1}. \tag{2.4}$$

Though it is not induced by *an* inner product, it is a distance metric on a Riemannian manifold, thus it is induced by a metric tensor—a family of (real) inner products. These concepts are further elaborated in Section 2.4.

2.1.3 Linear maps over matrices

Let $\mathcal{X} = \mathbb{C}^d$ and $\mathcal{Y} = \mathbb{C}^{d'}$ be complex Euclidean spaces. A linear map Φ from $\mathbb{M}_\mathcal{X}$ to $\mathbb{M}_\mathcal{Y}$ is a map such that

$$\Phi(\alpha X_1 + \beta X_2) = \alpha\Phi(X_1) + \beta\Phi(X_2) \in \mathbb{M}_\mathcal{Y}, \tag{2.5}$$

for all $X_1, X_2 \in \mathbb{M}_\mathcal{X}$ and $\alpha, \beta \in \mathbb{C}$. The set of all such linear maps from $\mathbb{M}_\mathcal{X}$ to $\mathbb{M}_\mathcal{Y}$ is denoted by $\mathrm{LM}(\mathcal{X}, \mathcal{Y})$. We will use *linear map* in an unqualified fashion to refer to linear maps between square matrices. Indeed matrices are linear maps between vectors, but for such linear maps we already have a well-defined name—matrices.

To each $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$, we may uniquely associate a linear map $\Phi^* \in \mathrm{LM}(\mathcal{Y}, \mathcal{X})$ such

$$\langle Y, \Phi(X)\rangle = \langle \Phi^*(Y), X\rangle, \tag{2.6}$$

for all $X \in \mathbb{M}_\mathcal{X}$ and $Y \in \mathbb{M}_\mathcal{Y}$, where $\langle\cdot,\cdot\rangle$ denotes the HS inner product. Such a map Φ^* is called the *adjoint map* of Φ. Linear maps are of particular importance

in quantum information. Two important linear maps are the trace map and its generalization, the *partial trace* map, which we discuss next. A more detailed discussion of specific kinds of linear maps is deferred to Section 2.3.2.

Trace and partial trace

Let $\mathcal{X}$ be a complex Euclidean space. The *trace map* is a linear map of the form $\mathrm{Tr} : \mathbb{M}_{\mathcal{X}} \to \mathbb{C}$, whose action on a square matrix is obtained by summing along its diagonal. The trace functional is invariant with respect to the choice of basis. The adjoint map of trace is given by multiplying the identity matrix: $\mathrm{Tr}^*(\alpha) = \alpha \mathbb{I}_{\mathcal{X}}$, for any $\alpha \in \mathbb{C}$. Let $\mathcal{Y}$ be another complex Euclidean space. One can then consider the linear map $\mathrm{Tr} \otimes \mathrm{Id}_{\mathcal{Y}} : \mathbb{M}_{\mathcal{X}\otimes\mathcal{Y}} \to \mathbb{M}_{\mathcal{Y}}$ whose action is defined as

$$(\mathrm{Tr} \otimes \mathrm{Id}_{\mathcal{Y}})(X \otimes Y) = \mathrm{Tr}[X] \cdot Y, \tag{2.7}$$

for any pair $X \in \mathbb{M}_{\mathcal{X}}, Y \in \mathbb{M}_{\mathcal{Y}}$. This map is called the *partial trace* (over $\mathcal{X}$) as it traces out the component associated with $\mathcal{X}$. Hence, assuming the input space has been labeled properly, we will denote this map as $\mathrm{Tr}_{\mathcal{X}}$, with the subscript indicating the space that is traced out. The adjoint of the partial trace is given by tensoring the identity matrix (of the traced-out space): $\mathrm{Tr}^*_{\mathcal{X}}(Y) = \mathbb{I}_{\mathcal{X}} \otimes Y$ for any $Y \in \mathbb{M}_{\mathcal{Y}}$. It follows that $\mathrm{Tr}[X \otimes Y] = \mathrm{Tr}[X]\,\mathrm{Tr}[Y]$.

Indeed this can be generalized to multipartite spaces as well. Let $\mathcal{Z} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ with $\mathcal{X}_i = \mathbb{C}^{d_i}$ for $i \in [n]$. Then $\mathrm{Tr}_{\mathcal{X}_k} : \mathbb{M}_{\mathcal{Z}} \to \mathbb{M}_{\mathcal{Z}\setminus\mathcal{X}_k}$ is the map that traces out the $\mathcal{X}_k$-component, where $\mathcal{Z}\setminus\mathcal{X}_k \equiv \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_{k-1} \otimes \mathcal{X}_{k+1} \otimes \cdots \otimes \mathcal{X}_n$. The adjoint map is given by an appropriate tensoring of $\mathbb{I}_{\mathcal{X}_k}$. Analogously, the partial trace map that traces out every space other than $\mathcal{X}_k$ is denoted as $\mathrm{Tr}_{\mathcal{Z}\setminus\mathcal{X}_k} : \mathbb{M}_{\mathcal{Z}} \to \mathbb{M}_{\mathcal{X}_k}$. Its action on $Z := X_1 \otimes \cdots \otimes X_n \in \mathbb{M}_{\mathcal{Z}}$ is defined as

$$\mathrm{Tr}_{\mathcal{Z}\setminus\mathcal{X}_k}[Z] = \mathrm{Tr}_{\mathcal{Z}\setminus\mathcal{X}_k}[X_1 \otimes \cdots \otimes X_k \otimes \cdots \otimes X_n] = \prod_{i=1, i\neq k}^{n} \mathrm{Tr}[X_i] X_k. \tag{2.8}$$

Such an operation will also be called *marginalization* of Z over all space except $\mathcal{X}_k$ or computing the $\mathcal{X}_k$-*marginal* of Z. Here too the adjoint is given by an appropriate tensoring of the identity matrices associated with the traced-out spaces. We will also use the alternate notation $\mathrm{Tr}_{\mathcal{Z}\setminus\mathcal{X}_k}[Z] \equiv [Z]_{\mathcal{X}_k}$ or $Z_{\mathcal{X}_k}$ when unambiguous. For a bipartite matrix $Z \in \mathbb{M}_{\mathcal{Z}}$ with $\mathcal{Z} = \mathcal{X} \otimes \mathcal{Y}$, this would read

$$\mathrm{Tr}_{\mathcal{X}}[Z] \equiv [Z]_{\mathcal{Y}} \quad \text{and} \quad \mathrm{Tr}_{\mathcal{Y}}[Z] \equiv [Z]_{\mathcal{X}}. \tag{2.9}$$

Finally, we remark that by linearity we can extend all the above concepts to matrices that are not in a tensor product form. We once again refer to [Wat18,

Mathematical prelimiaries] for further details.

2.1.4 Polar decomposition

Polar decomposition allows us to write an arbitrary matrix as the product of a unitary and a positive semidefinite matrix. Let $A \in \mathbb{M}_d$ be an arbitrary square matrix. Then the right and left polar decompositions of A are given by

$$A = U|A| \quad \text{and} \quad |A^*|U, \tag{2.10}$$

respectively. Here $U \in \mathbb{U}_d$ is a unitary matrix and $|A| = \sqrt{A^*A} \in \mathbb{P}_d$ is positive semidefinite. Note that the unitary U in the left and right decompositions, which is referred to as the *unitary factor*, can be chosen to be the same. If A is full rank, then U is unique, and $|A|$ (and hence $|A^*|$) is positive definite. The unique unitary factor U will be denoted as $U = \mathrm{Pol}\,(A)$. Thus, for full rank matrices, we have

$$U \equiv \mathrm{Pol}\,(A) := A|A|^{-1} \iff U^* = U^{-1} = |A|A^{-1}. \tag{2.11}$$

As discussed in Section 2.1.2, the trace-norm of a matrix is defined as $\|A\|_1 = \mathrm{Tr}[|A|]$. Let A be full-rank and let $U = \mathrm{Pol}\,(A)$. Observe that

$$\langle U, A\rangle = \mathrm{Tr}[U^*A] = \mathrm{Tr}[U^{-1}A] = \mathrm{Tr}[|A|A^{-1}A] = \mathrm{Tr}[|A|] = \|A\|_1, \tag{2.12}$$

which shows that the unitary that achieves the optimal value in the variational expression for trace-norm is indeed the polar factor of A.

The polar decomposition plays an important role in the definition of trace norm, and thereby in the definitions of trace distance and (Uhlmann) fidelity. It also plays crucial roles in the results discussed in Chapters 3 and 4.

2.1.5 Matrix geometric mean

We now discuss some basic properties of the geometric mean of two positive definite matrices [KA80]. For an excellent treatment of this topic, see [Bha09]. See [Liu+24] for quantum algorithms for matrix geometric means, along with a discussion on the role of matrix geometric means in various quantum information problems. [CS20] also discusses matrix geometric means in the context of quantum information.

Definition 2.1.2 (Geometric mean)**.** *Let $A, B \in \mathbb{P}_d^+$ be positive definite matrices. The geometric mean between A and B is defined as*

$$A\#B := A^{\frac{1}{2}}\sqrt{A^{-\frac{1}{2}}BA^{-\frac{1}{2}}}A^{\frac{1}{2}}. \tag{2.13}$$

The geometric mean is symmetric: $A\#B = B\#A$, and it reduces to $A\#B = A^{\frac{1}{2}}B^{\frac{1}{2}} = \sqrt{AB}$ if A and B commute. If A, B are singular positive semidefinite matrices, then their geometric mean can be defined by a limiting procedure:

$$A\#B := \lim_{\epsilon \to 0}(A + \epsilon \mathbb{I}_d)\#(B + \epsilon \mathbb{I}_d). \tag{2.14}$$

However, in this thesis, we typically deal with the geometric means of positive definite matrices. The geometric mean also enjoys various other properties, some of which we list here.

Proposition 2.1.3 (Properties of Geometric mean)**.** *Let $A, B \in \mathbb{P}_d^+$ and $A\#B$ be their geometric mean. The following statements hold true.*

1. $(A\#B)^{-1} = A^{-1}\#B^{-1}$.
2. $(ZAZ^*)\#(ZBZ^*) = Z(A\#B)Z^*$ *for any invertible* Z.
3. $X = A\#B$ *is the unique positive definite solution to the matrix Riccati equation*

$$B = XA^{-1}X. \tag{2.15}$$

The proofs are available in [Bha09].

2.1.6 Star product

We now discuss the *star product*, which is a non-commutative and non-associative product over positive semidefinite matrices [LS13] defined as follows.

Definition 2.1.4 (Star product)**.** *Let $P, Q \in \mathbb{P}_d$. The star-product between P and Q is defined as*

$$P \star Q := Q^{\frac{1}{2}}PQ^{\frac{1}{2}}. \tag{2.16}$$

The star product turns out to be convenient while working with positive semidefinite matrices, and specifically the Bures manifold as it *compactifies* various expressions of interest. We collect some elementary properties of the star product in the following proposition.

Proposition 2.1.5. *The following statements regarding the star product hold true for positive semidefinite matrices.*

1. $P, Q \geq 0 \Rightarrow P \star Q \geq 0$.
2. $(P \star Q) \star Q^{-1} = P$ *if* Q *is invertible.*
3. $(P \star Q)^\intercal = P^\intercal \star Q^\intercal$.

Proof. The first two statements follow directly from the definition. For the last one, observe that

$$(P \star Q)^\intercal = \left(Q^{\frac{1}{2}} P Q^{\frac{1}{2}}\right)^\intercal = \sqrt{Q}^\intercal P^\intercal \sqrt{Q}^\intercal = \sqrt{Q^\intercal} P^\intercal \sqrt{Q^\intercal} = P^\intercal \star Q^\intercal. \tag{2.17}$$

Here we have used the fact that $\sqrt{Q^\intercal} = \sqrt{Q}^\intercal$. To see this, observe that $Q^\intercal = (\sqrt{Q}\sqrt{Q})^\intercal = \sqrt{Q}^\intercal \sqrt{Q}^\intercal$. Since positive semidefinite matrices have unique positive semidefinite square roots, we have $\sqrt{Q}^\intercal = \sqrt{Q^\intercal}$. □

When multiple star-products are chained without parentheses, we evaluate from left to right:

$$A \star B \star C \equiv (A \star B) \star C := C^{\frac{1}{2}} \left(B^{\frac{1}{2}} A B^{\frac{1}{2}}\right) C^{\frac{1}{2}}. \tag{2.18}$$

Observe that for $\{P_i\}_{i \in [n]} \subset \mathbb{P}_d$, we have $P_1 \star P_2 \star \cdots \star P_n = Z^* Z$ where $Z = P_1^{\frac{1}{2}} P_2^{\frac{1}{2}} \cdots P_n^{\frac{1}{2}}$. Frequently we take the star product of $A, B \in \mathbb{P}_{\mathcal{X} \otimes \mathcal{Y}}$ where the second argument is a tensor product of the form $B = X \otimes \mathbb{I}_\mathcal{Y}$ or $B = \mathbb{I}_\mathcal{X} \otimes Y$. In this case, we use the shorthands

$$A \star_\mathcal{X} X := A \star (X \otimes \mathbb{I}_\mathcal{Y}) \quad \text{and} \quad A \star_\mathcal{Y} Y := A \star (\mathbb{I}_\mathcal{X} \otimes Y). \tag{2.19}$$

Indeed this can be generalized to multipartite systems. Let $\mathcal{Z} := \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, $A \in \mathbb{P}_\mathcal{Z}$, and $X \in \mathbb{P}_{\mathcal{X}_k}$ for some $k \in [n]$. Then we denote

$$A \star_{\mathcal{X}_k} X = A \star (\mathbb{I}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_{k-1}} \otimes X \otimes \mathbb{I}_{\mathcal{X}_{k+1} \otimes \cdots \otimes \mathcal{X}_n}). \tag{2.20}$$

Essentially we *star* X to the part of A in the component space $\mathcal{X}_k$ alone. Recall the shorthand we introduced for marginals: for $A \in \mathbb{P}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}$, we use $[A]_{\mathcal{X}_k}$ or $A_{\mathcal{X}_k}$ to denote the marginal of A on the space $\mathcal{X}_k$, which is obtained by tracing out every other subsystem. Indeed it follows that

$$[A \star_{\mathcal{X}_k} X]_{\mathcal{X}_k} = [A]_{\mathcal{X}_k} \star X, \tag{2.21}$$

for any $k \in [n]$ and $X \in \mathfrak{X}_k$. For bipartite systems, this can be made explicit as

$$\begin{aligned} [A \star_{\mathfrak{x}} X]_{\mathfrak{X}} &:= \mathrm{Tr}_{\mathfrak{y}}[A \star_{\mathfrak{x}} X] = \mathrm{Tr}_{\mathfrak{y}}[A] \star X = A_{\mathfrak{X}} \star_{\mathfrak{x}} X \\ [A \star_{\mathfrak{y}} Y]_{\mathfrak{y}} &:= \mathrm{Tr}_{\mathfrak{X}}[A \star_{\mathfrak{y}} Y] = \mathrm{Tr}_{\mathfrak{X}}[A] \star Y = A_{\mathfrak{y}} \star_{\mathfrak{y}} Y. \end{aligned} \tag{2.22}$$

The star product allows us a convenient shorthand for matrix geometric mean. Recall that for $A, B, C \in \mathbb{P}_d^+$, we have $B = CAC \iff C = A^{-1}\#B$, which can be rewritten as

$$B = A \star C^2 \iff C = A^{-1}\#B. \tag{2.23}$$

Equivalently, for any pair $A, B \in \mathbb{P}_d^+$, it holds that

$$A \star (A^{-1}\#B)^2 = B \quad \text{and} \quad B \star (B^{-1}\#A)^2 = A. \tag{2.24}$$

We now summarize relevant concepts from convex optimization.

2.2 Convex optimization and semidefinite programs

Convex optimization plays a central role in modern optimization theory. R.T. Rockafellar, one of the leading scholars in optimization theory, said that *"the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity"* [Roc93].

In some sense, convex problems are *easy* to solve and analyze. Fast algorithms for solving convex problems are at the heart of advancements in optimization and machine learning. Convex problems feature prominently in this thesis, and thus we now review the basics of convex optimization.

2.2.1 Convexity

The authoritative sources for convex analysis and optimization are Rockafellar [Roc70], Boyd and Vandenberghe [BV04], and Bubeck [Bub15]. See Watrous [Wat18, Mathematical Preliminaries] for an introduction to convexity in the context of quantum information.

Let $\mathcal{X} := \mathbb{C}^d$. A set $\mathcal{C} \subseteq \mathcal{X}$ is *convex* if for all $t \in [0, 1]$ we have

$$(1-t)x + ty \in \mathcal{C} \quad \text{for all } x, y \in \mathcal{C}. \tag{2.25}$$

Geometrically, this is equivalent to requiring that the line segment between any two points in the set lies entirely within the set. A function $f : \mathcal{C} \to \mathbb{R}$ is called convex if

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \tag{2.26}$$

for all $x, y \in \mathcal{C}$ and $t \in [0, 1]$. If the inequality holds strictly whenever $x \neq y$ and $t \in (0, 1)$, then we say that the function is *strictly convex*. We say function $g : \mathcal{C} \to \mathbb{R}$ is (strictly) concave if and only if $-g$ is (strictly) convex. A (differentiable) function $f : \mathcal{C} \to \mathbb{R}$ is called *strongly convex* with parameter $m > 0$ if

$$f(y) - f(x) - \mathrm{D}f(x)(y - x) \geq \frac{m}{2}\|y - x\|_2^2, \tag{2.27}$$

for all $x, y \in \mathcal{C}$. Here $\mathrm{D}f(x)$ is the total derivative of f at x (see Section 2.4.5 for further details on total derivatives). Strictly (and therefore strongly) convex functions have a unique minimum. Optimization algorithms, such as gradient descent, have provable convergence guarantees for strong convex functions.

Having defined convex sets and functions, we can define a *convex optimization problem*. At its simplest, a convex (optimization) problem involves minimizing a

convex function f over a convex set $\mathcal{C}$:

$$\begin{aligned} &\text{minimize} : f(x), \\ &\text{subject to} : x \in \mathcal{C}. \end{aligned} \tag{2.28}$$

An important subclass of convex optimization problems are *semidefinite programs* which feature prominently in this thesis, and thus we discuss it next.

2.2.2 Semidefinite programs

Semidefinite programming is a field of convex optimization where we are concerned with optimizing a linear function (recall, all linear functions are convex) over the intersection of the cone of positive semidefinite matrices with an affine space. Such a set is called a *spectrahedron* [RG95; Vin14]. See Chiribella [Chi23] for a study of various spectrahedra of importance in quantum information.

Let us now formalize the notion of semidefinite programs. Let $\mathcal{X}, \mathcal{Y}$ be complex Euclidean spaces. A semidefinite program (SDP) can specified by a triple (Φ, A, B) [Wat18] where $\Phi : \mathbb{M}_{\mathcal{X}} \to \mathbb{M}_{\mathcal{Y}}$ is a Hermitian preserving linear map, $A \in \mathbb{H}_{\mathcal{X}}$ and $B \in \mathbb{H}_{\mathcal{Y}}$ are Hermitian matrices. We adopt the SDP formalism outlined by Watrous [Wat18].

Given a triple (Φ, A, B) of the above form, we may associate two optimization problems with it, which we call the *primal* and *dual problems*.

Primal problem	**Dual problem**
$\text{maximize} : \langle A, X \rangle$	$\text{minimize} : \langle B, Y \rangle$
$\text{subject to} : \Phi(X) = B,$	$\text{subject to} : \Phi^*(Y) \geq A.$
$X \in \mathbb{P}_{\mathcal{X}}.$	$Y \in \mathbb{H}_{\mathcal{Y}}.$

(2.29)

The sets of all operators that satisfy the respective constraints are called *primal feasible* set $\mathcal{A}$ and *dual feasible* set $\mathcal{B}$:

$$\begin{aligned} \mathcal{A} &= \{X \in \mathbb{P}_{\mathcal{X}} : \Phi(\mathrm{X}) = B\}, \\ \mathcal{B} &= \{Y \in \mathbb{H}_{\mathcal{Y}} : \Phi^*(\mathrm{Y}) \geq A\}. \end{aligned} \tag{2.30}$$

The *primal optimum* α and *dual optimum* β are then defined as

$$\alpha := \sup_{X \in \mathcal{A}} \langle A, X \rangle \quad \text{and} \quad \beta := \inf_{Y \in \mathcal{B}} \langle B, Y \rangle. \tag{2.31}$$

In the case that $\mathcal{A} = \varnothing$ or $\mathcal{B} = \varnothing$, we set $\alpha := -\infty$ or $\beta := \infty$ respectively.

With semidefinite programs, there exist certain notions of duality. The first notion has already manifested in the form of the *dual problem* above. Another

set of manifestations is *weak duality* and *strong duality*. The property of weak duality, which holds for all semidefinite programs, is that the primal optimum is always bounded above by the dual optimum: $\alpha \leq \beta$. Strong duality describes the situation where the inequality is saturated, which as one might expect, is not exhibited by all SDPs. *Slater's conditions* are two sufficient conditions for strong duality.

Theorem 2.2.1 (Slater's theorem for semidefinite programs [Wat18])**.** *Let $\Phi : \mathbb{M}_{\mathcal{X}} \to \mathbb{M}_{\mathcal{Y}}$ be a Hermitian-preserving map, and let $A \in \mathbb{H}_{\mathcal{X}}$ and $B \in \mathbb{H}_{\mathcal{Y}}$ be Hermitian matrices. Letting $\mathcal{A}, \mathcal{B}, \alpha$, and β be defined as above for the semidefinite program (Φ, A, B), the following two statements hold true:*

1. *If $\mathcal{A} \neq \varnothing$ and there exists a Hermitian operator $Y \in \mathbb{H}_{\mathcal{Y}}$ such that $\Phi^*(Y) > A$, then $\alpha = \beta$, and moreover there exists a primal-feasible operator $X \in \mathcal{A}$ such that $\langle A, X \rangle = \alpha$.*
2. *If $\mathcal{B} \neq \varnothing$ and there exists a positive definite operator $X \in \mathbb{P}_{\mathcal{X}}^{+}$ such that $\Phi(X) = B$, then $\alpha = \beta$, and moreover there exists a dual-feasible operator $Y \in \mathcal{B}$ such that $\langle B, Y \rangle = \beta$.*

The satisfaction of either of the above two statements implies strong duality, $\alpha = \beta$. If strong duality holds and there exist primal and dual feasible operators $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ achieving equality

$$\langle A, X \rangle = \alpha = \beta = \langle B, Y \rangle, \tag{2.32}$$

then a certain relation between these two operators, namely *complementary slackness*, exists.

Theorem 2.2.2 (Complementary slackness for semidefinite programs [Wat18])**.** *Let $\Phi : \mathbb{M}_{\mathcal{X}} \to \mathbb{M}_{\mathcal{Y}}$ be a Hermitian-preserving map, and let $A \in \mathbb{H}_{\mathcal{X}}$ and $B \in \mathbb{H}_{\mathcal{Y}}$ be Hermitian matrices. Let $\mathcal{A}$ and $\mathcal{B}$ be the primal-feasible and dual-feasible sets associated with the semidefinite program (Φ, A, B), and suppose that $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ are operators satisfying $\langle A, X \rangle = \langle B, Y \rangle$. It holds that*

$$\Phi^*(Y)X = AX. \tag{2.33}$$

Having summarized the prerequisites from linear algebra and convex optimization, we now discuss quantum information preliminaries.

2.3 Quantum information preliminaries

In this section, we discuss the prerequisites from quantum information. The classic textbook for quantum information is Nielsen and Chuang [NC10]. For a treatment whose tenor is followed by this thesis, see Watrous [Wat18]. Other excellent textbooks for this topic are Wilde [Wil13] and Tomamichel [Tom15]. For a detailed exposition of the geometric aspects of quantum information, see Bengtsson and Życzkowski [BŻ17].

This section begins by discussing quantum states and density matrices, the mathematical representations of states in finite-dimensional quantum systems. We then discuss *quantum channels*, the mathematical representation of the most general physical processes governing quantum systems. We will then discuss a particular representation of quantum channels, called *Choi matrices*. Subsequently, we discuss the different *figures of merit* or *distance measures* used in quantum information, with the most important of them, to this thesis, being *quantum fidelities*.

2.3.1 Quantum states

Throughout this thesis, we will consider only finite-dimensional quantum systems. The *state* of a quantum system is mathematically described by a density matrix, which is a positive semidefinite matrix of unit trace. We use $\mathbb{D}_d$ to denote the set of all $d \times d$ density matrices.

A *pure* state is a density matrix of rank one, and they form the extreme points of the convex set of quantum states. Any pure state can be written in the form $\rho = |\psi\rangle\langle\psi|$ for a unit complex vector $|\psi\rangle \in \mathbb{C}^d$. Thus we usually denote pure states by just the *state vector* $|\psi\rangle$.

Remark 2.3.1. Typically in quantum information, we use the term *state* to refer to density matrices (or unit vectors). However, in this thesis, we will use the word state to refer to positive semidefinite matrices that are not necessarily unit trace. One can implicitly view these as weighted density matrices, with the weight given by the trace of the positive semidefinite matrix.

We now discuss the notion of a *purification* of a state.

Definition 2.3.2. *Let $\mathcal{X}, \mathcal{Y}$ be complex spaces and let $P \in \mathbb{P}_{\mathcal{X}}$. A vector $|u\rangle \in \mathcal{X} \otimes \mathcal{Y}$ is called a* purification *of P if*

$$P = [|u\rangle\langle u|]_{\mathcal{X}} \equiv \mathrm{Tr}_{\mathcal{Y}}[|u\rangle\langle u|]. \tag{2.34}$$

The space $\mathcal{Y}$ is called the *auxiliary* space and for a purification to exist, a necessary and sufficient condition is $\dim(\mathcal{Y}) \geq \operatorname{rank}(P)$. We will typically choose $\mathcal{Y} \cong \mathcal{X}$.

Purifications are not unique, and we denote the set of purifications of a state $P \in \mathbb{P}_d$ as $\operatorname{Pur}(P)$. Moreover, there exists a unitary equivalence between purifications of the same state. That is, if $u, v \in \mathcal{X} \otimes \mathcal{Y}$ are purifications of P then there exists a unitary $U \in \mathbb{U}_{\mathcal{Y}}$ such that

$$|u\rangle = (\mathbb{I}_{\mathcal{X}} \otimes U)|v\rangle. \tag{2.35}$$

Given the eigendecomposition $P = \sum_{i=1}^{d} \lambda_i |u_i\rangle\langle u_i|$, the vector $|u\rangle := \sum_{i=1}^{n} \sqrt{\lambda_i}|u_i\rangle|u_i\rangle \in \mathcal{X} \otimes \mathcal{Y}$ is a purification of P (for $\mathcal{Y} \cong \mathcal{X}$). For proofs and further details, we refer to [Wat18; Wil13].

2.3.2 Quantum channels

The most general way of describing the evolution of a quantum system is via linear maps that are *Completely Positive* and *Trace-preserving*. Such maps are called CPT maps or *quantum channels*. We now discuss some important subclasses of linear maps between matrices. See Watrous [Wat18] for detailed descriptions and proofs of the following statements.

Let $\mathcal{X} = \mathbb{C}^d$ and $\mathcal{Y} = \mathbb{C}^{d'}$ and $\operatorname{LM}(\mathcal{X}, \mathcal{Y})$ denote the set of all linear maps from $\mathbb{M}_{\mathcal{X}}$ to $\mathbb{M}_{\mathcal{Y}}$ and let $\Phi \in \operatorname{LM}(\mathcal{X}, \mathcal{Y})$.

1. Φ is called *Hermitian preserving* if $\Phi(H) \in \mathbb{H}_{\mathcal{Y}}$ for all $H \in \mathbb{H}_{\mathcal{X}}$.

2. Φ is called *positive* if $\Phi(P) \in \mathbb{P}_{\mathcal{Y}}$ for all $P \in \mathbb{P}_{\mathcal{X}}$.

3. Φ is called *completely positive* if for an *auxiliary* space $\mathcal{Z} = \mathbb{C}^k$ of any finite dimension $k \in \mathbb{N}$, the map $\Phi \otimes \operatorname{Id}_{\mathcal{Z}}$ is positive:

$$(\Phi \otimes \operatorname{Id}_{\mathcal{Z}})(P) \in \mathbb{P}_{\mathcal{Y}\otimes\mathcal{Z}} \quad \text{for all } P \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Z}}. \tag{2.36}$$

   Here $\operatorname{Id}_{\mathcal{Z}} : \mathbb{M}_{\mathcal{Z}} \to \mathbb{M}_{\mathcal{Z}}$ is the *identity superoperator* whose action is defined as $\operatorname{Id}_{\mathcal{Z}}(Z) := Z$ for any $Z \in \mathbb{M}_{\mathcal{Z}}$.

4. Φ is called trace-preserving if for all $X \in \mathbb{M}_{\mathcal{X}}$

$$\operatorname{Tr}[\Phi(X)] = \operatorname{Tr}[X]. \tag{2.37}$$

5. Φ is called unital if it maps identity to matrix: $\Phi(\mathbb{I}_{\mathcal{X}}) = \mathbb{I}_{\mathcal{Y}}$.

Thus a quantum channel is a linear map that satisfies both Eq. (2.36) and Eq. (2.37) simultaneously. The set of all CP, CPU, and CPT maps from $\mathbb{M}_{\mathcal{X}}$ to $\mathbb{M}_{\mathcal{Y}}$ are denoted by $\mathrm{CP}(\mathcal{X},\mathcal{Y}), \mathrm{CPU}(\mathcal{X},\mathcal{Y})$, and $\mathrm{CPT}(\mathcal{X},\mathcal{Y})$ respectively.

The *adjoint* of any linear map $\Phi \in \mathrm{LM}(\mathcal{X},\mathcal{Y})$ is the unique linear map $\Phi^* \in \mathrm{LM}(\mathcal{Y},\mathcal{X})$ such that

$$\langle Y, \Phi(X)\rangle = \langle \Phi^*(Y), X\rangle \tag{2.38}$$

for all $X \in \mathbb{M}_{\mathcal{X}}$ and $Y \in \mathbb{M}_{\mathcal{Y}}$. A linear map is completely positive if and only if its adjoint is completely positive and it is trace-preserving if and only if its adjoint is unital. A linear map that is both trace-preserving and unital is called *bistochastic*.

In the above form, a linear map is a rather abstract object. There are other, more concrete ways, of representing quantum channels such as Kraus operators [HK69; Kra71], Choi matrix [Cho75; Jam72], and Stinespring representation [Sti55]. See Watrous [Wat18], Wilde [Wil13], and Tomamichel [Tom15] for detailed expositions.

Choi matrices play a central role in Chapter 5, where optimizations and projections on channels and maps are carried out through their Choi representations. We now explore their key properties. The *Choi-Jamiołkowski isomorphism* bijectively relates linear maps in $\mathrm{LM}(\mathcal{X},\mathcal{Y})$ to bipartite matrices in $\mathbb{M}_{\mathcal{X}\otimes\mathcal{Y}}$. More specifically, the *Choi matrix* of a linear map $\Phi \in \mathrm{LM}(\mathcal{X},\mathcal{Y})$ is defined as

$$J(\Phi) := \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|) = (\mathrm{Id}_{\mathcal{X}} \otimes \Phi)(\Omega) \quad \in \mathbb{M}_{\mathcal{X}\otimes\mathcal{Y}}, \tag{2.39}$$

where $\{|i\rangle\}_{i\in[d]}$ are the computational (standard) basis vectors which span $\mathcal{X} = \mathbb{C}^d$ and $\Omega = |\omega\rangle\langle\omega|$ for the unnormalized maximally entangled state $|\omega\rangle = \sum_{i=1}^{d} |i\rangle|i\rangle$ (in the basis used to define the Choi matrix). Conversely, for any $X \in \mathbb{M}_{\mathcal{X}}$, one can obtain $\Phi(X)$ from the Choi matrix $J(\Phi)$ as

$$\Phi(X) = \mathrm{Tr}_{\mathcal{X}}[J(\Phi)\cdot(X^{\intercal}\otimes\mathbb{I}_{\mathcal{Y}})] \equiv [J(\Phi)\cdot(X^{\intercal}\otimes\mathbb{I}_{\mathcal{Y}})]_{\mathcal{Y}}. \tag{2.40}$$

where the transpose is taken in the basis used to define the Choi matrix, which we will always take as the computational basis. Some important facts regarding the Choi-Jamiołkowski isomorphism follows.

1. A linear map is Hermitian-preserving if and only if its Choi matrix is Hermitian.

2. A linear map is completely positive if and only if its Choi matrix is positive semidefinite.

3. A linear map is trace-preserving if and only if the *input-marginal* of its Choi matrix is the identity matrix: $[J(\Phi)]_{\mathcal{X}} = \mathbb{I}_{\mathcal{X}}$.

4. A linear map is unital if and only if its *output-marginal* is the identity matrix: $[J(\Phi)]_{\mathcal{Y}} = \mathbb{I}_{\mathcal{Y}}$.

Thus a linear map $\Phi \in \mathrm{M}(\mathcal{X}, \mathcal{Y})$ is a quantum channel if and only if its Choi matrix is positive semidefinite and has $\mathcal{X}$-marginal to be identity:

$$\Phi \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y}) \quad \iff \quad J(\Phi) \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}} \text{ and } \mathrm{Tr}_{\mathcal{Y}}[J(\Phi)] = \mathbb{I}_{\mathcal{X}}. \tag{2.41}$$

The set of Choi matrices corresponding to the quantum channels (CPT maps) and CPU maps will be denoted by $\mathrm{J}_{\mathrm{CPT}}(\mathcal{X}, \mathcal{Y}), \mathrm{J}_{\mathrm{CPU}}(\mathcal{X}, \mathcal{Y}) \subset \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$ respectively.

Typically we would be interested in the application of a completely positive map $\Phi \in \mathrm{CP}(\mathcal{X}, \mathcal{Y})$ on a positive semidefinite input state $P \in \mathbb{P}_d$. In such a scenario, one can use the star product to compactify Eq. (2.40) as follows.

$$\Phi(P) = [J(\Phi) \cdot (P^{\intercal} \otimes \mathbb{I}_{\mathcal{Y}})]_{\mathcal{Y}} = [J(\Phi) \star_{\mathcal{X}} P^{\intercal}]_{\mathcal{Y}}. \tag{2.42}$$

The RHS succinctly describes how to get the action of the channel from its Choi matrix—*star* the input state on the input space and then consider only the output space (by discarding the input space).

Transposition and Choi matrices

We now discuss some ideas about the transposition of Choi matrices of linear maps. Some of the following ideas have been explored in Bengtsson and Życzkowski [BŻ17, Chapter 11]. Recall that both transpose operation and definition of Choi matrices are basis dependent. Whenever a transpose operation and a Choi matrix appear together, the basis used to define the Choi matrix and transpose map must agree. In this section, without loss of generality, we assume the basis to be the standard (computational) basis.

For an arbitrary linear map $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$, define the *transpose twirl of* Φ as

$$\Phi^{\intercal} \equiv \intercal \circ \Phi \circ \intercal \in \mathrm{LM}(\mathcal{X}, \mathcal{Y}), \tag{2.43}$$

where $\intercal$ denotes the transpose map with respect to the standard basis. Note that in the RHS of Eq. (2.43), the *right-most* transpose is to be taken with respect to the standard basis of $\mathcal{X}$ and the *left-most* transpose is to be taken with respect to the standard basis of $\mathcal{Y}$.

Observe that for any $X \in \mathbb{M}_{\mathcal{X}}$ we have $\Phi^{\intercal}(X) = \Phi(X^{\intercal})^{\intercal}$. We now show that the Choi matrices of Φ and $\Phi^{\intercal}$ are mutual transposes, which justifies the notation.

Proposition 2.3.3. *Let $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$ be an arbitrary map and denote $\Phi^\intercal \equiv \intercal \circ \Phi \circ \intercal \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$. It then holds*

$$J(\Phi)^\intercal = J(\Phi^\intercal). \tag{2.44}$$

Proof. The Choi matrix $J(\Phi^\intercal)$ can be written as

$$J(\Phi^\intercal) := \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes \Phi^\intercal(|i\rangle\langle j|) = \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes \Phi(|j\rangle\langle i|)^\intercal, \tag{2.45}$$

where in the last equality we have used the fact that $\Phi^\intercal = \intercal \circ \Phi \circ \intercal$. Take transpose across to get

$$J(\Phi^\intercal)^\intercal = \sum_{i,j=1}^{d} |j\rangle\langle i| \otimes \Phi(|j\rangle\langle i|) = J(\Phi). \tag{2.46}$$

which is equivalent to $J(\Phi^\intercal) = J(\Phi)^\intercal$ as required. □

It is this transpose relation between the channels that motivates the notation $\Phi^\intercal$. We call a map Φ *transpose-preserving* if $\Phi(X^\intercal) = \Phi(X)^\intercal$ for all $X \in \mathbb{M}_\mathcal{X}$. We now show that a map is transpose-preserving if and only if its Choi matrix is symmetric: $J(\Phi)^\intercal = J(\Phi)$.

Theorem 2.3.4. *Let $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$. Then Φ is transpose-preserving if and only if its Choi matrix is symmetric:*

$$\Phi(X^\intercal) = \Phi(X)^\intercal \text{ for all } X \in \mathbb{M}_\mathcal{X} \quad \iff \quad J(\Phi)^\intercal = J(\Phi). \tag{2.47}$$

Proof. Towards proving the forward direction, assume Φ is transpose-preserving. We then have

$$J(\Phi) := \sum_{i,j=1}^{n} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|) = \sum_{i,j=1}^{n} |i\rangle\langle j| \otimes \Phi(|j\rangle\langle i|)^\intercal. \tag{2.48}$$

Take transpose across to get

$$J(\Phi)^\intercal = \sum_{i,j=1}^{n} |j\rangle\langle i| \otimes \Phi(|j\rangle\langle i|) = J(\Phi), \tag{2.49}$$

which proves the forward direction. For the reverse direction, assume $J(\Phi)^\intercal =$

$J(\Phi)$. For any $X \in \mathbb{M}_{\mathcal{X}}$, we then have

$$\begin{aligned}\Phi(X^{\intercal}) = [J(\Phi) \cdot_x (X^{\intercal})^{\intercal}]_{\mathcal{Y}} &= [J(\Phi)^{\intercal} \cdot_x (X^{\intercal})^{\intercal}]_{\mathcal{Y}} \\ &= [(J(\Phi) \cdot_x X^{\intercal})^{\intercal}]_{\mathcal{Y}} = [J(\Phi) \cdot_x X^{\intercal}]_{\mathcal{Y}}^{\intercal} = \Phi(X)^{\intercal}.\end{aligned} \tag{2.50}$$

Here we have used the shorthand $A \cdot_x X \equiv A \cdot (X \otimes \mathbb{I}_{\mathcal{Y}})$ for $A \in \mathbb{M}_{\mathcal{X}\otimes\mathcal{Y}}, X \in \mathbb{M}_{\mathcal{X}}$, and $Y \in \mathbb{M}_{\mathcal{Y}}$. In the second equality, we have used the assumption $J(\Phi)^{\intercal} = J(\Phi)$, in the third equality we have used $(A^{\intercal}B^{\intercal}) = (BA)^{\intercal}$ followed by cyclic property of partial trace, and in the fourth equality we have used the fact that partial tracing commutes with transpose: $\mathrm{Tr}_{\mathcal{X}}[(X \otimes Y)^{\intercal}] = \mathrm{Tr}_{\mathcal{X}}[X^{\intercal} \otimes Y^{\intercal}] = \mathrm{Tr}[X]Y^{\intercal}$, which is extended linearly to arbitrary arguments. Having proven the reverse direction, we conclude the proof. □

We now discuss some properties of the transpose twirl of a map.

Proposition 2.3.5. *Let $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$. It holds that*

1. *Φ is Hermitian preserving if and only if $\Phi^{\intercal}$ is Hermitian preserving.*
2. *Φ is completely positive if and only if $\Phi^{\intercal}$ is completely positive.*
3. *Φ is trace-preserving if and only if $\Phi^{\intercal}$ is trace-preserving.*

Proof. We will use properties of the Choi matrices $J(\Phi)$ and $J(\Phi^{\intercal})$. For the first statement, observe that

$$\Phi \text{ is HP} \iff J(\Phi) \in \mathbb{H}_{\mathcal{X}\otimes\mathcal{Y}} \iff J(\Phi)^{\intercal} = J(\Phi^{\intercal}) \in \mathbb{H}_{\mathcal{X}\otimes\mathcal{Y}} \iff \Phi^{\intercal} \text{ is HP}. \tag{2.51}$$

Similarly, we have

$$\Phi \text{ is CP} \iff J(\Phi) \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}} \iff J(\Phi)^{\intercal} = J(\Phi^{\intercal}) \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}} \iff \Phi^{\intercal} \text{ is CP}. \tag{2.52}$$

For the final statement, observe that

$$\Phi \text{ is TP} \iff [J(\Phi)]_{\mathcal{X}} = \mathbb{I}_{\mathcal{X}} \iff [J(\Phi^{\intercal})]_{\mathcal{X}} = \mathbb{I}_{\mathcal{X}} \iff \Phi^{\intercal} \text{ is TP}. \tag{2.53}$$

The second bidirectional implication follows from the fact that $J(\Phi^{\intercal}) = J(\Phi)^{\intercal}$ and transpose map commutes with partial trace, which implies $[J(\Phi^{\intercal})]_{\mathcal{X}} = [J(\Phi)]_{\mathcal{X}}^{\intercal} = \mathbb{I}_{\mathcal{X}}$. □

An immediate corollary is that the transpose twirl of a quantum channel is a quantum channel. Furthermore, the transpose twirl commutes with the adjoint operation, as shown in the following proposition.

Proposition 2.3.6. *For any linear map $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$ it holds that*

$$(\Phi^\intercal)^* = (\Phi^*)^\intercal. \tag{2.54}$$

Proof. The proof easily follows from observing that for arbitrary (compatible) maps Ψ_1 and Ψ_2, we have $(\Psi_1 \circ \Psi_2)^* = \Psi_2^* \circ \Psi_1^*$. We then have

$$\Phi^{\intercal *} = (\intercal \circ \Phi \circ \intercal)^* = \intercal \circ \Phi^* \circ \intercal = \Phi^{*\intercal}. \tag{2.55}$$

In the second equality, we have noted that the transpose map is self-adjoint. □

Recall that, by Choi-Jamiołkowski isomorphism, every bipartite matrix $C \in \mathbb{M}_{\mathcal{X}\otimes\mathcal{Y}}$ defines a linear map from $\mathbb{M}_\mathcal{X}$ to $\mathbb{M}_\mathcal{Y}$. Indeed, by symmetry, it must also define a linear map from $\mathbb{M}_\mathcal{Y}$ to $\mathbb{M}_\mathcal{X}$. The action of these maps can be written as follows.

$$X \mapsto [C \cdot (X^\intercal \otimes \mathbb{I}_\mathcal{Y})]_\mathcal{Y} \quad \text{and} \quad Y \mapsto [C \cdot (\mathbb{I}_\mathcal{X} \otimes Y^\intercal)]_\mathcal{X}. \tag{2.56}$$

Denote these maps as $\Phi_C \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$ and $\Psi_C \in \mathrm{LM}(\mathcal{Y}, \mathcal{X})$ respectively. Note that in the second case, we have multiplied the *input* $Y^\intercal$ in the *second* space to maintain compatibility. By Proposition 2.3.3, we have

$$\Phi_C^\intercal = \Phi_{C^\intercal} \quad \text{and} \quad \Psi_C^\intercal = \Psi_{C^\intercal}. \tag{2.57}$$

Furthermore, as we show in the following proposition, $\Phi_C = \Psi_{C^\intercal}^*$ for Hermitian preserving maps. That is, the HP map from $\mathbb{M}_\mathcal{X}$ to $\mathbb{M}_\mathcal{Y}$ associated with any $C \in \mathbb{H}_{\mathcal{X}\otimes\mathcal{Y}}$ is the adjoint of the HP map $\mathbb{H}_\mathcal{Y}$ to $\mathbb{M}_\mathcal{X}$ associated with $C^\intercal$.

Proposition 2.3.7. *Let $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$ be a Hermitian preserving map and let $J(\Phi) \in \mathbb{H}_{\mathcal{X}\otimes\mathcal{Y}}$ be its Choi matrix. It then holds*

$$J(\Phi)^\intercal = J(\Phi^*). \tag{2.58}$$

where $\Phi^* \in \mathrm{LM}(\mathcal{Y}, \mathcal{X})$ is the adjoint map of Φ.

Proof. Let $J(\Phi)^\intercal \in \mathbb{H}_{\mathcal{X}\otimes\mathcal{Y}}$ define the linear map Ψ from $\mathbb{M}_\mathcal{Y}$ to $\mathbb{M}_\mathcal{X}$ with the action:

$$\Psi(Y) := [J(\Phi)^\intercal (\mathbb{I}_\mathcal{X} \otimes Y^\intercal)]_\mathcal{X}, \tag{2.59}$$

for any $Y \in \mathbb{M}_\mathcal{Y}$. We aim to show that $\Psi(Y) = \Phi^*(Y)$, which is equivalent to showing that $\langle Y, \Phi(X)\rangle = \langle \Psi(Y), X\rangle$, for all $X \in \mathbb{M}_\mathcal{X}$ and $Y \in \mathbb{M}_\mathcal{Y}$. Observe that

$$
\begin{aligned}
\langle X, \Psi(Y)\rangle &= \mathrm{Tr}\left[X^* \Psi(Y)\right] \\
&= \mathrm{Tr}\left[[J(\Phi)^\intercal(\mathbb{I}_\mathcal{X} \otimes Y^\intercal)]_\mathcal{X} X^*\right] && \text{definition of } \Psi(Y) \\
&= \mathrm{Tr}\left[J(\Phi)^\intercal(\mathbb{I}_\mathcal{X} \otimes Y^\intercal)(X^* \otimes \mathbb{I}_\mathcal{Y})\right] && \text{adjoint of } \mathrm{Tr}_\mathcal{Y} \\
&= \mathrm{Tr}\left[J(\Phi)^\intercal(X^* \otimes Y^\intercal)\right] \\
&= \mathrm{Tr}\left[J(\Phi)(\overline{X} \otimes Y)\right] && \mathrm{Tr}[A^\intercal] = \mathrm{Tr}[A] \text{ for any } A \\
&= \mathrm{Tr}\left[J(\Phi)(\overline{X} \otimes \mathbb{I}_\mathcal{Y})(\mathbb{I}_\mathcal{X} \otimes Y)\right] && (2.60) \\
&= \mathrm{Tr}\left[[J(\Phi)(\overline{X} \otimes \mathbb{I}_\mathcal{Y})]_\mathcal{Y} Y\right] && \text{adjoint of tensoring } \mathbb{I}_\mathcal{X} \\
&= \mathrm{Tr}\left[[J(\Phi)((X^*)^\intercal \otimes \mathbb{I}_\mathcal{Y})]_\mathcal{Y} Y\right] \\
&= \mathrm{Tr}[\Phi(X^*)Y] \\
&= \mathrm{Tr}[\Phi(X)^* Y] && \Phi \text{ is Hermitian-preserving} \\
&= \langle \Phi(X), Y\rangle,
\end{aligned}
$$

for any $X \in \mathbb{M}_\mathcal{X}$ and $Y \in \mathbb{M}_\mathcal{Y}$, which implies $\Psi = \Phi^*$. Thus, for any HP map Φ, we have $J(\Phi)^\intercal = J(\Phi^*)$ as claimed. □

Thus, for a Hermitian preserving map $\Phi \in \mathrm{LM}(\mathcal{X}, \mathcal{Y})$, we have the following equalities:

$$J(\Phi)^\intercal = J(\Phi^\intercal) = J(\Phi^*) = J(\Phi^{\intercal *})^\intercal \tag{2.61}$$

Some examples of transpose-preserving maps include the trace map, the partial trace map, and the depolarizing channel.

2.3.3 Metrics and figures of merit

To make quantitative statements about a quantum system, we must be able to compare two quantum states. More formally, we need functions that take in two quantum states and return real numbers which can be used to quantify the closeness of these quantum states. Formally such a function is called a *figure of merit* (FoM). We now briefly discuss some popular figures of merit used in quantum information. We begin with quantum fidelities, a class of figures of merit with attractive properties, which form the basis of much of this thesis. We will then briefly discuss other figures of merit for the sake of completeness.

Quantum fidelities

Quantum fidelities are the non-commutative generalizations (a.k.a. *quantizations*) of *classical fidelity*, which is more commonly known as the Bhattacharyya coefficient. The classical fidelity between two positive vectors $p, q \in \mathbb{R}^d_+$ is defined

as

$$\mathrm{F}_{\mathrm{Cl}}(p,q) := \left\langle p^{\frac{1}{2}}, q^{\frac{1}{2}} \right\rangle = \sum_{i=1}^{d} \sqrt{p_i q_i}, \tag{2.62}$$

where square roots of vectors are defined elementwise: $\left[p^{\frac{1}{2}}\right]_i := [p]_i^{\frac{1}{2}}$. Observe that the classical fidelity can be seen as the sum of the geometric mean of the two positive vectors p and q.

There is no unique way to quantize classical fidelity. We now discuss three well-studied quantizations of classical fidelity. These are *Holevo fidelity, Uhlmann fidelity,* and *Matsumoto fidelity.*

Holevo fidelity

Holevo fidelity [Kho72] is perhaps the most straightforward way to quantize classical fidelity:

$$\mathrm{F}_{\mathrm{Cl}}(p,q) := \left\langle p^{\frac{1}{2}}, q^{\frac{1}{2}} \right\rangle \mapsto \left\langle P^{\frac{1}{2}}, Q^{\frac{1}{2}} \right\rangle = \mathrm{Tr}\left[P^{\frac{1}{2}} Q^{\frac{1}{2}}\right] =: \mathrm{F}^{\mathrm{H}}(P,Q). \tag{2.63}$$

That is, we simply promote the Euclidean inner product of positive vectors to the Euclidean (a.k.a. Hilbert–Schmidt) inner product of positive semidefinite matrices.

Matsumoto fidelity

The Matsumoto fidelity [Mat10; Mat14; CS20] is obtained by taking the *geometric mean* part literally, which leads to the following definition:

$$\mathrm{F}^{\mathrm{M}}(P,Q) := \mathrm{Tr}[P\#Q] := \mathrm{Tr}\left[P^{\frac{1}{2}} \sqrt{P^{-\frac{1}{2}} Q P^{-\frac{1}{2}}}\, P^{\frac{1}{2}}\right], \tag{2.64}$$

where $P\#Q$ is the matrix geometric mean between P and Q.

Uhlmann fidelity

However the most popular quantum fidelity is the *Uhlmann fidelity* [Uhl76; Joz94]:

$$\mathrm{F}^{\mathrm{U}}(P,Q) := \left\| P^{\frac{1}{2}} Q^{\frac{1}{2}} \right\|_1 = \mathrm{Tr}\left[\sqrt{Q^{\frac{1}{2}} P Q^{\frac{1}{2}}}\right]. \tag{2.65}$$

The popularity of Uhlmann fidelity is exemplified by the fact that it is often referred to as simply *fidelity.* We collect some important properties of Uhlmann fidelity in the following proposition.

Proposition 2.3.8. *The Uhlmann fidelity satisfies the following properties.*

1. *Joint concavity [Wil13, Property 9.2.2] and [Wat18, Corollary 3.26]. For any $P_0, P_1, Q_0, Q_1 \in \mathbb{P}_d$, and any $\lambda \in [0,1]$, we have*

$$\mathrm{F}^{\mathrm{U}}((1-\lambda)P_0+\lambda P_1, (1-\lambda)Q_0+\lambda Q_1) \geq \lambda\,\mathrm{F}^{\mathrm{U}}(P_0,Q_0)+(1-\lambda)\,\mathrm{F}^{\mathrm{U}}(P_1,Q_1). \tag{2.66}$$

2. *Strict concavity in either variable [BJL19; BJL18]. For any $P, Q_0, Q_1 \in \mathbb{P}_d^+$ such that $Q_0 \neq Q_1$ and $\lambda \in (0,1)$, it holds that*

$$\mathrm{F}^{\mathrm{U}}(P, \lambda Q_0 + (1-\lambda)Q_1) > \lambda\,\mathrm{F}^{\mathrm{U}}(P,Q_0) + (1-\lambda)\,\mathrm{F}^{\mathrm{U}}(P,Q_1). \tag{2.67}$$

3. *The gradient of fidelity is given by [BJL19; BJL18; BŻ17]*

$$\nabla_Q\,\mathrm{F}^{\mathrm{U}}(P,Q) \equiv \nabla \mathrm{F}_P(Q) = \frac{1}{2}Q^{-1}\#P, \tag{2.68}$$

where we denoted $\mathrm{F}_P(Q) \equiv \mathrm{F}^{\mathrm{U}}(P,Q)$ to be the Uhlmann fidelity of $Q > 0$ with a fixed state $P > 0$.

For any $P, Q \in \mathbb{P}_d$, it holds that [Mat10; CS20]

$$\mathrm{F}^{\mathrm{M}}(P,Q) \leq \mathrm{F}^{\mathrm{H}}(P,Q) \leq \mathrm{F}^{\mathrm{U}}(P,Q). \tag{2.69}$$

Hilbert–Schmidt distance

The Hilbert–Schmidt distance between two matrices A and B is defined as

$$\|A-B\|_2 := \sqrt{\langle A-B, A-B\rangle} = \sqrt{\mathrm{Tr}[(A-B)^*(A-B)]}, \tag{2.70}$$

where $\langle X, Y\rangle := \mathrm{Tr}[X^*Y]$ is the Hilbert–Schmidt (Euclidean) inner product on the space $\mathbb{M}_d$. The main benefit of the Hilbert–Schmidt distance is the mathematical ease of handling. It is also the natural distance of the Riemannian manifold of Hermitian matrices once equipped with the *Euclidean metric*.

2.3.4 Data processing inequality

In the previous section, we discussed various figures of merits used in quantum information. However, not all FoMs are created equal, with some being more equal[1] than others. Essentially, not every FoM is useful for data processing tasks.

[1] or *inequal*, as the title suggests.

The guiding principle for choosing a FoM is as follows.

Data subjected noise should not become more distinguishable.

Since FoMs can characterize distinguishability and noise is modeled through channels in quantum information, the above statement can be mathematically summarized via the *data processing inequality* (DPI):

$$\mathrm{D}(\rho, \sigma) \geq \mathrm{D}(\Lambda(\rho), \Lambda(\sigma)), \tag{2.71}$$

where ρ, σ are quantum states, Λ is a quantum channel, and D is a figure of merit. Indeed when it comes to certain quantum information tasks such as quantum communication, we need FoMs that obey the above inequality for any triple (Λ, ρ, σ).

Some examples of FoMs that obey DPI include the three fidelities mentioned in the previous section, trace distance [NC01, Theorem 9.2], Umegaki relative entropy [Ume62; Lin75; Uhl77]. An important non-example is the HS distance, which does not obey DPI [Oza00].

Suppose D is an FoM that obeys DPI. We say a triple (Λ, ρ, σ) *saturates DPI* for D if

$$\mathrm{D}(\rho, \sigma) = \mathrm{D}(\Lambda(\rho), \Lambda(\sigma)). \tag{2.72}$$

Finding conditions for exact [Pet86b; Pet88; Hay+04; LRD17] and approximate [FR15; Jun+18; Wil15; STH16; CS22a] saturation forms a crucial research direction in quantum information theory.

Uhlmann fidelity and quantum relative entropy are two figures of merit for which saturation of DPI is important in the context of this thesis. Petz showed [Pet86b; Pet88] that saturation of DPI for Umegaki relative entropy is equivalent to the existence of the so-called *Petz recovery map*, whose geometric interpretation is studied in Chapter 5. The saturation of DPI for Uhlmann fidelity (more generally, α-sandwiched Rényi divergences) was studied in [LRD17] and the algebraic condition for saturation of DPI derived therein forms a central piece of the closed-forms for projections with respect to Bures distance, which are introduced and studied in Chapter 5.

2.4 Riemannian geometry

Certain results discussed in this thesis rely on the Riemannian geometry of the Bures–Wasserstein manifold. Thus we briefly (and slightly informally) discuss necessary concepts from Riemannian geometry, while pointing to Lee [Lee18] and Do Carmo and Flaherty Francis [DF92] for a more thorough and rigorous treatment. See Boumal [Bou23] for an excellent treatment of optimization on manifolds.

In this thesis, we will be mostly interested in the *Bures* manifold (a.k.a Bures–Wasserstein manifold) of positive definite matrices. However, a more familiar manifold is the manifold of Hermitian matrices equipped with the Hilbert–Schmidt metric, whose geometry is analogous to the familiar geometry of Euclidean vector spaces like $\mathbb{R}^d$. Thus each concept we discuss here would be exemplified by instances from the Euclidean manifold (of Hermitian matrices) and the Bures manifold. We will use 'Bures' and 'Bures–Wasserstein' (abbreviated as BW) interchangeably in this chapter and the rest of this thesis. The name 'Wasserstein' comes from the fact that the 2-Wasserstein distance between centered Gaussian probability measures is equal to the Bures distance between their covariance matrices. See [BJL19] for details of this connection.

2.4.1 Riemannian manifolds

Informally, a manifold is a set that *looks flat* when *zoomed in* sufficiently. The notions of *looks flat* and *zoomed in* are formalized using a set (called *atlas*) of homeomorphisms (called *charts*) from the manifold to open subsets $\mathbb{R}^m$ of appropriate dimension m.

However, for our purposes, a discussion on charts, atlases, and other fundamental concepts would be an overkill. Instead, we reason that the set of Hermitian matrices and positive definite matrices are *smooth manifolds* (see [Lee12] for a precise definition) based on the two following facts.

1. A finite-dimensional vector space is a smooth manifold (see [Lee12, Examples 1.5, 1.6, 1.7] and [AMS08, Sec. 3.1.4, 3.1.5]).

2. Any open subset of a smooth manifold is a smooth manifold (called an *open submanifold*) [Lee12, Example 1.8].

Recall that the set of Hermitian matrices $\mathbb{H}_d$ is a real vector space of dimension d^2, which makes $\mathbb{H}_d$ a smooth manifold. Moreover, the set of positive definite matrices $\mathbb{P}_d^+$ is an open subset of $\mathbb{H}_d$. Thus we have $\mathbb{P}_d^+$ to be a smooth manifold.

A Riemannian manifold is a smooth manifold $\mathcal{M}$ equipped with a *metric tensor* $\mathfrak{g}$, which endows the tangent space at every point $p \in \mathcal{M}$ with a real inner product $\mathfrak{g}_p$. The metric tensor can be used to measure angles between curves on a manifold. The concepts of tangent space and metric tensors are detailed next.

2.4.2 Tangent space

We now discuss the *tangent space* at a point on a smooth manifold $\mathcal{M}$. While the formal definition requires concepts from differential geometry, for our purposes the following would suffice.

For a point $p \in \mathcal{M}$, the tangent space $\mathrm{T}_p\mathcal{M}$ is the vector space of all *tangent vectors* at a given point $p \in \mathcal{M}$. One can visualize a tangent vector $v \in \mathrm{T}_p\mathcal{M}$ as an *initial velocity vector* of a smooth curve starting at p, tracing out a path that stays in $\mathcal{M}$, at least infinitesimally.

For any finite-dimensional vector space $\mathcal{V}$, the tangent space $\mathrm{T}_v\mathcal{V}$ at any $v \in \mathcal{V}$ is isomorphic to $\mathcal{V}$ itself: $\mathrm{T}_v\mathcal{V} \cong \mathcal{V}$ [Lee12, Proposition 3.8]. Thus for any $H \in \mathbb{H}_d$, we have $\mathrm{T}_H\mathbb{H}_d \cong \mathbb{H}_d$. If $\mathcal{U}$ is an open subset of $\mathcal{V}$, then $\mathcal{U}$ is also a smooth manifold, with $\mathrm{T}_u\mathcal{U} \cong \mathrm{T}_u\mathcal{V} \cong \mathcal{V}$ [Lee12, Propositions 3.6 and 3.7]. Thus for any $P \in \mathbb{P}_d^+$, the tangent space $\mathrm{T}_P\mathbb{P}_d^+ \cong \mathrm{T}_P\mathbb{H}_d \cong \mathbb{H}_d$.

Essentially the above two sections have been to establish that $\mathbb{P}_d^+$ and $\mathbb{H}_d$ are smooth manifolds and the tangent space at any point in the above manifolds is isomorphic to $\mathbb{H}_d$. These facts have been summarized in [CS22b, Section 2]. Also see [Jen03; httb]. The disjoint union of the tangent spaces is called the *tangent bundle*:

$$\mathrm{T}\mathcal{M} := \bigsqcup_{p \in \mathcal{M}} \mathrm{T}_p\mathcal{M}. \tag{2.73}$$

2.4.3 Metric tensor

Recall that the tangent space is a vector space. If the underlying manifold is modeled after real vector spaces (which is indeed the case for all Riemannian manifolds), then one can endow the tangent space with a real inner product (cf. Section 2.1.2) to turn each tangent space into a real inner product space. This is exactly what a Riemannian metric does. A Riemannian metric $\mathfrak{g}$ associates a smoothly varying real inner product to each tangent space $\mathrm{T}_p\mathcal{M}$. More formally

$$p \mapsto \mathfrak{g}(p) \equiv \mathfrak{g}_p \equiv \langle \cdot, \cdot \rangle_p, \tag{2.74}$$

where $\mathfrak{g}_p(\cdot, \cdot) \equiv \langle \cdot, \cdot \rangle_p$ is a real inner product on the tangent space $\mathrm{T}_p\mathcal{M}$.

By endowing the manifold with a metric tensor, we can perform various geometric operations on the manifold such as measuring the length of curves and

angles between curves. Moreover, the geometry of the manifold changes as we change the metric tensor. For the subsequent parts of this section, we discuss two metrics—the HS metric and the Bures metric. The HS metric can be thought of as the generalization of the *flat* Euclidean metric on $\mathbb{R}^n$ to $\mathbb{H}_d$. To understand the geometry of the Bures manifold (the manifold of positive definite matrices equipped with the Bures metric tensor), we refer to [BŻ17]. For further details of some of the concepts discussed in this section, we refer to [BJL19; MMP18; Han+21].

The standard choice of metric tensor over the vector space of Hermitian matrices is the Hilbert–Schmidt inner product:

$$\mathfrak{g}_H^{\mathrm{HS}}(U, V) = \langle U, V \rangle := \mathrm{Tr}[UV], \tag{2.75}$$

for all $H \in \mathbb{H}_d$ and $U, V \in \mathrm{T}_H \mathbb{H}_d \cong \mathbb{H}_d$.

Another important metric is the Bures metric defined over the manifold of positive definite matrices $\mathbb{P}_d^+$, which is an open subset of Hermitian matrices $\mathbb{H}_d$. The Bures metric tensor is defined as (see [BJL19, Eq. 29])

$$\begin{aligned} \mathfrak{g}_P^{\mathrm{Bu}}(U, V) \equiv \langle U, V \rangle_P^{\mathrm{Bu}} &:= \mathrm{Re}\,\mathrm{Tr}[\mathcal{L}_P(U) P \mathcal{L}_P(V)] \\ &= \frac{1}{2}\mathrm{Tr}[\mathcal{L}_P(U)V] = \frac{1}{2}\mathrm{Tr}[\mathcal{L}_P(V)U], \end{aligned} \tag{2.76}$$

where, for $P > 0$, $\mathcal{L}_P(U)$ is the (linear) *Lyapunov operator* which is implicitly defined as the unique Hermitian solution to the matrix Lyapunov equation $U = P\mathcal{L}_P(U) + \mathcal{L}_P(U)P$ [Syl84; BR97]. We refer to [MMP18; Bha09] for further properties of the Lyapunov operator. For completeness, we derive the above form of the metric tensor of the Bures manifold starting from the infinitesimal form [Ben98, Eq. 9.43]

$$\mathrm{B}(P, P + \mathrm{d}P) = \frac{1}{2}\mathrm{Tr}[\mathrm{d}P \mathcal{L}_P(\mathrm{d}P)] \tag{2.77}$$

(a.k.a square of line element) in Appendix A.1. This expression gives the squared Bures distance between P and $P + \mathrm{d}P$, where $\mathrm{d}P$ is interpreted as a tangent vector and $P + \mathrm{d}P$ is to be interpreted as an infinitesimally close point to P.

As one would expect, the Hilbert–Schmidt metric tensor leads to the Hilbert–Schmidt distance over Hermitian matrices, and the Bures metric leads to the Bures distance over positive matrices.

2.4.4 Geodesics and distances

We content ourselves with informal descriptions of geodesics in general followed by explicit definitions of geodesics of interest to us. For the formal and rigorous definition of geodesics, we refer to Lee [Lee18] and Vishnoi [Vis18].

Let $p, q \in \mathcal{M}$. Informally, any smooth *length-minimizing* curve[2] between p and q is a geodesic [Lee12, Theorem 6.6]. The converse is true only locally. If p and q are sufficiently close, then the geodesic is the curve of the minimal length [Lee12, Theorem 6.12].

The explicit forms of the geodesics with respect to the HS metric on $\mathbb{H}_d$ and Bures metric on $\mathbb{P}_d^+$ are given as

$$\gamma_{HK}^{\mathrm{Euc}}(t) := (1-t)H + tK \qquad \text{for all } H, K \in \mathbb{H}_d, \tag{2.78}$$

and

$$\begin{aligned} \gamma_{PQ}^{\mathrm{Bu}}(t) &:= [(1-t)\mathbb{I} + tP^{-1}\#Q]P[(1-t)\mathbb{I} + tP^{-1}\#Q] \\ &= P \star [(1-t)\mathbb{I} + tP^{-1}\#Q]^2, \qquad \text{for all } P, Q \in \mathbb{P}_d^+, \end{aligned} \tag{2.79}$$

where $t \in [0,1]$. See Vishnoi [Vis18] and Lee [Lee18] for the Euclidean geodesic definition and Bhatia, Jain, and Lim [BJL19, Eq. 39] for the definition of the Bures geodesic. It is easy to see that the endpoints are given at $t = 0$ and $t = 1$ respectively. The Euclidean geodesic can be extended over the whole real interval (as opposed to just $t \in [0,1]$) while still staying in $\mathbb{H}_d$. This is however not the case with the Bures geodesic. There exists $t \in \mathbb{R}$ such that $\gamma_{PQ}^{\mathrm{Bu}}(t)$ is singular, thus not positive definite. To see this, choose $t = \frac{1}{1-\lambda}$ where λ is any eigenvalue of $P^{-1}\#Q$.

The distance between two points is given by the infimum of the length of all smooth curves connecting the points. Indeed the HS distance and Bures distance are given by

$$\mathrm{d_{HS}}(H,K) = \|H - K\|_2 \quad \text{and} \quad \mathrm{d_{Bu}}(P,Q) = \sqrt{\mathrm{Tr}[P+Q] - 2\,\mathrm{F^U}(P,Q)}, \tag{2.80}$$

for any $H, K \in \mathbb{H}_d$ and $P, Q \in \mathbb{P}_d$.

2.4.5 Total derivative, gradient, and Riemannian gradient

We begin with a short digression on finite-dimensional inner product spaces and their duals. Let $(\mathcal{V}, \langle\cdot,\cdot\rangle)$ be a finite-dimensional real (or complex) inner product space. Any linear functional on $\mathcal{V}$ can be uniquely identified with a *dual vector* (a.k.a. covector) by Riesz representation theorem. The dual space of $\mathcal{V}$ is a vector

[2] Additional qualifiers exist. See [Lee18].

space consisting of all linear functionals on elements of $\mathcal{V}$, denoted by $\mathcal{V}^*$. These spaces are isomorphic and the identifications are made via the inner product:

$$g_u \in \mathcal{V}^* \text{ is dual of } u \in \mathcal{V} \quad \text{if and only if} \quad g_u(v) = \langle u, v \rangle \quad \text{for all } v \in \mathcal{V}. \tag{2.81}$$

The spaces $\mathcal{V}$ and $\mathcal{V}^*$ are mutual duals with $(\mathcal{V}^*)^* = \mathcal{V}$. The dual of a *ket* vector $|v\rangle \in \mathcal{V}$ is typically denoted by its *bra* vector $\langle v| \in \mathcal{V}^*$.

Total derivative

We now discuss the *Fréchet derivative* (a.k.a. total derivative) of real-valued functions over Hermitian matrices. See Bhatia [Bha13, Appendix] for a review of derivatives of functions over matrices.

Let $\mathcal{X} \subseteq \mathcal{V}$ for a real inner product[3] space $\mathcal{V}$ and let $f : \mathcal{X} \to \mathbb{R}$ be a real-valued function. f is said to be *differentiable* at $x \in \mathcal{X}$ if there exists a linear map $Z : \mathcal{X} \to \mathbb{R}$ such that

$$\lim_{\|v\| \to 0} \frac{\|f(x+v) - f(x) - Zv\|}{\|v\|} = 0 \tag{2.82}$$

for all $v \in \mathcal{V}$. Such a Z is called the *Fréchet derivative* or *total derivative* of f at x and is denoted as $\mathrm{D}f(x)$. If f is differentiable at every point in its domain, we say it is differentiable. The action of $\mathrm{D}f(x)$ on an element v can be explicitly computed as

$$\mathrm{D}f(x)(v) := \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} f(x + tv) \tag{2.83}$$

Observe that $\mathrm{D}f(x) : \mathcal{X} \to \mathbb{R}$ is a linear functional for every $x \in \mathcal{X}$. Or equivalently, $\mathrm{D}f : \mathcal{X} \to \mathcal{V}^*$. If f is a linear function, then $\mathrm{D}f(x) = f$ for all $x \in \mathcal{X}$. We also use the equivalent notation $\mathrm{D}f_x \equiv \mathrm{D}f(x)$ for the total derivative.

Gradient

The *gradient* of f at x is the dual of its derivative at x. That is, for any point $x \in \mathcal{X}$, we have the gradient $\nabla f(x)$ to be the unique vector in $\mathcal{X}$ such that

$$\mathrm{D}f(x)(v) = \langle \nabla f(x), v \rangle \quad \text{for all } v \in \mathcal{X}. \tag{2.84}$$

One may then denote $\nabla f : \mathcal{X} \to \mathcal{V}$ to be the function that assigns the vector $\nabla f(x)$ to each $x \in \mathcal{X}$.

[3] For our purposes it suffices to consider real inner product spaces.

Riemannian gradient

Observe that the gradient is obtained by taking the dual of the derivative with respect to the *Euclidean* inner product. The *Riemannian gradient* of a function $f : \mathcal{M} \to \mathbb{R}$ defined on a Riemannian manifold $(\mathcal{M}, \mathfrak{g})$ is the tangent vector $\overline{\nabla} f(x) \in \mathrm{T}_x\mathcal{M}$ defined as

$$\mathrm{D}f(x)(v) = \left\langle \overline{\nabla} f(x), v \right\rangle_x \tag{2.85}$$

for each $v \in \mathrm{T}_x\mathcal{M}$. Here the tangent space inner product $\langle \cdot, \cdot \rangle_x \equiv \mathfrak{g}_x(\cdot, \cdot)$ is defined via the Riemannian metric tensor $\mathfrak{g}$.

Let $\mathfrak{g}$ and $\mathfrak{h}$ be two different Riemannian metrics on a smooth manifold $\mathcal{M}$. Due to the metric independence of the total derivative, we can write the following for a function $f : \mathcal{M} \to \mathbb{R}$ differentiable at $x \in \mathcal{M}$:

$$\mathrm{D}f(x)(v) = \langle \nabla_{\mathfrak{g}} f(x), v \rangle_x^{\mathfrak{g}} = \langle \nabla_{\mathfrak{h}} f(x), v \rangle_x^{\mathfrak{h}}. \tag{2.86}$$

Consider the manifold $\mathbb{P}_d^+$. Choose one of the metrics to be the Hilbert–Schmidt metric and the other to be the Bures metric. We then have

$$\mathrm{D}f(P)(V) = \langle \nabla f(P), V \rangle = \langle \overline{\nabla} f(P), V \rangle_P^{\mathrm{Bu}}, \tag{2.87}$$

where we have used ∇f to denote the (Euclidean) gradient and $\overline{\nabla} f$ to denote the Bures gradient. This relation allows one to explicitly compute the Bures gradient, as follows (see also [Han+21; MMP18]).

Proposition 2.4.1. *Let $f : \mathbb{P}_d^+ \to \mathbb{R}$ be a differentiable function, ∇f denote the Euclidean gradient, and $\overline{\nabla} f$ denote the Bures gradient. Then we have*

$$\nabla f(P) = \frac{1}{2} \mathcal{L}_P \left(\overline{\nabla} f(P) \right) \tag{2.88}$$

and equivalently

$$\overline{\nabla} f(P) = 2\mathcal{L}_P^{-1}(\nabla f(P)) = 2[\nabla f(P) P + P \nabla f(P)]. \tag{2.89}$$

Proof. We begin with Eq. (2.87), which states that the Bures gradient and Euclidean gradient must be related via the relation $\langle \nabla f(P), V \rangle = \langle \overline{\nabla} f(P), V \rangle_P$ for all $V \in \mathrm{T}_P \mathbb{P}_d^+ \cong \mathbb{H}_d$. By Eq. (2.76), this is equivalent to the requirement

$$\langle \overline{\nabla} f(P), V \rangle_P := \frac{1}{2} \operatorname{Tr}[\mathcal{L}_P(\overline{\nabla} f(P)) V] = \operatorname{Tr}[\nabla f(P) V] =: \langle \nabla f(P), V \rangle, \tag{2.90}$$

for all Hermitian V, which implies $\nabla f(P) = \frac{1}{2} \mathcal{L}_P(\overline{\nabla} f(P))$ as claimed. Inverting

the equation, we get the second relation. □

2.4.6 Riemannian exponential and logarithmic maps

We now discuss the *exponential* and *logarithmic* maps on a Riemannian manifold. Informally, the exponential map maps tangent vectors to points on the manifold, while its inverse, the logarithmic map, maps points on the manifold to the tangent space. We discuss these concepts in detail.

Riemannian exponential map

Recall that geodesics can be (locally) uniquely defined by the initial position $x \in \mathcal{M}$ and an initial *velocity* (which is a tangent vector) $v \in \mathrm{T}_p\mathcal{M}$. Such a geodesic is denoted as

$$\gamma_{x,v} : [0, 1] \to \mathcal{M}, \tag{2.91}$$

where the input parameter $t \in [0, 1]$ can be thought of as the *time*. The Riemannian exponential map *at* x maps the pair v to $\gamma_{x,v}(1)$:

Definition 2.4.2 (Riemannian exponential)**.** *Let* $x \in \mathcal{M}$*. The* Riemannian exponential map *at* x *is defined as*

$$\mathrm{Exp}_x[v] := \gamma_{x,v}(1), \tag{2.92}$$

where $v \in \mathrm{dom}(\mathrm{Exp}_x) \subseteq \mathrm{T}_x\mathcal{M}$*.*

Informally, the exponential map takes in a tangent vector $v \in \mathrm{T}_x\mathcal{M}$ and maps it to the point where one would *reach* if one *starts* at $x \in \mathcal{M}$ and *travels* for a *unit time* along the direction specified by v.

We strengthen our intuition by studying the Euclidean exponential map. Recall that for a Euclidean space $\mathcal{X} = \mathbb{R}^d$, the tangent space at every point is isomorphic to $\mathcal{X}$. Thus one can freely *add* a tangent vector to a point on the manifold as both are elements of the vector space $\mathcal{X}$. The Euclidean exponential map is defined as

$$\mathrm{Exp}_x^{\mathrm{Euc}}[v] := x + v. \tag{2.93}$$

Indeed the form is familiar to us from elementary vector calculus. We now look at the definition of the exponential map for the Bures–Wasserstein manifold [MMP18; Han+21].

Definition 2.4.3 (Bures exp. map)**.** *The exponential map at P for the Bures manifold is defined as*

$$\mathrm{Exp}_P^{\mathrm{Bu}}[V] := P \star [\mathbb{I} + \mathcal{L}_P(V)]^2 = [\mathbb{I} + \mathcal{L}_P(V)] \cdot P \cdot [\mathbb{I} + \mathcal{L}_P(V)], \tag{2.94}$$

where $P \in \mathbb{P}_d^+$ and $V \in \mathrm{dom}(\mathrm{Exp}_P^{\mathrm{Bu}}) \subset \mathrm{T}_P\mathbb{P}_d^+ \cong \mathbb{H}_d$.

For any $P \in \mathbb{P}_d^+$, we have $\mathrm{dom}(\mathrm{Exp}_P) = \{V : \mathbb{I} + \mathcal{L}_P(V) \in \mathbb{P}_d^+\} \subset \mathrm{T}_P\mathbb{P}_d^+ \cong \mathbb{H}_d$. See [MMP18, Prop. 9] for details. As we will see in Section 2.4.7, the exponential map plays a crucial role in gradient descent over Riemannian manifolds.

Riemannian logarithmic map

The inverse of the Riemannian exponential map, as the name indicates, is the Riemannian logarithmic map. The two maps are diffeomorphic to each other. The logarithmic map at x takes in another point $y \in \mathcal{M}$ and maps it to a tangent vector in $\mathrm{T}_x\mathcal{M}$. The Euclidean logarithmic map is given as

$$\mathrm{Log}_x[y] := y - x. \tag{2.95}$$

Recall that this is familiar from vector calculus as the vector whose tip points at y with its tail at x. We now look at the Riemannian log map for the Bures manifold [MMP18, Prop. 9].

Definition 2.4.4 (Bures log. map)**.** *Let $P \in \mathbb{P}_d^+$. The logarithmic map at P, $\mathrm{Log}_P : \mathbb{P}_d^+ \to \mathrm{T}_P\mathbb{P}_d^+$, for the Bures manifold is defined as*

$$\begin{aligned} \mathrm{Log}_P^{\mathrm{Bu}}[Q] &:= \mathcal{L}_P^{-1}(P^{-1}\#Q - \mathbb{I}) \\ &= P \cdot [P^{-1}\#Q - \mathbb{I}] + [P^{-1}\#Q - \mathbb{I}] \cdot P \\ &= \sqrt{PQ} + \sqrt{QP} - 2P, \end{aligned} \tag{2.96}$$

for any $Q \in \mathbb{P}_d^+$, and we used the fact that $\mathcal{L}_P^{-1}(X) = XP + PX$.

It is easy to see that for any $Q \in \mathbb{P}_d^+$, we have

$$\begin{aligned} \mathrm{Exp}_P^{\mathrm{Bu}}\left[\mathrm{Log}_P^{\mathrm{Bu}}[Q]\right] &= P \star [\mathbb{I} + \mathcal{L}_P(\mathcal{L}_P^{-1}(P^{-1}\#Q - \mathbb{I}))]^2 \\ &= P \star (P^{-1}\#Q)^2 = Q, \end{aligned} \tag{2.97}$$

and for any $V \in \mathrm{dom}(\mathrm{Exp}_P^{\mathrm{Bu}})$ (which implies $\mathbb{I} + \mathcal{L}_P(V) \in \mathbb{P}_d^+$) we have

$$\begin{aligned} \mathrm{Log}_P^{\mathrm{Bu}}\left[\mathrm{Exp}_P^{\mathrm{Bu}}[V]\right] &= \mathcal{L}_P^{-1}(P^{-1}\#\mathrm{Exp}_P^{\mathrm{Bu}}[V] - \mathbb{I}) \\ &= \mathcal{L}_P^{-1}((\mathbb{I} + \mathcal{L}_P(V)) - \mathbb{I}) = V, \end{aligned} \tag{2.98}$$

where we used the fact that

$$\mathrm{Exp}_P^{\mathrm{Bu}}[V] := P \star [\mathbb{I} + \mathcal{L}_P(V)]^2 \quad \Longrightarrow \quad \mathbb{I} + \mathcal{L}_P(V) = P^{-1}\#\mathrm{Exp}_P^{\mathrm{Bu}}[V], \tag{2.99}$$

which follows from the uniqueness of the positive definite solution to the matrix Riccati equation (Proposition 2.1.3).

Having summarized the Riemannian exponential and logarithmic maps, we now look at them in action in *Riemannian Gradient Descent*.

2.4.7 Riemannian gradient descent

In this section, we briefly summarize *gradient descent* (GD), the workhorse of modern optimization algorithms and then discuss *Riemannian gradient descent* (RGD)—the generalization of GD to non-Euclidean manifolds.

Let us first recall gradient descent informally. Let f be a real-valued differentiable function. A point $x \in \mathrm{dom}(f)$ is said to be *stationary* if the derivative of f at x is zero. Our goal is to find such a stationary point. Gradient descent is an iterative method that moves in the direction of *steepest descent*, as indicated by the negative gradient of the function. With a suitably chosen step size and after a sufficient number of iterations, the algorithm yields (a sufficiently good approximation of) a stationary point.

The process of choosing the next point is governed by the *update rule*. Gradient descent starts from an *initial point*—often chosen randomly or based on prior knowledge—and generates a sequence of points that (under suitable conditions) converges to a stationary point of the function. For convex functions, every stationary point is also a global minimum.

Formally, let $\mathcal{X} = \mathbb{C}^d$ and $f : \mathcal{X} \to \mathbb{R}$ be a differentiable function. The update rule of gradient descent for f is defined as follows.

Definition 2.4.5 (GD update rule)**.** *Let $f : \mathcal{X} \to \mathbb{R}$ be differentiable and let x_t be the current point. Then the update rule is*

$$x_{t+1} := x_t - \eta_t \nabla f(x_t), \tag{2.100}$$

where $\eta_t \in \mathbb{R}_+$ is the step size *or learning rate at round t and $\nabla f : \mathcal{X} \to \mathcal{X}$ is the gradient of f.*

Let us analyze the update rule. Recall that, formally, $\nabla f(x_t)$ belongs to the tangent space of x_t:

$$\nabla f(x_t) \in \mathrm{T}_{x_t}\mathcal{X} \cong \mathcal{X}. \tag{2.101}$$

Thus, at round t, we are taking a step of size η_t in the direction dictated by the negative gradient. Indeed the above expression can be equivalently written as

$$x_{t+1} = \mathrm{Exp}_{x_t}^{\mathrm{Euc}}[-\eta_t \nabla f(x_t)]. \tag{2.102}$$

where $\mathrm{Exp}_x^{\mathrm{Euc}} : \mathrm{T}_x\mathcal{X} \to \mathcal{X}$ is the Euclidean exponential map at $x \in \mathcal{X}$. The update rule for gradient descent on a general Riemannian manifold is obtained by the appropriate generalization of Eq. (2.102).

Definition 2.4.6 (RGD update rule)**.** *Let $(\mathcal{M}, \mathfrak{g})$ be a Riemannian manifold, $f : \mathcal{M} \to \mathbb{R}$ be differentiable and let x_t be the current point. The update rule for RGD is*

$$x_{t+1} := \mathrm{Exp}_{x_t}^{\mathfrak{g}}\left[-\eta_t \overline{\nabla} f(x_t)\right], \tag{2.103}$$

where $\eta_t \in \mathbb{R}_+$ is the step size at round t, $\mathrm{Exp}_{x_t}^{\mathfrak{g}}$ is the Riemannian exponential map at x_t, and $\overline{\nabla} f(x_t) \in \mathrm{T}_{x_t}\mathcal{M}$ is the Riemannian gradient of f.

In this thesis, we are interested in Riemannian gradient descent over the Bures manifold. Thus we shall now summarize the relevant formulae pertaining to this manifold.

Proposition 2.4.7 (RGD update rule for Bures manifold.)**.** *Let* $(\mathcal{M}, \mathfrak{g}^{\mathrm{Bu}})$ *be the Bures manifold and* $f : \mathcal{M} \to \mathbb{R}$ *be a differentiable function. For a step size* $\eta_t \in \mathbb{R}_+$*, the update rule for Riemannian gradient descent over the Bures manifold is*

$$\begin{aligned} P_{t+1} := \mathrm{Exp}^{\mathrm{Bu}}_{P_t}\left[-\eta_t \overline{\nabla} f(P_t)\right] &= [\mathbb{I} - 2\eta_t \nabla f(P_t)] \cdot P_t \cdot [\mathbb{I} - 2\eta_t \nabla f(P_t)] \\ &\equiv P_t \star [\mathbb{I} - 2\eta_t \nabla f(P_t)]^2, \end{aligned} \tag{2.104}$$

provided $\eta_t \overline{\nabla} f(P_t) \in \mathrm{dom}(\mathrm{Exp}^{\mathrm{Bu}}_{P_t})$.

Proof. We aim to show that $\mathrm{Exp}^{\mathrm{Bu}}_{P_t}\left[-\eta_t \overline{\nabla} f(P_t)\right] = P_t \star [\mathbb{I} - 2\eta_t \nabla f(P_t)]^2$. To this end, first recall the form of the Bures exponential map as defined in Eq. (2.94):

$$\mathrm{Exp}^{\mathrm{Bu}}_P(V) = P \star [\mathbb{I} + \mathcal{L}_P(V)]^2 \tag{2.105}$$

for a tangent vector $V \in \mathrm{T}_P\mathcal{M}$. Choose $P = P_t$ to be the current iterate and $V = -\eta_t \overline{\nabla} f(P_t)$ to be the (scaled) Bures gradient of f at P_t. By Eq. (2.89), we have $\overline{\nabla} f(P_t) = 2\mathcal{L}_P^{-1}(\nabla f(P_t))$. Substituting, we get

$$\begin{aligned} \mathrm{Exp}^{\mathrm{Bu}}_{P_t}[-\eta_t \overline{\nabla} f(P_t)] &= P_t \star [\mathbb{I} + \mathcal{L}_{P_t}(-\eta_t \overline{\nabla} f(P_t))]^2 \\ &= P_t \star [\mathbb{I} - \eta_t \mathcal{L}_{P_t}(2\mathcal{L}_{P_t}^{-1} \nabla f(P_t))]^2 \\ &= P_t \star [\mathbb{I} - 2\eta_t \nabla f(P_t)]^2, \end{aligned} \tag{2.106}$$

where we have used the fact that $\mathcal{L}_P$ is a linear operator for any P. This concludes the proof. □

Chapter 3

Averaging fidelities

In this chapter, we discuss the first set of results with this thesis. This problem concerns finding the state that maximizes average Uhlmann fidelity over a given collection of states. We first present the problem statement. This is followed by a detailed statement of results, and then we look at how the results are derived. An illustration of the summary of the results is given in Figure 3.1.

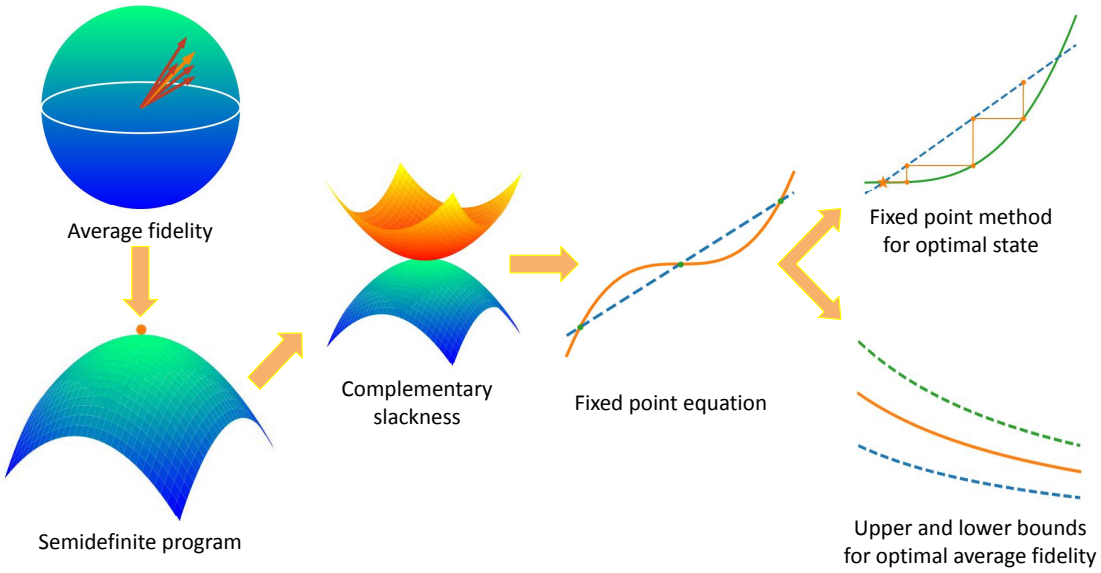

Figure 3.1: The problem of finding the state that maximizes the average fidelity is framed as a semidefinite program that exhibits complementary slackness relations. These relations lead to a fixed-point equation satisfied by the optimal state from which we construct a fixed-point iteration algorithm for the optimal state and heuristic near-optimal estimator which is optimal when all the states commute. Finally, we present upper bounds for optimal average fidelity achieved by any state.

3.1 Problem statement

We begin by stating the main problem of interest. Let $\mathcal{R} := \{\rho_1, \ldots, \rho_n\} \subset \mathbb{D}_d$ be a collection of density matrices and let $w_1, \ldots, w_n \geq 0$ with $\sum_{i=1}^n w_i = 1$ be associated probability weights, or equivalently, $w := (w_1, \ldots, w_n) \in \Delta_n$ be a probability vector. We shall denote this distribution as $(\mathcal{R}, w)$. We are then interested in finding the density matrix $\sigma_\sharp \in \mathbb{D}_d$ which maximizes *average Uhlmann fidelity* over $\mathcal{R}$:

$$\sigma_\sharp := \operatorname*{argmax}_{\sigma \in \mathbb{D}_d} f(\sigma) := \operatorname*{argmax}_{\sigma \in \mathbb{D}_d} \sum_{i=1}^n w_i \, \mathrm{F}^{\mathrm{U}}(\rho_i, \sigma), \tag{3.1}$$

where $\mathrm{F}^{\mathrm{U}}(P, Q) := \|P^{\frac{1}{2}} Q^{\frac{1}{2}}\|_1$ denotes the Uhlmann fidelity between $P, Q \in \mathbb{P}_d$ and $f(\sigma)$ denotes the *average Uhlmann fidelity* of σ over the distribution $\mathcal{R}$. Because we almost exclusively deal with the Uhlmann fidelity in this chapter, we will suppress the 'Uhlmann' part and refer to the function as just *fidelity* and will drop the superscript: $\mathrm{F}^{\mathrm{U}} \equiv \mathrm{F}$. If all the states in the ensemble are full-rank, by the strict concavity of fidelity (see Proposition 2.3.8), the average fidelity function is also strictly concave and thus admits a unique maximizer.

As noted earlier, we will study the situation where $\mathcal{R}$ is a collection of density matrices, the weights w form a probability vector, and the optimal state $\sigma_\sharp$ is a density matrix as this is typically the scenario of interest in quantum information. We note that most of the results in this chapter, especially the semidefinite program, fixed-point equation, and algorithms, extend to the case where the following relaxations are made.

1. The ensemble $\mathcal{R}$ is allowed to have positive semidefinite matrices as elements instead of just density matrices. The weight vector w is allowed to be an arbitrary positive vector: $w \in \mathbb{R}^n_+$.

2. The optimal state $\sigma_\sharp$ is allowed to be of arbitrary (but fixed) trace, instead of unit trace.

3.2 Informal statement of results

We now informally state the results, after which we will study each result in detail. We first present a semidefinite program (SDP) which solves Problem 3.1 in Section 3.3. The SDP exhibits strong duality and, when all the states in the ensemble are full rank, it exhibits complementary slackness. However, numerically solving SDPs can quickly grow intractable, especially since the SDP involves optimizing over matrices of dimension $(n+1)d$. This difficulty can be circumvented

by the following observation. Complementary slackness of the SDP implies a fixed-point equation that is uniquely satisfied by the optimal state:[1]

$$\sigma_\sharp = \frac{1}{f(\sigma_\sharp)} \sum_{i=1}^{n} w_i \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}}. \tag{3.2}$$

We use this to develop two fixed-point iteration algorithms that converge to the optimal state for any full-rank starting point. The two algorithms are defined by the two fixed-point iteration maps Λ and Ω of the form

$$\begin{aligned} \Lambda(\sigma) &:= \Pi \left[\sum_{i=1}^{n} w_i \sqrt{\sigma^{\frac{1}{2}} \rho_i \sigma^{\frac{1}{2}}} \right], \\ \Omega(\sigma) &:= \Pi \left[\sigma^{-\frac{1}{2}} \left(\sum_{i=1}^{n} w_i \sqrt{\sigma^{\frac{1}{2}} \rho_i \sigma^{\frac{1}{2}}} \right)^2 \sigma^{-\frac{1}{2}} \right], \end{aligned} \tag{3.3}$$

where $\Pi[A] := A/\operatorname{Tr}[A]$ is used to normalize (nonzero) positive semidefinite matrices to density matrices. The notation Π is indicative of the fact that the operation $A \mapsto A/\operatorname{Tr}[A]$ is bona-fide *projection* with respect to Bures distance (or equivalently fidelity). To elaborate, for a given $P \in \mathbb{P}_d$ the minimization problem

$$\begin{aligned} \text{minimize}: \quad & \mathrm{B}(P, \sigma) \\ \text{subject to}: \quad & \sigma \in \mathbb{D}_d, \end{aligned} \tag{3.4}$$

is is solved at at $\sigma = P/\operatorname{Tr}[P]$. Here $\mathrm{B}(P, Q) := \operatorname{Tr}[P + Q] - 2\,\mathrm{F}(P, Q)$ is the squared Bures distance between $P, Q \in \mathbb{P}_d$. Thus the projection of an arbitrary positive semidefinite matrix to the set of density matrices is given by trace-normalization. This result (see Corollary 5.2.6) and its generalizations are discussed in Chapter 5. A comparison of the numerical performances of these algorithms (standard numerical solvers for SDP and fixed-point) is presented in Figure 3.2.

The sequence of states $\{\Lambda^k(\sigma)\}_{k=0}^{\infty}$ and $\{\Omega^k(\sigma)\}_{k=0}^{\infty}$ is seen to converge to the optimal state $\sigma_\sharp$ for any full-rank initial state $\sigma \in \mathbb{D}_d$. Here we define $\Lambda^0(\sigma) := \sigma$ and $\Lambda^k(\sigma) := \Lambda\left(\Lambda^{k-1}(\sigma)\right)$ for all integers $k \geq 1$. Similar notation is followed for the map Ω. We observe that the fixed-point method works when all the states in the ensemble are full-rank. If one is interested in optimizing over rank-deficient states, first depolarize the states by a small factor to obtain full-rank states and then use the fixed-point algorithms. Alternatively, we may simply use the SDP

[1] Note that there is a different (and arguably shorter) way to derive the above fixed-point equation by directly applying the *Karush-Kuhn-Tucker* (KKT) conditions [BV04] to the optimization problem $\max_{\sigma \geq 0, \operatorname{Tr}[\sigma]=1} f(\sigma)$. However, one has to separately show that the optimum is achieved at a full-rank state.

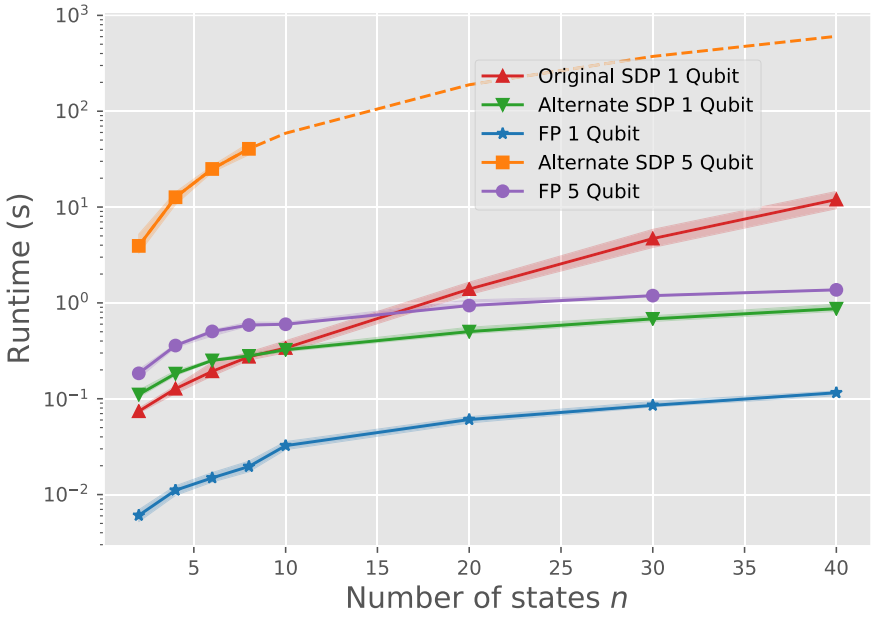

Figure 3.2: Performance comparison of SDP and FP algorithm for finding the state maximizing average fidelity. Runtime plotted as a function of number of states for original SDP Eq. (3.10), alternate SDP Eq. (B.1), and fixed-point iteration algorithm Eq. (3.37) for two different dimensions. For 1-qubit states ($d = 2$), we plot the runtime of all three methods. For 5-qubit states ($d = 32$), we omit the original SDP and plot the alternate SDP only up to $n = 8$ due to computational intractability. Each marker is the median of 50 iterations and shaded regions correspond to interquartile regions. The dashed line (for alternate SDP 5-qubit) represents extrapolation from numerical data.

which yields the optimum even when the states are rank-deficient.

If all the states in the ensemble $\mathcal{R}$ commute pairwise—i.e., $[\rho_i, \rho_j] = 0$ for any pair $i, j \in [n]$—then a simple analytic expression exists for the optimal state $\sigma_\sharp$:

$$\sigma_\sharp = \sigma' := \Pi \left[\left(\sum_{i=1}^{n} w_i \rho_i^{\frac{1}{2}} \right)^2 \right]. \tag{3.5}$$

Essentially, the square root of the optimal state in this scenario is equal to the weighted average of the square roots of the matrices supporting the distribution. The state σ', called the *Commuting estimator*, can also serve as an easy-to-compute near-optimal heuristic approximation even in cases where the ensemble does not commute. One can show that [FB16; BJL19], σ' is the optimal state for arbitrary (not necessarily commuting) distributions if we replace the Uhlmann fidelity with Holevo fidelity $\mathrm{F}^{\mathrm{H}}(P, Q) = \mathrm{Tr}\left[P^{\frac{1}{2}} Q^{\frac{1}{2}}\right]$ in the optimization problem of interest to get

$$f'(\sigma) := \sum_{i=1}^{n} w_i \, \mathrm{F}^{\mathrm{H}}(\rho_i, \sigma) = \sum_{i=1}^{n} w_i \, \mathrm{Tr}\left[\rho_i^{\frac{1}{2}} \sigma^{\frac{1}{2}}\right]. \tag{3.6}$$

Numerically we see σ' to be a good approximation to $\sigma_\sharp$. In particular, it is better than the mean estimator (defined as $\sigma_{\mathrm{M}} := \sum_{i=1}^{n} w_i \rho_i$). The superiority of the Commuting estimator over the mean estimator is seen especially when the states are close to each other, which is the case in Bayesian tomography–a proposed application for the results of this chapter.

We also present analytic upper bounds for the maximum value of the average fidelity $f(\sigma)$ for any state $\sigma \in \mathrm{D}(\mathcal{H})$:

$$f(\sigma) \leq \sqrt{\sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)} \leq \sqrt{f(\sigma_{\mathrm{M}})}. \tag{3.7}$$

We call these bounds *Product bound* and *Average bound* respectively.

Note that the average fidelity $f(\sigma')$ of the Commuting estimator σ' is a lower bound on the maximum average fidelity. Bringing the Product bound into the picture, we have lower and upper bounds for optimal average fidelity $f(\sigma_\sharp)$:

$$f(\sigma') \leq f(\sigma_\sharp) \leq \sqrt{\sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)}. \tag{3.8}$$

These bounds are tight as both the upper and lower bounds coincide with optimal average fidelity $f(\sigma)$, i.e., the inequalities become equalities when all the states in the ensemble commute pairwise:

$$[\rho_i, \rho_j] = 0 \text{ for all } i, j \in [n] \implies f(\sigma') = f(\sigma_\sharp) = \sqrt{\sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)}. \tag{3.9}$$

3.3 SDP for optimal average fidelity

The semidefinite program for optimal average fidelity can be seen as a generalization of Watrous's SDP for fidelity [Wat18]. We begin with formally defining the SDP whose primal optimum is the optimal average fidelity Eq. (3.1). Moreover, this SDP also provides the optimal state.

Definition 3.3.1 (SDP for optimal average fidelity)**.** *Let $\mathcal{H} = \mathbb{C}^d$ and $\mathcal{R} = \{\rho_1, \ldots, \rho_n\} \subset \mathbb{D}_{\mathcal{H}}$ be a collection of quantum states and $w \in \Delta^n$ be a probability vector. Let $\mathcal{X} = \bigoplus_{i=1}^{n+1} \mathcal{H}$ and $\mathcal{Y} = \bigoplus_{i=1}^{n} \mathcal{H} \oplus \mathbb{C}$ be Hilbert spaces with dimensions $(n+1)d$ and $nd+1$ respectively.*

Define the semidefinite program (Φ, A, B) as

$$A = \frac{1}{2}\begin{pmatrix} 0 & \cdots & 0 & w_1\mathbb{I}_{\mathcal{H}} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & w_n\mathbb{I}_{\mathcal{H}} \\ w_1\mathbb{I}_{\mathcal{H}} & \cdots & w_n\mathbb{I}_{\mathcal{H}} & 0 \end{pmatrix} \in \mathbb{H}_{\mathcal{X}}, \quad B = \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \rho_n & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \in \mathbb{H}_{\mathcal{Y}}, \tag{3.10}$$

and the Hermitian preserving map $\Phi : \mathbb{M}_{\mathcal{X}} \to \mathbb{M}_{\mathcal{Y}}$ which acts as

$$\Phi\begin{pmatrix} P_1 & \cdot & \cdots & \cdot \\ \cdot & \ddots & & \vdots \\ \vdots & & P_n & \cdot \\ \cdot & \cdots & \cdot & Q \end{pmatrix} = \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & P_n & 0 \\ 0 & \cdots & 0 & \mathrm{Tr}(Q) \end{pmatrix}. \tag{3.11}$$

The map Φ acts like the identity on the $d \times d$ principal block diagonal submatrices, except the last $d \times d$ submatrix which is traced. Every other element is zeroed out. That is

$$\Phi\left(\sum_{i,j=1}^{n+1} |i\rangle\langle j| \otimes R_{ij}\right) = \left(\sum_{i=1}^{n} |i\rangle\langle i| \otimes R_{ii}\right) \oplus \mathrm{Tr}[R_{n+1,n+1}], \tag{3.12}$$

for $R_{i,j} \in \mathbb{M}_{\mathcal{H}}$. The corresponding adjoint map $\Phi^* : \mathbb{M}_{\mathcal{Y}} \to \mathbb{M}_{\mathcal{X}}$ has the action

$$\Phi^*\begin{pmatrix} \rho_1 & \cdot & \cdots & \cdot \\ \cdot & \ddots & & \vdots \\ \vdots & & \rho_n & \cdot \\ \cdot & \cdots & \cdot & q \end{pmatrix} = \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \rho_n & 0 \\ 0 & \cdots & 0 & q\mathbb{I}_{\mathcal{H}} \end{pmatrix}, \tag{3.13}$$

where $q \in \mathbb{C}$. We now discuss how the optimal average fidelity $\max_{\sigma \in \mathbb{D}_d} f(\sigma)$ is an upper bound for the primal objective function $\langle A, \mathrm{X}\rangle$ of the above SDP.

Lemma 3.3.2. *The primal objective function of the semidefinite program (Φ, A, B) from Definition 3.3.1 is bounded above by the optimal average fidelity* $\max_{\sigma \in \mathbb{D}_d} f(\sigma)$.

Proof. Under the constraint $\Phi(\mathrm{X}) = B$, any primal feasible $\mathrm{X} \in \mathcal{A}$ must have the

form

$$\mathrm{X} = \begin{pmatrix} \rho_1 & R_{12} & \cdots & R_{1n} & X_1 \\ R_{12}^* & \rho_2 & \cdots & R_{2n} & X_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ R_{1n}^* & R_{2n}^* & \cdots & \rho_n & X_n \\ X_1^* & X_2^* & \cdots & X_n^* & \sigma \end{pmatrix} \geq 0. \tag{3.14}$$

Here, each of the block submatrices ρ_i, σ, X_i, and $R_{ij} \in \mathbb{M}_d$ are square matrices, with ρ_i and σ being positive semidefinite as well, for all $i, j \in [n]$. Let us briefly note what these square matrices are.

The collection $\{\rho_i\}$ are the fixed density matrices in the ensemble $\mathcal{R}$ and $\sigma \in \mathbb{D}_{\mathcal{H}}$ is the actual objective matrix which is varied over the set of all d-dimensional quantum states $\mathbb{D}_d$. The specific form of R_{ij}s will be discussed later and the form of the matrices X_is are discussed next.

The positivity of X necessarily implies the positivity of each principal submatrix. In particular,

$$M_i := \begin{pmatrix} \rho_i & X_i \\ X_i^* & \sigma \end{pmatrix} \geq 0 \quad \text{for all } i = 1, \ldots, n. \tag{3.15}$$

By [Wat18, Lemma 3.18], we have $M_i \geq 0$ if and only if $X_i = \rho_i^{\frac{1}{2}} U_i \sigma^{\frac{1}{2}}$ for some U_i with $\|U_i\| \leq 1$. Hence, by the variational property of the trace norm (see Proposition 2.1.1), we have that $M_i \geq 0$ necessarily implies $\mathrm{Re}(\mathrm{Tr}(X_i)) \leq \mathrm{F}(\rho_i, \sigma)$. Therefore for any primal feasible point X, we have the following chain of inequalities involving the objective function $\langle A, \mathrm{X} \rangle$:

$$\langle A, \mathrm{X} \rangle = \sum_{i=1}^{n} w_i \,\mathrm{Re}(\mathrm{Tr}(X_i)) \leq \sum_{i=1}^{n} w_i \,\mathrm{F}(\rho_i, \sigma) \leq \max_{\sigma \in \mathbb{D}_d} f(\sigma) =: f(\sigma_\sharp), \tag{3.16}$$

where $\sigma_\sharp$ is the optimal state. That is the value of the objective function $\langle A, \mathrm{X} \rangle$ is bounded above by the optimal average fidelity. □

Essentially, the maximization occurs at two levels: at the first level, each X_i is varied to maximize the real part of its trace, constrained by the value σ (along with each fixed ρ_i). The degree of freedom we allow σ—to vary all over $\mathbb{D}_d$—defines the second level of maximization.

We now prove, in two different ways, that the inequality Eq. (3.16) is saturated. The first proof, which makes use of the form of X and the fact that we are optimizing over the closed and bounded set of density matrices, culminates in the following theorem.

Theorem 3.3.3. *For any ensemble $\mathcal{R}$, all the inequalities in Eq. (3.16) are saturated by the semidefinite program (Φ, A, B).*

Proof. We first establish that for any state $\sigma \in \mathbb{D}_d$, there exists a primal feasible point $\mathrm{X}(\sigma) \geq 0$ such that $\langle A, \mathrm{X}(\sigma)\rangle = f(\sigma)$. To see this, fix $\sigma \in \mathbb{D}_d$ arbitrary and consider the $(n+1)d \times d$ matrix $Z \equiv Z(\sigma)$ of the form

$$Z = \begin{pmatrix} \rho_1^{\frac{1}{2}} U_1 \\ \rho_2^{\frac{1}{2}} U_2 \\ \vdots \\ \rho_n^{\frac{1}{2}} U_n \\ \sigma^{\frac{1}{2}} \end{pmatrix}. \tag{3.17}$$

Here, each $U_i = \mathrm{Pol}\left(\rho_i^{\frac{1}{2}} \sigma^{\frac{1}{2}}\right)$ is the *optimal* unitary such that $\mathrm{Re}\,\mathrm{Tr}\left[\rho_i^{\frac{1}{2}} U_i \sigma^{\frac{1}{2}}\right] = \mathrm{Tr}\left[\rho_i^{\frac{1}{2}} U_i \sigma^{\frac{1}{2}}\right] = \mathrm{F}(\rho_i, \sigma)$. Now consider

$$\mathrm{X}(\sigma) := ZZ^* = \begin{pmatrix} \rho_1 & \rho_1^{\frac{1}{2}} U_1 U_2^* \rho_2^{\frac{1}{2}} & \cdots & \rho_1^{\frac{1}{2}} U_1 U_n^* \rho_n^{\frac{1}{2}} & \rho_1^{\frac{1}{2}} U_1 \sigma^{\frac{1}{2}} \\ \rho_2^{\frac{1}{2}} U_2 U_1^* \rho_1^{\frac{1}{2}} & \rho_2 & \cdots & \rho_2^{\frac{1}{2}} U_2 U_n^* \rho_n^{\frac{1}{2}} & \rho_2^{\frac{1}{2}} U_2 \sigma^{\frac{1}{2}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_n^{\frac{1}{2}} U_n U_1^* \rho_1^{\frac{1}{2}} & \rho_n^{\frac{1}{2}} U_n U_2^* \rho_2^{\frac{1}{2}} & \cdots & \rho_n & \rho_n^{\frac{1}{2}} U_n \sigma^{\frac{1}{2}} \\ \sigma^{\frac{1}{2}} U_1^* \rho_1^{\frac{1}{2}} & \sigma^{\frac{1}{2}} U_2^* \rho_2^{\frac{1}{2}} & \cdots & \sigma^{\frac{1}{2}} U_n^* \rho_n^{\frac{1}{2}} & \sigma \end{pmatrix}. \tag{3.18}$$

We see that $\mathrm{X}(\sigma) \geq 0$ for any state σ and it follows that for such a choice of $\mathrm{X}(\sigma)$, we have $\langle A, \mathrm{X}(\sigma)\rangle = f(\sigma)$.

Since $\mathbb{D}_d$ is closed and bounded, the continuous function $f(\sigma)$ attains its maximum on $\mathbb{D}_d$. Therefore there exists an optimal $\sigma_\sharp \in \mathbb{D}_d$ achieving the optimal average fidelity. The optimal primal feasible point $\mathrm{X}(\sigma_\sharp)$ such that $\langle A, \mathrm{X}(\sigma_\sharp)\rangle = f(\sigma_\sharp)$ can be then constructed using Eq. (3.18). □

The second proof makes use of Slater's condition. As we show in Theorem 3.3.4, the SDP (Φ, A, B) always satisfies the first of Slater's conditions, thereby ensuring strong duality and the existence of a primal feasible point that achieves optimal value. Moreover, when Slater's conditions are satisfied—which occurs when all the states in the ensemble are full rank—the complementary slackness relations hold.

We formulate the satisfaction of Slater's conditions by the SDP (Φ, A, B) as a separate theorem which also accounts for complementary slackness. The following theorem shows that the above SDP exhibits strong duality and, if all the states in $\mathcal{R}$ are full rank, complementary slackness. The fixed-point equation,

and thereby the fixed-point algorithm for the optimal state, will then arise from complementary slackness.

Theorem 3.3.4. *The semidefinite program (Φ, A, B) exhibits strong duality. Moreover, complementary slackness holds if all the states in $\mathcal{R}$ are full rank.*

Proof. Recall that the first Slater's condition requires the primal constraint to be satisfied ($\mathcal{A} \neq \varnothing$) and the dual constraint to be strictly satisfied ($\Phi^*(\mathrm{Y}) > A$). Note that for any quantum state $\rho \in \mathbb{D}_d$, the matrix

$$\mathrm{X} = \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \rho_n & 0 \\ 0 & \cdots & 0 & \rho \end{pmatrix} \in \mathbb{P}_{\mathcal{H}}, \tag{3.19}$$

is a primal feasible point as it satisfies $\Phi(\mathrm{X}) = B$, thereby ensuring $\mathcal{A} \neq \varnothing$. For the second part, choose Y to be the identity matrix $\mathrm{Y} := \mathbb{I}_{\mathcal{Y}} = \bigoplus_{i=1}^{n} \mathbb{I}_{\mathcal{H}} \oplus 1$ and observe that

$$\Phi^*(\mathrm{Y}) - A = \begin{pmatrix} \mathbb{I}_{\mathcal{H}} & 0 & \cdots & -\frac{1}{2}w_1\mathbb{I}_{\mathcal{H}} \\ 0 & \ddots & & \vdots \\ \vdots & & \mathbb{I}_{\mathcal{H}} & -\frac{1}{2}w_n\mathbb{I}_{\mathcal{H}} \\ -\frac{1}{2}w_1\mathbb{I}_{\mathcal{H}} & \cdots & -\frac{1}{2}w_n\mathbb{I}_{\mathcal{H}} & \mathbb{I}_{\mathcal{H}} \end{pmatrix} > 0. \tag{3.20}$$

This is by [HJ85, Theorem 6.1.10], which states that if a Hermitian matrix M with strictly positive diagonal entries is strictly diagonally dominant i.e., $|M(i,i)| > \sum_{j \neq i} |M(i,j)|$ for all rows i, then M is positive definite. Since $w_i < 1$ for any $i \in [n]$, we see that $\Phi^*(\mathrm{Y}) - A$ satisfies this criterion and therefore is positive definite.

Of course, there could be the case where all weights except a particular weight w_i are 0. However, in such a case the optimization problem is trivially solved by choosing the optimal state to be $\sigma_\sharp = \rho_i$. We do not consider such a *deterministic* scenario.

Hence, the semidefinite program (Φ, A, B) satisfies the first Slater condition which in turn ensures strong duality. Moreover, this optimum α is always achieved for some $\mathrm{X}_\sharp \in \mathcal{A}$. The second condition, in contrast, need not always hold. One sees that if any state ρ_i is *not* full rank, then there exists no $\mathrm{X} \in \mathbb{P}_{\mathcal{X}}$ satisfying $\Phi(\mathrm{X}) = B$, as B itself is rank deficient and therefore complementary slackness will not hold. However, for the case where all states in the ensemble $\mathcal{R}$ are full rank,

then the second condition is also satisfied which in turn leads to complementary slackness. To show that $\mathcal{B}$ is non-empty, we may use the same Y as above and X as defined above satisfies $\Phi(\mathrm{X}) = B$ while being positive definite for any full rank quantum state ρ.

Therefore, if all the states $\{\rho_i\}$ are full rank, there exist primal and dual optimal operators $\mathrm{X}_\sharp \in \mathcal{A}$ and $\mathrm{Y}_\sharp \in \mathcal{B}$ such that

$$\langle A, \mathrm{X}_\sharp \rangle = \alpha = \beta = \langle B, \mathrm{Y}_\sharp \rangle. \tag{3.21}$$

Such a condition implies complementary slackness between primal and dual optimal points which manifests as

$$\Phi^*(\mathrm{Y}_\sharp)\mathrm{X}_\sharp = A\mathrm{X}_\sharp. \tag{3.22}$$

This concludes the proof. □

The second proof of the saturation of inequality Eq. (3.16) follows from the fact that the semidefinite program (Φ, A, B) exhibits strong duality. We now focus on Eq. (3.22), which allows us to infer insights about the properties of the optimal state $\sigma_\sharp$, including that it satisfies a fixed-point equation. This is formally discussed in the next theorem.

Theorem 3.3.5 (Fixed point equation for optimal state)**.** *If all the states in $\mathcal{R}$ are full rank, the following equation holds:*

$$\sigma_\sharp = \frac{1}{f(\sigma_\sharp)} \sum_{i=1}^{n} w_i \rho_i^{\frac{1}{2}} U_i \sigma_\sharp^{\frac{1}{2}} = \frac{1}{f(\sigma_\sharp)} \sum_{i=1}^{n} w_i \sigma_\sharp^{\frac{1}{2}} \rho_i^{\frac{1}{2}} U_i, = \frac{1}{f(\sigma_\sharp)} \sum_{i=1}^{n} w_i \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}} \tag{3.23}$$

for unitaries $U_i = \mathrm{Pol}\left(\rho_i^{\frac{1}{2}} \sigma_\sharp^{\frac{1}{2}}\right) \in \mathbb{U}_d$ and the optimal state $\sigma_\sharp$. This, in turn, implies

$$\sigma_\sharp = \Pi\left(\sum_{i=1}^{n} w_i \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}}\right), \tag{3.24}$$

where Π is the trace normalization map defined as $\Pi(A) := A/\mathrm{Tr}(A)$

Proof. Let $\mathrm{X}_\sharp \in \mathbb{P}_\mathcal{X}$ and $\mathrm{Y}_\sharp \in \mathbb{H}_\mathcal{Y}$ be the optimal primal and dual points satisfying

$$\langle A, \mathrm{X}_\sharp \rangle = \alpha = \beta = \langle B, \mathrm{Y}_\sharp \rangle = f(\sigma_\sharp). \tag{3.25}$$

Decompose $\mathrm{X}_\sharp$ and $\Phi^*(\mathrm{Y}_\sharp)$ as

$$\mathrm{X}_\sharp = \begin{pmatrix} \rho_1 & R_{12} & \cdots & R_{1n} & X_1 \\ R_{21} & \rho_2 & \cdots & R_{2n} & X_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ R_{n1} & R_{n2} & \cdots & \rho_n & X_n \\ X_1^* & X_2^* & \cdots & X_n^* & \sigma_\sharp \end{pmatrix}, \Phi^*(\mathrm{Y}_\sharp) = \begin{pmatrix} Y_1 & 0 & \cdots & 0 & 0 \\ 0 & Y_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & Y_n & 0 \\ 0 & 0 & \cdots & 0 & z\mathbb{1}_{\mathcal{H}} \end{pmatrix}, \quad (3.26)$$

where $R_{ij} = R_{ji}^*$. By complementary slackness, we have $\Phi^*(\mathrm{Y}_\sharp)\mathrm{X}_\sharp = A\mathrm{X}_\sharp$, which is equivalent to

$$\begin{pmatrix} Y_1\rho_1 & Y_1R_{12} & \cdots & Y_1R_{1n} & Y_1X_1 \\ Y_2R_{21} & Y_2\rho_2 & \cdots & Y_2R_{2n} & Y_2X_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_nR_{n1} & Y_nR_{n2} & \cdots & Y_n\rho_n & Y_nX_n \\ zX_1^* & zX_2^* & \cdots & zX_n^* & z\sigma_\sharp \end{pmatrix} = \frac{1}{2} \begin{pmatrix} w_1X_1^* & w_1X_2^* & \cdots & w_1X_n^* & w_1\sigma_\sharp \\ w_2X_1^* & w_2X_2^* & \cdots & w_2X_n^* & w_2\sigma_\sharp \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_nX_1^* & w_nX_2^* & \cdots & w_nX_n^* & w_n\sigma_\sharp \\ S_1 & S_2 & \cdots & S_n & \bar{X} \end{pmatrix} \quad (3.27)$$

where $S_i := w_i\rho_i + \sum_{j=1, j\neq i}^n w_jR_{ji}$ and $\bar{X} := \sum_{i=1}^n w_iX_i$

Let us now deduce the forms of X_i, Y_i, and R_{ij}. The invertibility of ρ_i and Y_i implies that X_i, R_{ij} and $\sigma_\sharp$ are also full-rank. Putting together Eq. (3.15) and Eq. (3.17), we deduce that

$$X_i = \rho_i^{\frac{1}{2}}U_i\sigma_\sharp^{\frac{1}{2}} = \sqrt{\rho_i\sigma_\sharp} = \rho_i\#\sigma_\sharp^{-1} \cdot \sigma_\sharp. \quad (3.28)$$

where $U_i = \mathrm{Pol}\left(\rho_i^{\frac{1}{2}}\sigma^{\frac{1}{2}}\right)$. See Appendix A.2.1 for the proof regarding the last two equalities. From the form of X_i and the fact that $Y_iX_i = \frac{1}{2}w_i\sigma_\sharp$ we can deduce that $Y_i = \frac{1}{2}w_i\rho_i^{-1}\#\sigma_\sharp$. The relations necessarily implies $R_{ij} = \rho_i^{\frac{1}{2}}U_iU_j^*\rho_j^{\frac{1}{2}}$, which is in agreement with the forms for R_{ij} (as defined in Eq. (3.14)) we got in Eq. (3.18). The last block matrix equality from Eq. (3.27) is of particular importance, which can be rewritten as

$$2z\sigma_\sharp = \sum_{i=1}^n w_iX_i = \sum_{i=1}^n w_i\rho_i^{\frac{1}{2}}U_i\sigma_\sharp^{\frac{1}{2}} = \sum_{i=1}^n w_i\rho_i\#\sigma_\sharp^{-1} \cdot \sigma_\sharp. \quad (3.29)$$

We will later on derive the fixed-point algorithm from this relation. It follows that $\mathrm{Tr}(X_i) = \mathrm{F}(\rho_i, \sigma_\sharp)$, and tracing both sides, we obtain $2z = f(\sigma_\sharp)$, Hence we

have

$$f(\sigma_\sharp)\sigma_\sharp = \sum_{i=1}^{n} w_i \rho_i \# \sigma_\sharp^{-1} \cdot \sigma_\sharp = \sum_{i=1}^{n} w_i \left(\sigma_\sharp^{-1/2} \sqrt{\sigma_\sharp^{1/2} \rho_i \sigma_\sharp^{1/2}} \sigma_\sharp^{-1/2} \right) \cdot \sigma_\sharp \\ = \sum_{i=1}^{n} w_i \sigma_\sharp^{-1/2} \sqrt{\sigma_\sharp^{1/2} \rho_i \sigma_\sharp^{1/2}} \sigma_\sharp^{1/2}, \tag{3.30}$$

where in the second equality we used the definition of matrix geometric mean $A\#B := A^{\frac{1}{2}} \sqrt{A^{-\frac{1}{2}} B A^{-\frac{1}{2}}} A^{\frac{1}{2}}$ (see Proposition 2.1.3 for further details). Left and right multiply $\sigma_\sharp^{\frac{1}{2}}$ and $\sigma_\sharp^{-\frac{1}{2}}$ respectively, and rearrange to get

$$\sigma_\sharp = \frac{1}{f(\sigma_\sharp)} \sum_{i=1}^{n} w_i \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}}. \tag{3.31}$$

Equivalently, one may write

$$\sigma_\sharp = \Pi \left(\sum_{i=1}^{n} w_i \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}} \right) = \Pi \left(\sum_{i=1}^{n} w_i \left| \rho_i^{\frac{1}{2}} \sigma_\sharp^{\frac{1}{2}} \right| \right), \tag{3.32}$$

where $\Pi[A] := A/\operatorname{Tr}[A]$. This concludes the proof. □

We can now use expression Eq. (3.32) to construct an iterative fixed-point algorithm to obtain the optimal state. This expression has the property that after each iteration, we have a convex combination of positive definite operators which is then normalized to a density matrix. Therefore, we never leave the set of density matrices during the iteration process.

3.4 Iterative algorithms for optimal fidelity estimator

One may construct a fixed-point iterative algorithm from Eq. (3.32) as follows.

$$\sigma_{(k)} = \Lambda\left(\sigma_{(k-1)}\right) := \Pi \left(\sum_{i=1}^{n} w_i \sqrt{\sigma_{(k-1)}^{\frac{1}{2}} \rho_i \sigma_{(k-1)}^{\frac{1}{2}}} \right), \tag{3.33}$$

where $\Pi(A) = A/\operatorname{Tr}(A)$. We define the repeated action of Λ recursively as $\Lambda^k(\rho) = \Lambda(\Lambda^{k-1}(\rho))$. This algorithm is numerically seen to converge to the optimal state $\sigma_\sharp$ for all choices of full-rank initial states $\sigma_{(0)}$ and is much more tractable than solving the semidefinite program (Φ, A, B) numerically as we avoid optimizing over complex matrices of dimension $(n+1)d$. Though the fixed-point algorithm is numerically seen to always converge to the optimal state for random

initializations, a well-motivated ansatz is the Commuting estimator σ' Eq. (3.39) which is studied in the next subsection.

We now construct a second fixed-point algorithm based on a simple extension of a fixed-point algorithm developed in [Álv+16]. This algorithm was developed in the context of finding *Wasserstein barycenter* [AC11] over Gaussian probability measures with respect to the 2-Wasserstein distance [Mon81; Kan42; Vil+09; PZ19]. Bhatia, Jain, and Lim [BJL19] provide an excellent study of the connection between the 2-Wasserstein distance between Gaussian probability measures and the Bures distance between positive semidefinite matrices.

We also note that Zemel and Panaretos [ZP19] identified this fixed-point algorithm as a Riemannian gradient descent algorithm (with unit step size) on the Wasserstein manifold. The discussion in Section 3.5 is based on this identification.

Let us return to the construction of the second fixed-point algorithm. Bhatia, Jain, and Lim [BJL19] study the problem of minimizing average squared Bures distance $\mathrm{B}(P, Q) := \mathrm{Tr}(P + Q) - 2\,\mathrm{F}(P, Q)$. That is, given a collection of positive definite matrices $\mathcal{P} := \{P_1, \dots, P_n\} \subset \mathbb{P}_d^+$ and a probability vector $w \in \Delta_n$ over it, they find the solution to the optimization problem

$$\operatorname*{argmin}_{Q \in \mathbb{P}_d} \sum_{i=1}^{n} w_i \mathrm{B}(P_i, Q) = \operatorname*{argmin}_{Q \in \mathbb{P}_d} \sum_{i=1}^{n} w_i \left[\mathrm{Tr}[P_i + Q] - 2\,\mathrm{F}(P_i, Q)\right]. \tag{3.34}$$

Such a point is called the *Bures–Wasserstein barycenter* of the distribution $(w, \mathcal{P})$.

The solution $Q_\sharp$ to this problem satisfies a similar fixed-point equation to ours:

$$Q_\sharp = \sum_{i=1}^{n} w_i \sqrt{Q_\sharp^{\frac{1}{2}} P_i Q_\sharp^{\frac{1}{2}}}. \tag{3.35}$$

The fixed-point algorithm, which provably converges to $Q_\sharp$, is of the form

$$Q_{(k+1)} = \mathrm{K}\left(Q_{(k)}\right) = Q_{(k)}^{-\frac{1}{2}} \left(\sum_{i=1}^{n} w_i \sqrt{Q_{(k)}^{\frac{1}{2}} P_i Q_{(k)}^{\frac{1}{2}}} \right)^2 Q_{(k)}^{-\frac{1}{2}}, \tag{3.36}$$

which reduces to Eq. (3.35) at the fixed point $Q_\sharp$. We can modify this algorithm to obtain a second fixed-point algorithm which we describe formally in the following theorem.

Theorem 3.4.1 (Convergence of Ω fixed-point algorithm)**.** *Let $R = \{\rho_1, \ldots, \rho_n\} \in \mathbb{P}_d^+$ be a collection of full-rank states and $w \in \Delta_n$ be a probability vector. Consider the map $\Omega : \mathbb{P}_d^+ \to \mathbb{D}_d$ of the form*

$$\Omega(\sigma) := \Pi(\mathrm{K}(\sigma)) = \Pi\left(\sigma^{-\frac{1}{2}}\left(\sum_{i=1}^{n} w_i \sqrt{\sigma^{\frac{1}{2}}\rho_i\sigma^{\frac{1}{2}}}\right)^2 \sigma^{-\frac{1}{2}}\right), \tag{3.37}$$

where $\Pi(A) := A/\operatorname{Tr}(A)$. Then the collection of states $\{\Omega^k(\sigma)\}_{k=1}$, where we define $\Omega^{k+1}(\sigma) := \Omega(\Omega^k(\sigma))$, converges to the optimal state $\sigma_\sharp$ for any full-rank initial state $\sigma \in \mathbb{D}_d$.

Proof. The proof is a simple extension of [BJL19, Theorem 11], which proves that the fixed-point algorithm K (Eq. (3.36)) converges to Eq. (3.35). By noting that Π is a continuous function (as it is division by a non-zero positive scalar), it follows that the fixed-point algorithm Ω converges to $\sigma_\sharp$. □

Numerical results indicate that both fixed-point algorithms have comparable performance in terms of total runtime. We also note that [Che+20] establishes a linear rate of convergence for the fixed-point algorithm discussed in [BJL19; Álv+16]. Also see [BRT23] for a proof establishing linear convergence for a generalization of the average fidelity maximization problem. Since the convergence of Ω fixed-point algorithm can be proven theoretically, we prefer it over Λ fixed-point algorithm. In the following sections, when we refer to simply 'fixed-point algorithm', we mean the Ω fixed-point algorithm.

3.4.1 Heuristic approximations of optimal fidelity estimators

If all the states in $\mathcal{R}$ commute pairwise i.e., $[\rho_i, \rho_j] = 0$ for all $i, j \in [n]$, then there exists a simple analytic expression for the optimal state $\sigma_\sharp$. We begin with noting that when all the states commute pairwise, the problem reduces to a *classical* problem, as the problem can be considered in the common eigenbasis where we are now dealing with a collection of probability vectors. Hence we may take that the optimal solution is also a probability vector, or that the optimal state commutes with all $\rho_i \in \mathcal{R}$. We then have

$$f(\sigma_\sharp) \cdot \sigma_\sharp = \sum_{i=1}^{n} w_i \sqrt{\sigma_\sharp^{\frac{1}{2}}\rho_i\sigma_\sharp^{\frac{1}{2}}} = \sum_{i=1}^{n} w_i \rho_i^{\frac{1}{2}}\sigma_\sharp^{\frac{1}{2}} \tag{3.38}$$

Multiply both sides by $\sigma_\sharp^{-\frac{1}{2}}$ and squaring, we obtain

$$\sigma_\sharp = \sigma' := \Pi\left(\left(\sum_{i=1}^{n} w_i \rho_i^{\frac{1}{2}}\right)^2\right), \tag{3.39}$$

which we call the *Commuting estimator* σ'. This expression is consistent with the results of [FB16], which address the problem of finding optimal states for fidelity restricted to commuting states. The expression Eq. (3.39) also serves as a heuristic approximation in the general case. To see this, recall that $f(\sigma_\sharp)\cdot\sigma_\sharp = \sum_{i=1}^{n} w_i \rho_i^{\frac{1}{2}} U_i \sigma_\sharp^{\frac{1}{2}}$. Since $U_j = \exp(iH_j)$ for some Hermitian $H_j \in \mathbb{H}_\mathcal{H}$, we may write

$$f(\sigma_\sharp)\cdot\sigma_\sharp = \sum_{j=1}^{n} w_j \rho_j^{\frac{1}{2}} \left(\mathbb{I}_\mathcal{H} + iH_j + \frac{1}{2}(iH_j)^2 + \cdots\right)\sigma_\sharp^{\frac{1}{2}}. \tag{3.40}$$

Taking a 0th order approximation by ignoring all terms expect $\mathbb{I}_\mathcal{H}$, we obtain σ'. This approximation would make sense when all the states $\rho_i \in \mathcal{R}$ are 'close by', which is the case for Bayesian state estimation which is the proposed application of these results.

3.4.2 Upper and lower bounds on optimal average fidelity

We now present two different upper bounds on the maximum average fidelity achievable by any state for an arbitrary ensemble $(\mathcal{R}, p)$ (that may include rank-deficient states). We call these bounds the *Average bound* and the *Product bound*, respectively. The Average bound states that the square root of average fidelity of the Mean estimator $\sigma_\mathrm{M} = \sum_{i=1}^{n} w_i \rho_i$ bounds the average fidelity obtained by state σ from above:

$$f(\sigma) = \sum_{i=1}^{n} w_i \,\mathrm{F}(\rho_i, \sigma) \leq \sqrt{f(\sigma_\mathrm{M})}. \tag{3.41}$$

The Product bound states that for any state σ,

$$f(\sigma) \leq \sqrt{\sum_{i,j=1}^{n} w_i w_j \,\mathrm{F}(\rho_i, \rho_j)}. \tag{3.42}$$

The Product bound is tighter than the Average bound. These statements are formalized below. We first prove the following lemma which deals with the scenario where all the states in the ensemble are full-rank.

Lemma 3.4.2. *Let $\mathcal{R} = \{\rho_1, \ldots, \rho_n\} \subset \mathbb{D}_d$ be a collection of full-rank states and $w \in \Delta_n$ be a probability vector. Then for any state $\sigma \in \mathbb{D}_d$, it holds that*

$$f(\sigma) \leq \sqrt{\sum_{i=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)}. \tag{3.43}$$

When all the states in the ensemble commute pairwise, the inequality is saturated.

Proof. By Eq. (3.23), we have

$$f(\sigma_\sharp) \cdot \sigma_\sharp = \sum_{i=1}^{n} w_i \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}} = \sum_{i=1}^{n} w_i \sigma_\sharp^{\frac{1}{2}} \rho_i^{\frac{1}{2}} U_i. \tag{3.44}$$

Squaring and then left and right multiplying by $\sigma_\sharp^{-\frac{1}{2}}$, we have

$$\begin{aligned} f(\sigma_\sharp)^2 \cdot \sigma_\sharp &= \sigma_\sharp^{-\frac{1}{2}} \left(\sum_{i,j=1}^{n} w_i w_j \sigma_\sharp^{\frac{1}{2}} \rho_i^{\frac{1}{2}} U_i U_j^* \rho_j^{\frac{1}{2}} \sigma_\sharp^{\frac{1}{2}} \right) \sigma_\sharp^{-\frac{1}{2}} \\ &= \sum_{i,j=1}^{n} w_i w_j \rho_i^{\frac{1}{2}} U_i U_j^* \rho_j^{\frac{1}{2}}. \end{aligned} \tag{3.45}$$

Tracing both sides, we have

$$f(\sigma_\sharp)^2 = \mathrm{Tr}\left[\sum_{i,j=1}^{n} w_i w_j \rho_i^{\frac{1}{2}} U_i U_j^* \rho_j^{\frac{1}{2}} \right] = \sum_{i,j=1}^{n} w_i w_j \, \mathrm{Re}\,\mathrm{Tr}\left[\rho_i^{\frac{1}{2}} U_i U_j^* \rho_j^{\frac{1}{2}} \right]. \tag{3.46}$$

Noting that $\mathrm{Re}\,\mathrm{Tr}\left[\rho_i^{\frac{1}{2}} V \rho_j^{\frac{1}{2}}\right] \leq \mathrm{F}(\rho_i, \rho_j)$ for any unitary $V \in \mathrm{U}(\mathcal{H})$, we have

$$f(\sigma_\sharp)^2 \leq \sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j). \tag{3.47}$$

Taking square root over both sides, we obtain Eq. (3.43).

To see that the inequality is saturated in the commuting case, note that if ρ_i and $\sigma_\sharp$ commute, we have $U_i = \mathbb{I}_{\mathcal{H}}$. Hence Eq. (3.45) reduces to

$$f(\sigma_\sharp)^2 = \mathrm{Tr}\left[\sum_{i,j=1}^{n} w_i w_j \rho_i^{\frac{1}{2}} \rho_j^{\frac{1}{2}} \right] = \sum_{i,j=1}^{n} w_i w_j \mathrm{Tr}\left[\rho_i^{\frac{1}{2}} \rho_j^{\frac{1}{2}} \right] = \sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j), \tag{3.48}$$

where, for commuting states ρ_i and ρ_j, we have $\mathrm{F}(\rho_i, \rho_j) = \mathrm{Tr}\left[\rho_i^{\frac{1}{2}} \rho_j^{\frac{1}{2}}\right]$. Taking square root across, inequality saturation of Eq. (3.43) follows. □

Lemma 3.4.2 bounds optimal average fidelity in the case where all the states in the ensemble are full rank. This can be extended to ensembles with arbitrary (such as rank-deficient) states by continuity, which brings us to the Product bound for arbitrary ensemble.

Theorem 3.4.3 (Product bound). *Let $\mathcal{R} = \{\rho_1, \ldots, \rho_n\} \in \mathbb{D}_d$ be an arbitrary collection of quantum states and let $w \in \Delta_n$ be a probability vector. Then for any state $\sigma \in \mathbb{D}_d$, it holds that*

$$f(\sigma) \leq \sqrt{\sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)}. \tag{3.49}$$

Proof. Let $\mathcal{R}_\epsilon$ denote the ensemble obtained by depolarizing all the states in $\mathcal{R}$ by a factor of $\epsilon \in [0, 1]$:

$$\mathcal{R}_\epsilon := \{\rho_i'(\epsilon) = (1 - \varepsilon)\rho_i + \varepsilon \mathbb{I}_{\mathcal{H}}/d : \rho_i \in \mathcal{R}\}. \tag{3.50}$$

Let $J(\epsilon)$ denote the gap of the Product bound when the ensemble is depolarized by a factor of ϵ:

$$J(\epsilon) = \sqrt{\sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}\left(\rho_i', \rho_j'\right)} - \sum_{i=1}^{n} \mathrm{F}(\rho_i', \sigma_\sharp), \tag{3.51}$$

where $\sigma_\sharp$ is the optimal state over the ensemble $\mathcal{R}_\epsilon$ and we have dropped the ϵ while writing ρ_i' for brevity.

For any $\epsilon \in (0, 1]$, all of the states in $\mathcal{R}_\epsilon$ are full rank and thereby $J(\epsilon) \geq 0$ (as the Product bound holds for full-rank states by Lemma 3.4.2). We are interested in the case where $\epsilon = 0$, which implies $\mathcal{R}_\epsilon = \mathcal{R}$, thereby the ensemble can now contain rank-deficient states.

Note that both the terms in RHS of Eq. (3.51) are continuous functions of ϵ (for $\epsilon \in [0, 1]$) as they're both compositions of continuous functions (ρ_i' is continuous in ϵ and fidelity is continuous in its arguments). Hence we have $J(\epsilon)$ to be continuous in ϵ. By noting that $J(\epsilon)$ is continuous in ϵ along with the fact that $J(\epsilon) \geq 0$ for $\epsilon \in (0, 1]$, we conclude that $J(\epsilon = 0) \geq 0$. Equivalently, Product bound holds for arbitrary ensembles (that may include rank-deficient states). □

Remark 3.4.4. For an ensemble $(\mathcal{R}, p)$, the average fidelity of the Commuting estimator σ' Eq. (3.39) and the Product bound Eq. (3.49) are lower and upper

bounds to optimal average fidelity respectively:

$$f(\sigma') \leq f(\sigma_\sharp) \leq \sqrt{\sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)}. \tag{3.52}$$

Moreover, these bounds coincide when all the states in the ensemble $\mathcal{R}$ commute pairwise.

We now derive the *average bound*, which is a consequence of joint concavity of fidelity.

Theorem 3.4.5. *Let $\mathcal{R} = \{\rho_1, \ldots, \rho_n\} \in \mathbb{D}_d$ be an arbitrary collection of quantum states, $w \in \Delta_n$ be a probability vector, and $\sigma_{\mathrm{M}} = \sum_{i=1}^{n} w_i \rho_i$ be the Mean estimator. Then for any state $\sigma \in \mathbb{D}_d$, it holds that*

$$f(\sigma) \leq \sqrt{\sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)} \leq \sqrt{f(\sigma_{\mathrm{M}})}. \tag{3.53}$$

Proof. We have

$$f(\sigma_{\mathrm{M}}) := \sum_{i=1}^{n} w_i \, \mathrm{F}(\rho_i, \sigma_{\mathrm{M}}) = \sum_{i=1}^{n} w_i \, \mathrm{F}\left(\rho_i, \sum_{j=1}^{n} w_j \rho_j\right) \geq \sum_{i,j=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j), \tag{3.54}$$

where the inequality comes from the concavity of fidelity (see Proposition 2.3.8). Combining Eq. (3.49) and Eq. (3.54) and taking square root across, and noting that $f(\sigma) \leq f(\sigma_\sharp)$ for any state $\sigma \in \mathbb{D}_d$, we obtain Eq. (3.53). □

3.5 Ω Fixed-point algorithm as projected Riemannian gradient descent

As discussed in Chapter 2, Riemannian gradient descent generalizes gradient descent (on the Euclidean manifold) to Riemannian manifolds. RGD has found applications in various optimization problems, including in classical and quantum information theory [Wan+24; Hsu+24; VNM24].

We now show that the fixed-point algorithm $\Omega = \Pi \circ \mathrm{K}$ defined in Eq. (3.37) is can be seen an instance of projected Riemannian gradient descent on the Bures–Wasserstein manifold. This is based on the two following facts.

1. Each step of the K algorithm (Eq. (3.36)) can be seen as a single step of gradient descent on the Bures–Wasserstein manifold, a fact first noted

in [ZP19], and has since been studied in more detail in various works including [Che+20; Alt+21; Alt+23; KSS21].

2. As mentioned previously, the trace-normalization operation Π can be seen a Bures projection onto the set of density matrices.

The rest of this section is devoted to proving the above two facts.

Recall that the update rule of a general gradient descent algorithm on a Riemannian manifold $(\mathcal{M}, \mathfrak{g})$ can be written as

$$x_{t+1} = \operatorname{Exp}^{\mathfrak{g}}_{x_t}[\eta_t \overline{\nabla} f(x_t)], \tag{3.55}$$

where $f : \mathcal{M} \to \mathbb{R}$ is the function we are trying to minimize, x_t is iterate and η_t is the learning rate at the current (t^{th}) step. Here the *Riemannian gradient* $\overline{\nabla} f(x)$ is the dual of the total derivative $\mathrm{D}f(x)$ with respect to the tangent space inner product $\mathfrak{g}_x$ at x. That is, for each $x \in \mathcal{M}$, the Riemannian gradient $\overline{\nabla} f(x) \in \mathrm{T}_x\mathcal{M}$ is the unique tangent vector satisfying

$$\langle \overline{\nabla} f(x), v \rangle_x = \mathrm{D}f(x)(v), \tag{3.56}$$

for all tangent vectors $v \in \mathrm{T}_x\mathcal{M}$, with $\langle \cdot, \cdot \rangle^{\mathfrak{g}}_x \equiv \mathfrak{g}_x(\cdot, \cdot)$, and $\operatorname{Exp}^{\mathfrak{g}}_x : \mathrm{T}_x\mathcal{M} \to \mathcal{M}$ is the Riemannian exponential map at x. See Section 2.4.7 for further details. If the gradient descent is used for a constrained optimization problem over a feasible set $\mathcal{C} \subset \mathcal{M}$, then a *projected* gradient descent performs a *projection* after each step of the above form Eq. (3.55). The projection is to be with respect to the natural distance of the manifold $(\mathcal{M}, \mathfrak{g})$:

$$\Pi_{\mathcal{C}}[x] = \operatorname*{argmin}_{y \in \mathcal{C}} \mathrm{d}^2_{\mathfrak{g}}(y, x). \tag{3.57}$$

Thus a single step of a projected RGD can be written as

$$x_{t+1} = (\Pi_{\mathcal{C}} \circ \operatorname{Exp}^{\mathfrak{g}}_{x_t})(\eta_t \overline{\nabla} f(x_t)). \tag{3.58}$$

Here the manifold of interest is the Bures manifold and the feasible set is $\mathcal{C} = \mathbb{D}_d$. As mentioned before (see Eq. (3.4)) and proved in 5.2.6, the projection with respect to Bures distance onto the set of density matrices is given by trace-normalization: $\Pi_{\mathbb{D}_d}[P] = P/\operatorname{Tr}[P]$. Thus the trace-normalization we apply after each iteration of $\Omega = \Pi \circ \mathrm{K}$ fixed-point algorithm is in fact the Bures projection onto the set of density matrices.

We next show that the map K is the application of a single step of the Riemannian gradient descent for unit step-size on the Bures manifold of the average

squared Bures distance function.

$$\begin{aligned} b(\sigma) := \frac{1}{2}\sum_{i=1}^{n} w_i \mathrm{B}(\rho_i, \sigma) &= \frac{1}{2}\sum_{i=1}^{n} w_i [\mathrm{Tr}[\rho_i] + \mathrm{Tr}[\sigma] - 2\,\mathrm{F}(\rho_i, \sigma)] \\ &= \frac{1}{2}\sum_{i=1}^{n} \mathrm{Tr}[w_i \rho_i] + \mathrm{Tr}[\sigma] - 2f(\sigma). \end{aligned} \tag{3.59}$$

An inspection reveals that the minimization of the function b is equivalent to the maximization of f, provided the argument is constrained to be in $\mathbb{D}_d$. A straightforward computation reveals the form of the gradient of b:

$$\nabla b(\sigma) = \frac{\mathbb{I} - 2\nabla f(\sigma)}{2}. \tag{3.60}$$

The gradient of the average fidelity function f can be computed as

$$\nabla f(\sigma) = \nabla_\sigma \left(\sum_{i=1}^{n} w_i \,\mathrm{F}(\rho_i, \sigma) \right) = \left(\sum_{i=1}^{n} w_i \nabla_\sigma \,\mathrm{F}(\rho_i, \sigma) \right) = \frac{1}{2}\sum_{i=1}^{n} w_i \sigma^{-1} \# \rho_i, \tag{3.61}$$

where we have used the fact (see Proposition 2.3.8) that $\nabla_P \,\mathrm{F}(P, Q) = \frac{1}{2} P^{-1} \# Q$ for $P, Q > 0$.

We now show that the fixed-point algorithm K is an instance of RGD on the Bures manifold. By Proposition 2.4.7, we can write the update rule for RGD on the Bures manifold for a differentiable function $g : \mathbb{P}_d^+ \to \mathbb{R}$ as

$$Q_{t+1} := \mathrm{Exp}_{Q_t}^{\mathrm{Bu}}[-\eta_t \overline{\nabla} g(Q_t)] = Q_t \star [\mathbb{I} - 2\eta_t \nabla g(Q_t)]^2, \tag{3.62}$$

where $\overline{\nabla} g$ denotes the Riemannian gradient of the function g with respect to the Bures metric (see Section 2.4.5 for details). Choosing $g = b$ to be the average squared Bures distance function and current iterate as σ_t, we get

$$\begin{aligned} \sigma_{t+1} := \mathrm{Exp}_{\sigma_t}^{\mathrm{Bu}} \left[-\eta_t \overline{\nabla} b(\sigma_t)\right] &= \sigma_t \star [\mathbb{I} - 2\eta_t \nabla b(\sigma_t)]^2 \\ &= \sigma_t \star [\mathbb{I} - \eta_t (\mathbb{I} - 2\nabla f(\sigma_t)]^2 \\ &= \sigma_t \star \left[(1 - \eta_t)\mathbb{I} + \eta_t \sum_{i=1}^{n} w_i \sigma_t^{-1} \# \rho_i \right]^2. \end{aligned} \tag{3.63}$$

On choosing step size $\eta_t = 1$, we get

$$\begin{aligned}\sigma_{t+1} = \sigma_t \star [2\nabla f(\sigma_t)]^2 &= \sigma_t \star \left[\sum_{i=1}^{n} w_i \sigma_t^{-1} \# \rho_i\right]^2 \\ &= \sigma_t^{-\frac{1}{2}} \left(\sum_{i=1}^{n} w_i \sqrt{\sigma_t^{\frac{1}{2}} \rho_i \sigma_t^{\frac{1}{2}}}\right)^2 \sigma_t^{-\frac{1}{2}} = \mathrm{K}(\sigma_t),\end{aligned} \tag{3.64}$$

which is exactly the update rule for the K fixed-point algorithm (Eq. (3.36)).

By Corollary 5.2.6 we know that trace normalization is the projection (with respect to Bures distance) onto the set of density matrices, which means the Ω fixed-point algorithm can be written as

$$\sigma_{t+1} := \Omega(\sigma_t) \equiv \Pi[\mathrm{K}(\sigma_t)] = \Pi\left[\mathrm{Exp}_{\sigma_t}^{\mathrm{Bu}}\left[-\overline{\nabla} b(\sigma_t)\right]\right]. \tag{3.65}$$

Thus the fixed-point algorithm of interest can be seen as an instance of a *projected Riemannian gradient descent* on the Bures manifold.

The identification also allows us to readily adapt the convergence guarantees for the Ω FP algorithm. Recall in Theorem 3.4.1 we only showed asymptotic convergence of Ω by adapting the proof of asymptotic convergence of K from [BJL19]. However, by identifying the connection to RGD on Bures manifold, one could now extend more quantitative convergence guarantees thanks to the extensive work on this field in the optimal transport literature [Che+20; Alt+21; Alt+23].

We now show two other fixed-point algorithms can be interpreted as RGD on the Bures manifold.

3.6 Other fixed-point algorithms as Riemannian gradient descent

In the previous section, we showed that each iterate fixed-point algorithm K, can be written as (proportional to) $\nabla f(\sigma_t)\sigma_t \nabla f(\sigma_t)$. We then went on to show that this is equivalent to GD on the Bures manifold with respect to the average squared Bures distance function. Indeed on could have also defined the function $b'(\sigma) := \frac{1}{2}\mathrm{Tr}[\sigma] - f(\sigma)$, which is an equivalent problem as they share the optimum, provided we restrict σ to be a density matrix. One can also see that plugging it into the RGD update rule (Eq. (3.62)) would give the same K FP algorithm, which in turn yields Ω FP algorithm once we perform the projection as well.

This section aims to show that there exist other protocols that are well known in quantum information which are demonstrably RGD on the Bures manifold. These are the $R\rho R$ algorithm for state tomography [Hra97; Řeh+07] and the FP

algorithm for the *matrix projection problem* [BRT23].

3.6.1 $R\rho R$ algorithm as projected Riemannian gradient descent

Let us first formally state the problem of maximum likelihood quantum state tomography [Hra97; Řeh+07; RHJ01; Blu10b; LCL21; LC19]. Let $\mathcal{H} = \mathbb{C}^d$ and let $\mathcal{E} = \{E_i\}_{i\in[n]}$ be a POVM associated with $\mathcal{H}$. Suppose we have measurement data in the form $\mathsf{D} = (m_i)_{i\in[n]}$. That is, we have prepared and measured the system a total of m times and we observed the outcome i a total of m_i times (with $\sum_{i=1}^n m_i = m$).

We want to find the state $\rho \in \mathbb{D}_d$ which maximizes the *likelihood* of being the true state. This is formalized using the *likelihood* function, which is defined as

$$L(\rho) \equiv L(\rho|\mathsf{D}) = \Pr[\mathsf{D}|\rho] := \frac{m!}{\prod_{i=1}^n m_i!} \prod_{k=1}^n \langle E_i, \rho\rangle^{m_i}. \tag{3.66}$$

It is easier to work with the log of this function, as it is more numerically stable, and thus we may write

$$\log L(\rho) := c + \sum_{k=1}^n m_i \log\langle E_i, \rho\rangle, \tag{3.67}$$

where c is the constant term stemming from the log of the combinatorial factor, which may be ignored. One may divide throughout by m to normalize the frequencies $z_i := m_i/m$, and we arrive at the objective function of interest:

$$\ell(\rho) := \sum_{k=1}^n z_i \log\langle E_i, \rho\rangle. \tag{3.68}$$

The *maximum likelihood estimator* is the maximizer of this function over $\mathbb{D}_{\mathcal{H}}$:

$$\sigma_{\text{MLE}} := \operatorname*{argmax}_{\rho\in\mathbb{D}_{\mathcal{H}}} \ell(\rho). \tag{3.69}$$

σ_{MLE} is then reported as an estimate for the underlying true quantum state.

In the $R\rho R$ algorithm, the Eq. (3.68) is extremized to show that the *optimal state* would satisfy the fixed-point equation $\rho = R\rho R$ where

$$R \equiv R(\rho) := \nabla\ell(\rho) = \sum_{i=1}^n \frac{z_i}{\langle E_i, \rho\rangle} E_i \tag{3.70}$$

is the gradient of the objective function. We see that if there exists a ρ which is

consistent at the measurement data, then $R(\rho) = \mathbb{I}_d$. The $R\rho R$ algorithm, as the name suggests, proceeds in an iterative manner as follows

$$\rho_{t+1} = \Pi[R(\rho_t)\rho_t R(\rho_t)] = \Pi[\nabla\ell(\rho_t)\rho_t\nabla\ell(\rho_t)], \tag{3.71}$$

which is reminiscent of the form we had for K FP algorithm. To view the $R\rho R$ algorithm as an instance of the RGD algorithm, consider the augmented objective function of the form $\hat{\ell}(\rho) := \frac{1}{2}[\mathrm{Tr}[\rho] - \ell(\rho)]$, whose gradient is given by $\nabla\hat{\ell}(\rho) = \frac{1}{2}[\mathbb{I} - \nabla\ell(\rho)]$. Observe that any minimizer of the convex function $\hat{\ell}$ is a maximizer of the concave function ℓ and vice versa. Plug this into the update rule for RGD Eq. (3.62) and we have

$$\begin{aligned}\mathrm{Exp}_{\rho_t}^{\mathrm{Bu}}\left[-\eta_t\overline{\nabla}\hat{\ell}(\rho_t)\right] = \rho_t \star \left[\mathbb{I} - 2\eta_t\nabla\hat{\ell}(\rho_t)\right]^2 &= \rho_t \star [(1-\eta_t)\mathbb{I} + \eta_t\nabla\ell(\rho_t)]^2 \\ &= \rho_t \star [(1-\eta_t)\mathbb{I} + \eta_t R(\rho_t)]^2.\end{aligned} \tag{3.72}$$

On choosing $\eta_t = 1$, we get the RHS to be of the $R\rho R$ form.

The $R\rho R$ typically converges to optimum, but it is possible to construct datasets where it can get stuck in *loops* [Řeh+07]. A simple fix suggested in [Řeh+07] is to *dilute* the R operator at every step:

$$\rho_{t+1} := \Pi\left[\frac{\mathbb{I} + \epsilon R(\rho_t)}{1+\epsilon}\rho_t\frac{\mathbb{I} + \epsilon R(\rho_t)}{1+\epsilon}\right], \tag{3.73}$$

for a small positive $\epsilon \ll 1$. One can immediately see that this corresponds to a projected RGD with step size $\eta_t = \frac{\epsilon}{1+\epsilon}$.

Thus we have shown that the the $R\rho R$ and its diluted version can be interpreted as (projected) Riemannian gradient descent on the Bures manifold. To the best of our knowledge, this is the first time this identification has been made.

3.6.2 Fixed point algorithm for matrix projection as Riemannian Gradient Descent

In this section, we show that the FP algorithm for the *matrix projection problem* as studied in Brahmachari, Rubboli, and Tomamichel [BRT23] is also an instance of RGD on the Bures manifold. We first briefly describe the problem setting.

Let G be a finite group G with a projective unitary representation $\mathcal{U} := \{U_g\}_{g\in G} \subset \mathbb{U}_d$. The *commutant* of $\mathcal{U}$ is defined as

$$\mathrm{comm}(\mathcal{U}) := \{X \in \mathbb{M}_d : XU_g = U_g X \text{ for all } U_g \in \mathcal{U}\}. \tag{3.74}$$

The commutant of any subset of $\mathbb{M}_d$ is a *unital subalgebra* of $\mathbb{M}_d$; i.e., a subset of

$\mathbb{M}_d$ that is closed under scalar multiplication, addition, and operator composition, and that contains the identity element. See Watrous [Wat18, Section 1.1] for further details.

The intersections $\mathcal{H} \equiv \mathbb{H}_d \cap \mathrm{comm}(\mathcal{U})$ and $\mathcal{P} \equiv \mathbb{P}_d \cap \mathrm{comm}(\mathcal{U})$ are called *invariant Hermitian matrices* and *invariant positive matrices* respectively. Given a positive definite matrix $P \in \mathbb{P}_d^+$, the optimization problem of interest in [BRT23] is :

$$\begin{aligned} \text{minimize :} \quad & \mathrm{B}(P, Q) \\ \text{subject to :} \quad & Q \in \mathcal{P}. \end{aligned} \tag{3.75}$$

In words, we want to *project* P to $\mathcal{P}$ with respect to Bures distance. The *twirling map* $\Phi : \mathbb{H}_d \to \mathcal{H}$ is defined as

$$\Phi(H) := \frac{1}{|G|} \sum_{g \in G} U_g H U_g^*. \tag{3.76}$$

Indeed we also have that $\Phi(P) \in \mathcal{P}$ for any $P \in \mathbb{P}_d$.

The FP algorithm to solve the optimization problem in Eq. (3.75) has the update rule

$$Q_{t+1} = Q_t^{-\frac{1}{2}} \left(\Phi \left(\sqrt{Q_t^{\frac{1}{2}} P Q_t^{\frac{1}{2}}} \right) \right)^2 Q_t^{-\frac{1}{2}}, \tag{3.77}$$

which is based on the fact that optimal state $Q_\sharp \in \mathcal{P}$ satisfies the fixed-point equation

$$Q_\sharp = Q_\sharp^{-\frac{1}{2}} \left(\Phi \left(\sqrt{Q_\sharp^{\frac{1}{2}} P Q_\sharp^{\frac{1}{2}}} \right) \right)^2 Q_\sharp^{-\frac{1}{2}}. \tag{3.78}$$

We now show that the update rule can be seen as RGD on the Bures manifold with unit step size. Observe that Problem (3.75) is equivalent to the following optimization problem:

$$\begin{aligned} \text{minimize :} \quad & \mathrm{B}(P, \Phi(Q)) \\ \text{subject to :} \quad & Q \in \mathbb{P}_d, \end{aligned} \tag{3.79}$$

as every invariant positive matrix in $\mathcal{P}$ can be written as $\Phi(Q)$ for some $Q \in \mathbb{P}_d$. For a fixed $P \in \mathbb{P}_d^+$, define the functions $b, b' : \mathbb{P}_d \to \mathbb{R}$ as

$$b(Q) := \frac{1}{2}\mathrm{B}(P, Q) \quad \text{and} \quad b'(Q) := \frac{1}{2}\mathrm{B}(P, \Phi(Q)), \tag{3.80}$$

which allows us to write $b' = b \circ \Phi$. Let us now compute the gradient. Begin with the total derivative and use the chain rule to conclude that

$$\mathrm{D}b'_Q = \mathrm{D}b_{\Phi(Q)} \circ \mathrm{D}\Phi_Q = \mathrm{D}b_{\Phi(Q)} \circ \Phi, \tag{3.81}$$

where we have used the fact that $\mathrm{D}\Phi_Q = \Phi$ as Φ is a linear map. Moreover, for any $R \in \mathbb{P}_d^+$, we have

$$\mathrm{D}b_R(Z) = \frac{1}{2}\operatorname{Tr}[(\mathbb{I} - R^{-1}\#P)Z], \tag{3.82}$$

for any $Z \in \mathrm{T}_R\mathbb{P}_d^+ \cong \mathbb{H}_d$. Thus we have

$$\mathrm{D}b'_Q(Z) = \mathrm{D}b_{\Phi(Q)}(\Phi(Z)) = \frac{1}{2}\operatorname{Tr}[(\mathbb{I} - \Phi(Q)^{-1}\#P)\Phi(Z)]. \tag{3.83}$$

For the fixed-point algorithm, we will be interested in the cases where $Q \in \mathcal{P}$, which implies $\Phi(Q) = Q$. The fact that Φ is self-adjoint (with respect to HS inner product), allows us to write the gradient as

$$\nabla b'(Q) = \frac{\mathbb{I} - \Phi(Q^{-1}\#P)}{2}. \tag{3.84}$$

We now have all the ingredients to show that the update rule Eq. (3.77) is an instance of Riemannian gradient descent. Let $Q_t \in \mathcal{P}$ be the current iterate and $\eta_t \in [0,1]$ be the step size. Using Proposition 2.4.7, we have

$$\begin{aligned} Q_{t+1} = \operatorname{Exp}^{\mathrm{Bu}}_{Q_t}[-\eta_t \overline{\nabla} b'(Q_t)] &= Q_t \star [\mathbb{I} - 2\eta_t \nabla b'(Q_t)]^2 \\ &= Q_t \star \left[\mathbb{I} - 2\eta_t \frac{1}{2}\left(\mathbb{I} - \Phi(Q_t^{-1}\#P)\right)\right]^2 \\ &= Q_t \star [(1-\eta_t)\mathbb{I} + \eta_t \Phi(Q_t^{-1}\#P)]^2. \end{aligned} \tag{3.85}$$

Choosing a unit step size ($\eta_t = 1$), we have

$$\begin{aligned} Q_{t+1} = Q_t \star [\Phi(Q_t^{-1}\#P)]^2 &= Q_t \star \left[Q_t^{-\frac{1}{2}} \Phi\left(\sqrt{Q_t^{\frac{1}{2}} P Q_t^{\frac{1}{2}}}\right) Q_t^{-\frac{1}{2}}\right]^2 \\ &= Q_t^{-\frac{1}{2}} \left(\Phi\left(\sqrt{Q_t P Q_t}\right)\right)^2 Q_t^{-\frac{1}{2}}, \end{aligned} \tag{3.86}$$

where, in the second equality, we have used the fact that

$$\Phi(Q_t^{-1}\#P) = \Phi\left(Q_t^{-\frac{1}{2}} \sqrt{Q_t^{\frac{1}{2}} P Q_t^{\frac{1}{2}}} Q_t^{-\frac{1}{2}}\right) = Q_t^{-\frac{1}{2}} \Phi\left(\sqrt{Q_t^{\frac{1}{2}} P Q_t^{\frac{1}{2}}}\right) Q_t^{-\frac{1}{2}} \tag{3.87}$$

as $Q_t \in \mathcal{P}$ commutes with every $U_g \in \mathcal{U}$.

We have recovered the update rule Eq. (3.77), thereby showing that the FP algorithm for the matrix projection problem studied in [BRT23] is an instance of RGD on the Bures manifold.

3.7 On generalizations of the Λ fixed-point algorithm

Let us go back to the Λ FP algorithm (Eq. (3.33)). We are now at a position to identify it to be of the form

$$\Lambda(\sigma) = \Pi\left[\sum_{i=1}^{n} w_i \sqrt{\sigma^{\frac{1}{2}}\rho_i\sigma^{\frac{1}{2}}}\right] = \Pi\left[\sigma^{\frac{1}{2}}\nabla f(\sigma)\sigma^{\frac{1}{2}}\right], \tag{3.88}$$

where the second equality comes from the fact that $\nabla f(\sigma) = \sum_{i=1}^{n} w_i\sigma^{-1}\#\rho_i$.

It is not known whether this algorithm can be given an interpretation of gradient descent. However, we now identify that such *root-gradient-root* fixed-point algorithms have been studied in other instances as well. In You, Cheng, and Li [YCL22, Eq. 13], the authors are interested in minimizing the average divergence functions

$$g_{\mathrm{PR}}(\sigma) = \sum_{i=1}^{n} w_i D_\alpha^{\mathrm{PR}}(\rho_i\|\sigma) \quad \text{and} \quad g_{\mathrm{S}}(\sigma) = \sum_{i=1}^{n} w_i D_\alpha^{\mathrm{S}}(\rho_i\|\sigma), \tag{3.89}$$

where $D_\alpha^{\mathrm{PR}}(\rho\|\sigma) := \frac{1}{\alpha-1}\log \mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}]$ is the Petz-Rényi divergence [Pet86a] and $D_\alpha^{\mathrm{S}}(\rho\|\sigma) := \frac{1}{\alpha-1}\log \mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]$ is the sandwiched Rényi divergence [Mül+13; WWY14]. The main result of the paper is a mirror descent algorithm with Polyak step size.

Additionally, they briefly discuss a fixed-point algorithm which is of the form $\sigma_{t+1} = \Pi\left[-\sigma_t^{\frac{1}{2}}\nabla g(\sigma_t)\sigma_t^{\frac{1}{2}}\right]$ (see [YCL22, Eq. (13)]), which, they say to be a non-commutative generalization of an FP algorithm proposed in [Nak19] for the classical variant of the problem.

Indeed this FP algorithm can be easily converted to the form of Eq. (3.88) by an appropriate change of sign of the objective function. Thus it appears that FP algorithms of the form Eq. (3.88) have wider applicability than just in maximizing average fidelity.

It would be interesting to unify these methods and see if they can also be interpreted as a type of gradient descent. One could ask similar questions for other fixed-point algorithms of interest in quantum information[LCL21]. Additionally one could ask if projected GD on the Bures manifold is applicable for the functions g_{PR} and g_{S}. We leave these as open problems.

3.8 Applications

We now discuss some applications of the main results.

3.8.1 Bayesian quantum state tomography

In quantum tomography, we aim to reconstruct the state of a quantum system from measurement outcomes. There exists a multitude of different tomography methods such as linear inversion [DP07], projected least square [Guţ+20], maximum likelihood estimation [Hra97; RHJ01], hedged maximum likelihood estimation [Blu10a], compressed sensing tomography [Gro+10; Fla+12; Kli+16], and Bayesian quantum tomography [Blu10c; HH12; GCC16]. We now discuss how our results find application in Bayesian quantum tomography.

In tomography or state estimation, we are presented with *measurement data* $\mathsf{D} = (m_i)_{i=1}^n$, where m_i is the number of times we see E_i, which is an element of the n-outcome POVM $\mathcal{E}$ and $m = \sum_i m_i$ denotes the total number of measurements done. The likelihood of obtaining the data D given any true state ρ is then computed as

$$L(\rho) = \Pr(\mathsf{D}|\rho) = \frac{m!}{\prod_i (m_i!)} \prod_i \operatorname{Tr}(E_i \rho)^{m_i}. \tag{3.90}$$

As previously discussed, in maximum likelihood estimation, we are concerned with finding the state that maximizes this likelihood function:

$$\sigma_{\mathrm{MLE}} = \operatorname*{argmax}_{\sigma \in \mathbb{D}_d} L(\sigma). \tag{3.91}$$

In practical Bayesian state estimation [Gra+17], we instead begin with a collection of states $\mathcal{R} = \{\rho_i\}_{i=1}^n$ called *particles* and a prior distribution $u \in \Delta_n$ over them. Usually the prior is taken to be as uniform as possible, but our results work for arbitrary distributions so this is not a concern here. Using Bayes' rule we then compute the posterior distribution $w \in \Delta_n$ as

$$w_i \propto u_i L(\rho_i), \tag{3.92}$$

with the probabilities w_i being normalized afterwards. Once we have the posterior distribution, the Bayes estimator is the state that maximizes the posterior average fidelity over the ensemble $(\mathcal{R}, p)$.

The Bayes estimator is then reported as an estimate for the true state which generated the measurement data. The main results of this work tell us, given the posterior distribution, how to compute the Bayes estimator for fidelity, heuristic approximations for it, and bounds for the maximum average fidelity.

3.8.2 Secrecy measure in quantum key distribution and success probability of quantum interactive proof

Konig, Renner, and Schaffner [KRS09] provide direct operational interpretations to (conditional) min- and max-entropies. They also introduce a related secrecy measure for quantum key distribution protocols defined as follows.

$$p_{\text{sec}}(\mathcal{R}, w) := \max_{\sigma \in \mathbb{D}} \sum_{i=1}^{n} \sqrt{w_i}\, \mathrm{F}(\rho_i, \sigma), \tag{3.93}$$

where $\mathcal{R} = (\rho_1, \ldots, \rho_n)$ is an ensemble of states and $w \in \Delta_n$ is a probability vector. Observe this is exactly the optimization problem we study, with weights corresponding to square-root probabilities. Indeed, our fixed-point algorithm can be extended to consider arbitrary positive weights, so it works for this problem as well.

Rethinasamy, Agarwal, Sharma, and Wilde [Ret+23] shows that the square of this quantity corresponds to the maximum success probability with which a prover can pass a test for an ensemble of states being similar in a quantum interactive proof.

Hence our results can provide numerical solutions for the optimal solution $\sigma_\sharp$, approximately optimal solution σ', and tight bounds in these scenarios.

3.9 Numerical experiments

We consider four different numerical experiments in this work. First, we look at the performance (runtime) of the different methods to obtain the optimal state over random distributions of full-rank states (Fig. 3.2). We then simulate Bayesian tomography and showcase how our results can improve on state-of-the-art methods (Fig. 3.3). Finally, we numerically demonstrate the relative tightness of the bounds we derived (Fig. 3.4). In Appendix B.2, we compare the performance of two fixed-point methods Λ and Ω (Fig. B.1).

We consider the fixed-point methods to have converged (i.e., the stopping condition) if the spectral norm of the difference between two consecutive iterations is less than some tolerance ϵ. In Fig. 3.2, we choose $\epsilon = 10^{-4}$. We use CVXPY [DB16; Agr+18] to solve SDPs numerically. The numerics were done on Google Colab (single-core CPU at 2.3 GHz and approximately 12 GB RAM). The code is available on GitHub [Afh22].

3.9.1 Comparing performance of SDP and Omega FP algorithm for optimal average fidelity

The semidefinite program as defined in Definition 3.3.1 can be solved numerically using standard convex optimization libraries. However, owing to the size of the matrices over which we optimize—positive semidefinite matrices of size $(n+1)d \times (n+1)d$—the process is quite intractable for even moderately large n and d. This intractability can be partially resolved by defining an *alternate SDP* for optimal average fidelity which reformulates the problem into solving n SDPs over matrices of size $2d \times 2d$ (see Appendix B.1 for definition and details). The alternate SDP provides a more favorable scaling as the number of states n in the ensemble increases. Moreover, the runtime scales (roughly) linear in the number of states as compared to superlinear for the original SDP.

As seen in Fig. 3.2, the Ω fixed-point method vastly outperforms both SDPs in terms of runtime. Though both the alternate SDP and FP methods scale linearly in n, The FP method is orders of magnitude faster. In particular, for 5 qubit states ($d = 32$), the FP method was faster than alternate SDP by a factor of 68 on average. The plots also show that the time taken for the FP method to obtain a solution in the 5-qubit scenario is comparable to the time it takes the alternate SDP to solve the 1-qubit case. The figure also shows how intractable solving the original SDP can be. For even just 1 qubit ($d = 2$) and $n = 20$ states, it takes more time than the FP method takes for 5 qubits and $n = 20$. Moreover, the FP method can be easily parallelized at each iteration, as the n different terms (see Eq. (3.37)) in the sum can be computed in parallel, thereby further boosting performance.

3.9.2 Simulating Bayesian tomography

In the second set of experiments, we simulate Bayesian tomography. The results are presented in Fig. 3.3. Since simulating measurement and computing the posterior distribution is expensive and outside the scope of this paper, we assign posterior weights as follows. We begin with a randomly generated *true state* ρ_{T}. In Bayesian state estimation, we begin with a set of hypothesis states and an associated distribution. The posterior distribution would be peaked near the true state and as we increase the number of measurements, the peak sharpens.

To simulate this behavior, we introduce a parameter $\lambda \in [0, 1]$, and then for randomly generated $\{\rho_i'\}_{i=1}^n$, we choose our particles as $\rho_i = \lambda\rho_{\mathrm{T}} + (1-\lambda)\rho_i'$. As $\lambda \to 1$, the particles $\{\rho_i'\}_{i=1}^n$ are closer to the true state ρ_{T}. The unnormalized weights are then chosen as $\mathrm{F}(\rho_i, \rho_{\mathrm{T}})$, which is then normalized to serve as the

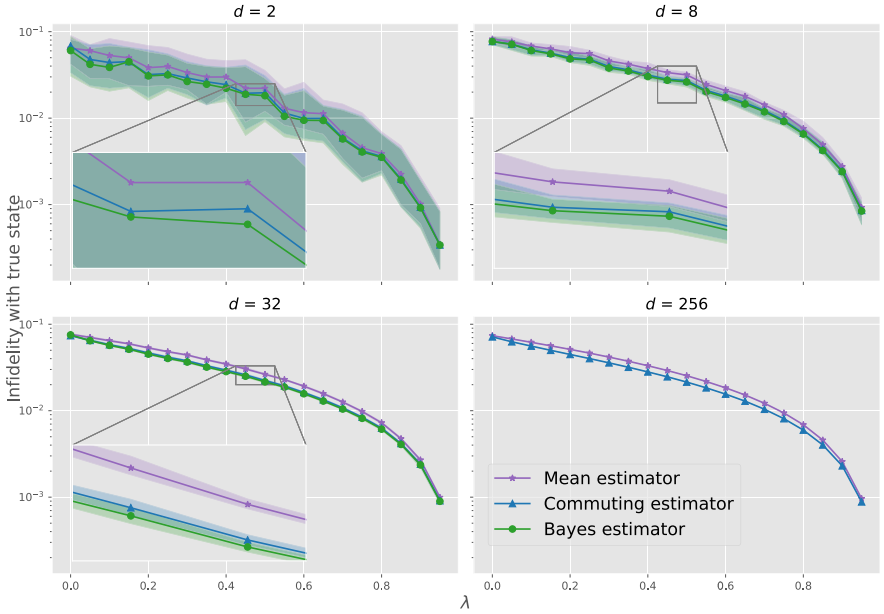

Figure 3.3: Infidelity with true state as a function of λ for various dimensions ($n = 20$) plotted in log scale. $\lambda \in [0, 1]$ is a parameter to simulate measurement counts with $\lambda = 0$ indicating 0 measurements and $\lambda = 1$ indicating infinitely many measurements. Each point is the median of 50 iterations. Interquartile regions are shaded. As seen in the inset plots, the Commuting estimator can serve as a high-quality proxy (better than the Mean estimator) to the Bayes estimator while being inexpensive to compute.

weights w_i:

$$w_i \propto \mathrm{F}(\rho_i, \rho_{\mathrm{T}}), \quad \sum_{i=1}^{n} w_i = 1. \tag{3.94}$$

This allows the assignment of higher weights to ρ_is that are closer to the true state ρ_{T}. We then compute the Mean estimator $\sigma_{\mathrm{M}} = \sum_{i=1}^{n} w_i \rho_i$, the Commuting estimator $\sigma' = \Pi\left(\left(\sum_{i=1}^{n} w_i \rho_i^{1/2}\right)^2\right)$ and the Bayes estimator (optimal estimator) $\sigma_\sharp$. In tomography, one is usually interested in the behavior of infidelity with the true state as a function of the number of measurements. To simulate this, we vary λ from 0 to 1, as $\lambda = 0$ would correspond to the hypothesis states being randomly initialized states (zero measurements) and $\lambda = 1$ would correspond to all the particles being equal to the true state (infinitely many measurements). Fig. 3.3 plots the infidelity of the Bayes estimator, Commuting estimator, and Mean estimator with the true state for various dimensions and $n = 20$. For higher dimensions, we drop the Bayes estimator due to the computational costs while noting that the Commuting estimator σ' remains a good alternative while being inexpensive to compute.

3.9.3 Tightness of bounds and other estimators with optimum

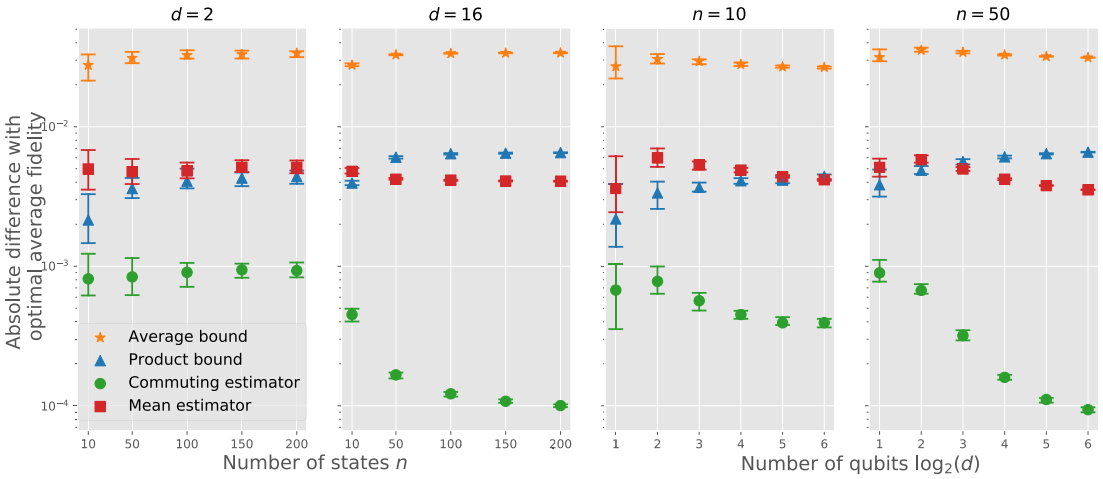

Figure 3.4: Tightness of quantities of interest with optimal average fidelity as a function of (a) number of states n and (b) number of qubits ($\log_2(d)$). We plot the *absolute difference* with the optimal average fidelity of (i) Average bound Eq. (3.53), (ii) Product bound Eq. (3.49), (iii) average fidelity of Commuting estimator Eq. (3.39), and (iv) average fidelity of Mean estimator $\sigma_{\mathrm{M}} = \sum_{i=1}^{n} w_i \rho_i$. Note that the Average bound and Product bound are upper bounds and the average fidelities of the Mean estimator and Commuting estimator are lower bounds. Each data point is the median of 50 iterations and ticks correspond to interquartile regions. See Eq. (3.95) for details on the quantities being plotted.

In Fig. 3.4, we numerically demonstrate the tightness of the two upper bounds and the fidelity achieved by the Commuting estimator and Mean estimator with the maximum average fidelity for randomly generated ensembles for various dimensions and number of states. More formally, for randomly generated ensembles $(\mathcal{R}, p)$, we plot the quantity $|f(\sigma_\sharp) - g|$, where

$$g = \begin{cases} \sqrt{f(\sigma_{\mathrm{M}})} & \text{(Average bound)}, \\ \sqrt{\sum_{i=1}^{n} w_i w_j \, \mathrm{F}(\rho_i, \rho_j)} & \text{(Product bound)}, \\ f(\sigma') & \text{(Commuting estimator)}, \\ f(\sigma_{\mathrm{M}}) & \text{(Mean estimator)}. \end{cases} \tag{3.95}$$

Note that the first two are upper bounds while the last two are lower bounds on optimal average fidelity $f(\sigma_\sharp)$. Since the average fidelity of the Mean estimator and Product bounds are different kinds of bounds (lower and upper respectively), their crossing is not unexpected, and it simply means that the lower bound gets closer to the optimum than the upper bound. As the plots show, the Product

bound and average fidelity of the Commuting estimator are quite close to the optimal average fidelity.

3.10 Conclusion

In this work, we present algorithms for identifying states that maximize average fidelity over arbitrary finite ensembles of quantum states. We have constructed semidefinite programs that solve this problem, from which we derive faster fixed-point algorithms for the scenario where all the states in the ensemble are full rank. The fixed-point methods are orders of magnitude faster than the semidefinite programs. We also derive heuristic approximations for the optimal state which are exact when the states in the ensemble commute pairwise. Furthermore, we derive novel upper and lower bounds for maximum average fidelity achievable by any quantum state, which are saturated when all the states commute pairwise. Finally, we present numerical experiments to complement our theoretical findings. These results solve open problems in Bayesian quantum tomography and are of independent theoretical interest.

An interesting question to ask is if the Λ FP algorithm can be endowed with an interpretation of gradient descent. Indeed, as we showed in Section 3.7, the same *root-gradient-root* approach has been studied elsewhere, and thus a unification might exist. A second problem we as is whether there exists other fixed-point algorithms of interest in quantum information that can be interpreted as Riemannian gradient descent. We have shown that the $R\rho R$ algorithm falls into this category, and we plan to study other similar iterative algorithms.

Chapter 4

Generalizing fidelities

In this chapter, we discuss the second set of results of this thesis, where we propose a family of fidelities we term *generalized fidelity*. Generalized fidelity is motivated by the geometry of the Bures–Wasserstein manifold and has the special property that it can recover various existing quantum fidelities like Uhlmann, Holevo, and Matsumoto fidelities. We will use 'Bures' and 'Bures–Wasserstein' (abbreviated as BW) interchangeably in the chapter.

The generalized fidelity between two positive matrices P and Q is parametrized by a third positive (definite) matrix R called the *base*. See Fig. 4.1 for visualization of the generalized fidelity between rebit states as a function of the base. Surprisingly, the generalized fidelity between two positive semidefinite matrices can be negative—or, more generally, complex.

The generalized fidelity between fixed P and Q shows remarkable geometric properties as we vary the base. We also show that convex combinations of generalized fidelities over different bases also define valid quantizations of classical fidelity and that generalized fidelities can be thought of as extreme points of this class of (further generalized) quantum fidelities. Based on this notion, we define *polar fidelity* which is a family of fidelities, parametrized by a single real number, that can recover Uhlmann-, Holevo-, and Matsumoto-fidelity. To our knowledge, this is the first such generalization. Other results we explore in this chapter include

1. A block-matrix characterization of generalized fidelity.
2. An 'Uhlmann-like' theorem for generalized fidelity which relates the generalized fidelity to specific purifications of the states involved.
3. Potential applications of generalized fidelity.

This chapter is structured as follows. We first begin with the definitions of generalized fidelity and generalized Bures distance. We then discuss some basic

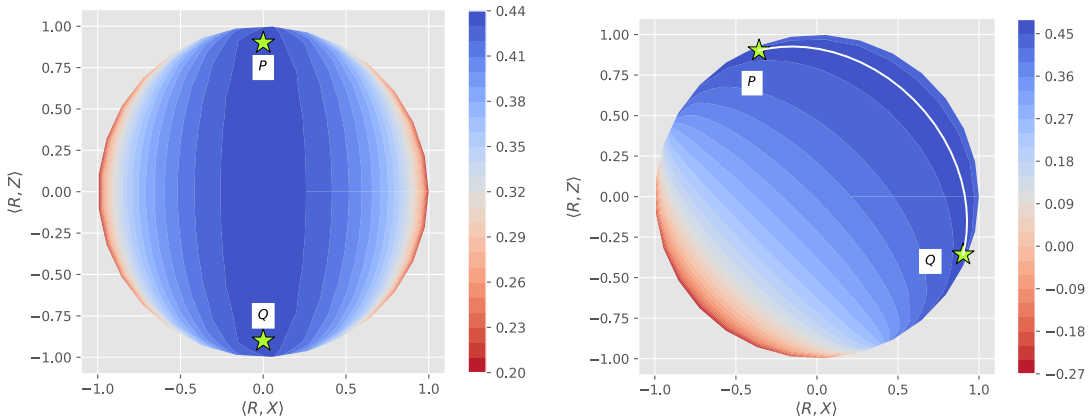

Figure 4.1: Contour plots of the generalized fidelity $\mathrm{F}_R(P, Q)$ as a function of the base R. Each point indicates the generalized fidelity between P and Q at R, where R is the rebit state with Bloch vector $(\langle R, X\rangle, 0, \langle R, Z\rangle)$. The first subfigure shows that the generalized fidelity between two commuting quantum states can be unequal to the Bhattacharyya coefficient (classical fidelity) for general bases. The second subfigure shows that the generalized fidelity between two states can be negative (generally complex-valued) and that the generalized fidelity attains its maximum value if the base is any point along the Bures–Wasserstein geodesic (white curve) between P and Q. Along this geodesic, the value of generalized fidelity $\mathrm{F}_R(P, Q)$ is equal to the Uhlmann fidelity $\mathrm{F}^{\mathrm{U}}(P, Q)$. Code to generate plots is available at [Afh24].

properties of generalized fidelity and then move on to some basic properties and geometric motivation for generalized Bures distance. We then study the geometric properties of generalized fidelity, which include various invariance and covariance behavior as a function of the base. We then discuss the above-mentioned additional results.

4.1 Generalized fidelity and generalized Bures distance

We now formally introduce generalized fidelity and generalized Bures distance. We begin with the definition and some basic properties of generalized fidelity. We then do the same for generalized Bures distance and finally conclude the section with the Riemannian-geometric motivation for the definition.

4.1.1 Generalized fidelity

Let us now define generalized fidelity.

Definition 4.1.1. *Let $P, Q, R \in \mathbb{P}_d^+$ be positive definite matrices. The* generalized fidelity between P and Q at R *is defined as*

$$\mathrm{F}_R(P,Q) := \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right] \tag{4.1}$$

Here and henceforth, R is referred to as the *base* of the generalized fidelity between P and Q. Throughout the analysis, we require the base R to be positive definite while P and Q can be rank-deficient. However, for most of our analysis, we consider P and Q to be invertible as well. The first thing to notice is that the matrix inside the trace is not necessarily Hermitian. Thus, generalized fidelity is complex-valued in general. Let us list some basic properties of generalized fidelity. Proofs are provided in Appendix C.1.

1. **Quantization of classical fidelity.** The generalized fidelity $\mathrm{F}_R(P,Q)$ is a quantization of classical fidelity for any triple $P, Q, R \in \mathbb{P}_d^+$. That is, if P, Q, and R mutually commute, the generalized fidelity reduces to the classical fidelity (Bhattacharyya coefficient) between P and Q.

2. **Conjugate symmetry.** Generalized fidelity is generally complex-valued and conjugate symmetric:

$$\mathrm{F}_R(P,Q) \in \mathbb{C} \quad \text{and} \quad \mathrm{F}_R(P,Q) = \mathrm{F}_R(Q,P)^* . \tag{4.2}$$

   Moreover, $\mathrm{F}_R(P,P) = \mathrm{Tr}[P]$ for any $P, R \in \mathbb{P}_d^+$. It follows that the generalized fidelity of a normalized state with itself is 1 at any base.

3. **Equivalent forms.** The generalized fidelity has the following equivalent forms.

$$\begin{aligned} \mathrm{F}_R(P,Q) :&= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right] \\ &= \mathrm{Tr}\left[Q^{\frac{1}{2}}U_Q U_P^* P^{\frac{1}{2}}\right] \\ &= \mathrm{Tr}\left[(R^{-1}\#Q)R(R^{-1}\#P)\right], \end{aligned} \tag{4.3}$$

   where $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}}R^{\frac{1}{2}}\right)$, $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}}R^{\frac{1}{2}}\right)$.

4. **Pure state simplification.** If both $P = |\psi\rangle\langle\psi|$ and $Q = |\phi\rangle\langle\phi|$ are pure states, then

$$\mathrm{F}_R(P,Q) = \mathrm{F}_R(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = \frac{\langle\psi,\phi\rangle\langle\phi,R\psi\rangle}{\mathrm{F}^{\mathrm{U}}(P,R)\,\mathrm{F}^{\mathrm{U}}(Q,R)}. \tag{4.4}$$

5. **Commutation with base implies positivity.** If the base R commutes with P or Q, then $\mathrm{F}_R(P,Q) \geq 0$.

6. **Multiplicativity.** Let $P_1, Q_1, R_1 \in \mathbb{P}_{d_1}^+$ and $P_2, Q_2, R_2 \in \mathbb{P}_{d_2}^+$. Then

$$\mathrm{F}_{R_1 \otimes R_2}(P_1 \otimes P_2, Q_1 \otimes Q_2) = \mathrm{F}_{R_1}(P_1, Q_1) \cdot \mathrm{F}_{R_2}(P_2, Q_2)\,. \tag{4.5}$$

7. **Additivity.** Let $P_1, Q_1, R_1 \in \mathbb{P}_{d_1}^+$ and $P_2, Q_2, R_2 \in \mathbb{P}_{d_2}^+$. Then

$$\mathrm{F}_{R_1 \oplus R_2}(P_1 \oplus P_2, Q_1 \oplus Q_2) = \mathrm{F}_{R_1}(P_1, Q_1) + \mathrm{F}_{R_2}(P_2, Q_2)\,. \tag{4.6}$$

8. **Unitary invariance.** For any triple $P, Q, R \in \mathbb{P}_d$ and any unitary $U \in \mathbb{U}_d$,

$$\mathrm{F}_R(P,Q) = \mathrm{F}_{URU^*}(UPU^*, UQU^*)\,. \tag{4.7}$$

9. **Unitary contravariance.** For any triple $P, Q, R \in \mathbb{P}_d^+$ and any unitary $U \in \mathbb{U}_d$,

$$\mathrm{F}_{URU^*}(P,Q) = \mathrm{F}_R(U^*PU, U^*QU)\,. \tag{4.8}$$

10. **Scaling.** For positive scalars $p, q, r \in \mathbb{R}_+$,

$$\mathrm{F}_{rR}(pP, qQ) = \sqrt{pq}\,\mathrm{F}_R(P,Q). \tag{4.9}$$

    Since the generalized fidelity is invariant with respect to any scaling of the base R, one can always choose the base to be a density matrix without loss of generality.

11. **Absolute value is bounded above by Uhlmann fidelity.** For any triple $P, Q, R \in \mathbb{P}_d^+$,

$$|\,\mathrm{F}_R(P,Q)| \leq \mathrm{F}^{\mathrm{U}}(P,Q)\,. \tag{4.10}$$

    This relation implies the nonnegativity of generalized Bures distance.

12. **Orthogonal support implies 0.** If any two matrices from P, Q, R have orthogonal support, then $\mathrm{F}_R(P,Q) = 0$.

13. **Reduction to other fidelities.** For specific choices of the base R, one can recover various named fidelities from generalized fidelity.

    - *Uhlmann fidelity:* If (but not only if) $R = P$ or $R = Q$,

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{U}}(P,Q) := \mathrm{Tr}\left[\sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}}\right]. \tag{4.11}$$

- *Holevo fidelity:* If $R = \mathbb{I}$,

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{H}}(P,Q) := \mathrm{Tr}\left[P^{\frac{1}{2}}Q^{\frac{1}{2}}\right]. \tag{4.12}$$

- *Matsumoto fidelity:* If (but not only if) $R = P^{-1}$ or $R = Q^{-1}$,

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q) := \mathrm{Tr}[P\#Q] = \mathrm{Tr}\left[P^{\frac{1}{2}}\sqrt{P^{-\frac{1}{2}}QP^{-\frac{1}{2}}}P^{\frac{1}{2}}\right]. \tag{4.13}$$

We note that these are not the only cases where generalized fidelity reduces to the above-mentioned fidelities. Other scenarios where this happens are discussed in Section 4.2.

Next, we define the generalized Bures distance and discuss some basic properties and the geometric motivation behind the definition. We also see how generalized fidelity naturally appears in this geometric motivation.

4.1.2 Generalized Bures distance

Having defined generalized fidelity, we are in a position to define the *generalized Bures distance*. After defining it, we will show how it can be naturally derived from the geometry of the Bures–Wasserstein manifold.

Definition 4.1.2. *Let $P, Q, R \in \mathbb{P}_d^+$ be positive definite matrices. The squared* generalized Bures(-Wasserstein) distance *between P and Q at R is defined as*

$$\mathrm{B}_R(P,Q) := \mathrm{Tr}[P+Q] - 2\,\mathrm{Re}\,\mathrm{F}_R(P,Q). \tag{4.14}$$

We use uppercase B to denote squared distance and lowercase b to denote distances. For example, we will henceforth use

$$\mathrm{b}^{\mathrm{U}}(P,Q) \equiv \sqrt{\mathrm{B}^{\mathrm{U}}(P,Q)} = \sqrt{\mathrm{Tr}[P+Q] - 2\,\mathrm{F}^{\mathrm{U}}(P,Q)} \tag{4.15}$$

to denote Bures distance. The superscript indicates the specific type of fidelity used. Similarly, we denote the generalized Bures distance as

$$\mathrm{b}_R(P,Q) \equiv \sqrt{\mathrm{B}_R(P,Q)} = \sqrt{\mathrm{Tr}[P+Q] - 2\,\mathrm{Re}\,\mathrm{F}_R(P,Q)}. \tag{4.16}$$

Essentially, the relation between generalized Bures distance and generalized fidelity is analogous to the relation between Bures distance and Uhlmann fidelity.

We now show that this definition is geometrically motivated. As mentioned before, the generalized Bures distance between P and Q at R is the distance

between P and Q if we *flatten the* Bures–Wasserstein manifold at R. This is formalized in the next section.

Geometric motivation for generalized Bures distance

We now provide a (Riemannian-)geometrically motivated derivation of the generalized Bures distance. Since the generalized fidelity features in the definition of the generalized Bures distance, this also provides a geometric interpretation for generalized fidelity.

Let $(\mathcal{M}, \mathfrak{g})$ be a Riemannian manifold. It is a defining property of Riemannian manifolds [DF92] that for any $x \in \mathcal{M}$, there is a neighborhood $\mathcal{N}_x \subseteq \mathcal{M}$ such that for any $y \in \mathcal{N}_x$, we have the natural distance between the points to be

$$\mathrm{d}^2(x, y) = ||\mathrm{Log}_x[y]||_x^2, \tag{4.17}$$

where $\mathrm{Log}_x[y] \in \mathrm{T}_x\mathcal{M}$ and where the norm is taken with respect to the inner product $\mathfrak{g}_x$ on the tangent space $\mathrm{T}_x\mathcal{M}$. It turns out that for the Bures–Wasserstein manifold, $\mathcal{N}_P = \mathbb{P}_d^+$ for any $P \in \mathbb{P}_d^+$. Thus, we can write

$$\mathrm{b}^{\mathrm{U}}(P, Q) = \|\mathrm{Log}_P[Q]\|_P = \|\mathrm{Log}_Q[P]\|_Q, \tag{4.18}$$

for any $P, Q \in \mathbb{P}_d^+$ where the norm is taken with respect to the Bures metric tensor in the corresponding tangent space.

One can visualize this process as follows. We linearize the *curved* Bures manifold $\mathbb{P}_d^+$ *about* the point P. This is equivalent to replacing every point $S \in \mathbb{P}_d^+$ with its image under the Riemannian log map $\mathrm{Log}_P[S] \in \mathrm{T}_P\mathbb{P}_d^+$. Thanks to the metric tensor, the tangent space has a natural distance between two tangent vectors $X, Y \in \mathrm{T}_P\mathbb{P}_d^+$ as $\|X - Y\|_P^2$. Therefore the squared distance between the tangent vectors $\mathrm{Log}_P[P]$ and $\mathrm{Log}_P[Q]$ is given by

$$\|\mathrm{Log}_P[P] - \mathrm{Log}_P[Q]\|_P^2 = \|\mathrm{Log}_P[Q]\|_P^2 = \mathrm{B}^{\mathrm{U}}(P, Q), \tag{4.19}$$

where we note that $\mathrm{Log}_S[S] = 0$ for any $S \in \mathbb{P}_d^+$.

To obtain generalized Bures distance and thereby generalized fidelity, we flatten our manifold not at P or Q, but at an arbitrary third point $R \in \mathbb{P}_d^+$. That is, we map the points $P, Q \in \mathbb{P}_d^+$ to the tangent space $\mathrm{T}_R\mathbb{P}_d^+$ via the Log_R map. We then compute the natural distance between these tangent vectors, which turns out to be exactly the generalized Bures distance between P and Q at R, as formalized in the following theorem.

Theorem 4.1.3. *Let $P, Q, R \in \mathbb{P}_d^+$ be chosen arbitrarily. Then*

$$\mathrm{B}_R(P,Q) = \|\mathrm{Log}_R[P] - \mathrm{Log}_R[Q]\|_R^2. \tag{4.20}$$

Proof. We will show that the RHS evaluates to the definition of the squared generalized Bures distance. For brevity, we will use the shorthand $P_R \equiv \mathrm{Log}_R[P]$ and $Q_R \equiv \mathrm{Log}_R[Q]$. Expanding the RHS, we have

$$\begin{aligned} \|P_R - Q_R\|_R^2 &= \langle P_R - Q_R, P_R - Q_R\rangle_R \\ &= \langle P_R, P_R\rangle_R + \langle Q_R, Q_R\rangle_R - \langle P_R, Q_R\rangle_R - \langle Q_R, P_R\rangle_R, \end{aligned} \tag{4.21}$$

where we used the bilinearity of the inner product to get the second line. Let us now analyze a single term, namely $\langle P_R, Q_R\rangle_R$, of the second line. By the definition of the Bures–Wasserstein metric (Eq. (2.76)), we have

$$\begin{aligned} \langle P_R, Q_R\rangle_R &:= \mathrm{Re}\,\mathrm{Tr}\left[\mathcal{L}_R(P_R)\cdot R\cdot \mathcal{L}_R(Q_R)\right] \\ &= \mathrm{Re}\,\mathrm{Tr}[\mathcal{L}_R(\mathrm{Log}_R[P])\cdot R\cdot \mathcal{L}_R(\mathrm{Log}_R[Q])]. \end{aligned} \tag{4.22}$$

Now recall that the definition of the BW Log_R map is $\mathrm{Log}_R[Q] := \mathcal{L}_R^{-1}(P^{-1}\#Q - \mathbb{I})$, and thus we have

$$\begin{aligned} \langle P_R, Q_R\rangle_R &= \mathrm{Re}\,\mathrm{Tr}\left[\mathcal{L}_R(\mathcal{L}_R^{-1}(R^{-1}\#P - \mathbb{I}))\cdot R\cdot \mathcal{L}_R(\mathcal{L}_R^{-1}(R^{-1}\#Q - \mathbb{I}))\right] \\ &= \mathrm{Re}\,\mathrm{Tr}[(R^{-1}\#P - \mathbb{I})R(R^{-1}\#Q - \mathbb{I})] \\ &= \mathrm{Re}\,\mathrm{Tr}[(R^{-1}\#P)R(R^{-1}\#Q)] - \mathrm{Tr}[(R^{-1}\#P)R] - \mathrm{Tr}[(R^{-1}\#Q)R] + \mathrm{Tr}[R] \\ &= \mathrm{Re}\,\mathrm{F}_R(Q,P) - \mathrm{F}^{\mathrm{U}}(P,R) - \mathrm{F}^{\mathrm{U}}(Q,R) + \mathrm{Tr}[R], \end{aligned} \tag{4.23}$$

where we used the fact that $\mathrm{Tr}[A(A^{-1}\#B)] = \mathrm{Tr}\left[\sqrt{A^{\frac{1}{2}}BA^{\frac{1}{2}}}\right] = \mathrm{F}^{\mathrm{U}}(A,B)$. We thus have

$$\begin{aligned} \|P_R\|_R^2 &= \mathrm{Tr}[P+R] - 2\,\mathrm{F}^{\mathrm{U}}(P,R) = \mathrm{B}^{\mathrm{U}}(P,R), \\ \|Q_R\|_R^2 &= \mathrm{Tr}[Q+R] - 2\,\mathrm{F}^{\mathrm{U}}(Q,R) = \mathrm{B}^{\mathrm{U}}(Q,R). \end{aligned} \tag{4.24}$$

Substituting all of these in Eq. (4.21), we get

$$\begin{aligned} \|P_R - Q_R\|_R^2 &= \mathrm{B}^{\mathrm{U}}(P,R) + \mathrm{B}^{\mathrm{U}}(Q,R) - 2\left(\mathrm{Tr}[R] - \mathrm{F}^{\mathrm{U}}(P,R) - \mathrm{F}^{\mathrm{U}}(Q,R) + \mathrm{Re}\,\mathrm{F}_R(P,Q)\right) \\ &= \mathrm{Tr}[P+Q] - 2\,\mathrm{Re}\,\mathrm{F}_R(P,Q) =: \mathrm{B}_R(P,Q). \end{aligned} \tag{4.25}$$

This concludes the proof. □

Thus, the geometric interpretation of generalized Bures distance is clear. It is the distance between the points P and Q if we *linearize* the manifold at R. Analogously, the generalized fidelity is the *fidelity part* of the generalized Bures

distance in this scenario. Thus, the various named fidelities have the following geometric interpretations.

- Uhlmann fidelity is the generalized fidelity if we linearize the manifold at P or Q (among other points).
- Holevo fidelity is the generalized fidelity if we linearize the manifold at $\mathbb{I}$.
- Matsumoto fidelity is the generalized fidelity if we linearize the manifold at P^{-1} or Q^{-1} (among other points).

The *other points* are discussed in Section 4.2. We now discuss some other properties of the generalized Bures distance.

Further properties of generalized Bures distance

We begin with an alternative form of generalized Bures distance.

Proposition 4.1.4. *Let $P, Q, R \in \mathbb{P}_d^+$. Then, the squared generalized Bures distance has the form*

$$\mathrm{B}_R(P,Q) = \left\| U_P^* P^{\frac{1}{2}} - U_Q^* Q^{\frac{1}{2}} \right\|_2^2 = \left\| P^{\frac{1}{2}} U_P - Q^{\frac{1}{2}} U_Q \right\|_2^2, \tag{4.26}$$

where $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}} R^{\frac{1}{2}}\right)$, $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}} R^{\frac{1}{2}}\right)$, and $\|A\|_2 := \sqrt{\mathrm{Tr}[A^* A]}$ denotes the Frobenius norm of a matrix A.

Proof. We have

$$\begin{aligned} \left\| U_P^* P^{\frac{1}{2}} - U_Q^* Q^{\frac{1}{2}} \right\|_2^2 &= \mathrm{Tr}\left[(U_P^* P^{\frac{1}{2}} - U_Q^* Q^{\frac{1}{2}})^* (U_P P^{\frac{1}{2}} - U_Q Q^{\frac{1}{2}}) \right] \\ &= \mathrm{Tr}\left[P + Q - P^{\frac{1}{2}} U_P U_Q^* Q^{\frac{1}{2}} + Q^{\frac{1}{2}} U_Q U_P^* P^{\frac{1}{2}} \right] \\ &= \mathrm{Tr}[P+Q] - 2\,\mathrm{Re}\,\mathrm{F}_R(P,Q) =: \mathrm{B}_R(P,Q). \end{aligned} \tag{4.27}$$

The second inequality of Eq. (4.26) follows from the adjoint invariance of the Frobenius norm. This concludes the proof. □

We now show that generalized Bures distance is a bona fide distance. A distance function d must satisfy three properties: symmetry, nonnegativity, and triangle inequality. See Section 2.1.2 for the definition of these properties. We now show that the generalized Bures distance satisfies the above three properties.

Theorem 4.1.5. *The generalized Bures distance* $\mathrm{b}_R(\cdot,\cdot)$ *at any* $R \in \mathbb{P}_d^+$ *satisfies*

1. *Symmetry:* $\mathrm{b}_R(P,Q) = \mathrm{b}_R(Q,P)$,
2. *Nonnegativity:* $\mathrm{b}_R(P,Q) \geq 0$ *with equality if and only if* $P = Q$.
3. *Triangle inequality: For any triple* $P, Q, S \in \mathbb{P}_d^+$, *we have*

$$\mathrm{b}_R(P,Q) \leq \mathrm{b}_R(P,S) + \mathrm{b}_R(S,Q). \tag{4.28}$$

Proof. The symmetry part follows directly from the definition. For the nonnegativity part, either observe that the generalized Bures distance is a norm in the tangent space $\mathrm{T}_R\mathbb{P}_d^+$ or observe that by Proposition 4.1.4, we have

$$\mathrm{b}_R(P,Q) = \left\| U_P^* P^{\frac{1}{2}} - U_Q^* Q^{\frac{1}{2}} \right\|_2 \geq 0, \tag{4.29}$$

where $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}} R^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}} R^{\frac{1}{2}}\right)$. We now show that equality is achieved if and only if $P = Q$. The direction $P = Q \implies \mathrm{b}_R(P,Q) = 0$ follows trivially. In the other direction, we want to prove

$$\mathrm{b}_R(P,Q) = 0 \implies P = Q. \tag{4.30}$$

To this end, observe that

$$\mathrm{b}_R(P,Q) = \left\| U_P^* P^{\frac{1}{2}} - U_Q^* Q^{\frac{1}{2}} \right\|_2 = 0 \quad \implies \quad U_P^* P^{\frac{1}{2}} = U_Q^* Q^{\frac{1}{2}}. \tag{4.31}$$

Right multiply both sides by $R^{\frac{1}{2}}$ to obtain

$$U_P^* P^{\frac{1}{2}} R^{\frac{1}{2}} = U_Q^* Q^{\frac{1}{2}} R^{\frac{1}{2}}. \tag{4.32}$$

By Lemma A.2.1, we have that the above equality is the same as

$$\sqrt{R^{\frac{1}{2}} P R^{\frac{1}{2}}} = \sqrt{R^{\frac{1}{2}} Q R^{\frac{1}{2}}}. \tag{4.33}$$

Squaring both sides and right-left multiplying by $R^{-\frac{1}{2}}$, we get $P = Q$.

To prove that the generalized Bures distance satisfies the triangle inequality,

choose arbitrary $P, Q, R, S \in \mathbb{P}_d^+$. We then have,

$$\begin{aligned} \mathrm{b}_R(P,Q) = \left\| U_P^* P^{\frac{1}{2}} - U_Q^* Q^{\frac{1}{2}} \right\|_2 &= \left\| U_P^* P^{\frac{1}{2}} - U_S^* S^{\frac{1}{2}} + U_S^* S^{\frac{1}{2}} + U_Q^* Q^{\frac{1}{2}} \right\|_2 \\ &\leq \left\| U_P^* P^{\frac{1}{2}} - U_S^* S^{\frac{1}{2}} \right\|_2 + \left\| U_S^* S^{\frac{1}{2}} - U_Q^* Q^{\frac{1}{2}} \right\|_2 \\ &= \mathrm{b}_R(P,S) + \mathrm{b}_R(S,Q). \end{aligned} \tag{4.34}$$

Here $U_S := \mathrm{Pol}\left(R^{\frac{1}{2}} S^{\frac{1}{2}}\right)$ and we have used triangle inequality for the Frobenius norm. □

We now mention a different way of showing triangle inequality and the saturation of the nonnegativity by invoking the tangent space $\mathrm{T}_R\mathbb{P}_d^+$. By Theorem 4.1.3, we have

$$\mathrm{b}_R(P,Q) = \left\| \mathrm{Log}_R[P] - \mathrm{Log}_R[Q] \right\|_R = 0 \quad \Longrightarrow \quad \mathrm{Log}_R[P] = \mathrm{Log}_R[Q]. \tag{4.35}$$

Now recall that the map Log_R is a diffeomorphism from $\mathcal{M}$ to its image and, therefore, is invertible, which necessarily means $P = Q$. For the triangle inequality we have that for any quadruple $P, Q, R, S \in \mathbb{P}_d^+$, we have

$$\begin{aligned} \mathrm{b}_R(P,Q) &= \left\| \mathrm{Log}_R[P] - \mathrm{Log}_R[Q] \right\|_R \\ &= \left\| \mathrm{Log}_R[P] - \mathrm{Log}_R[S] + \mathrm{Log}_R[S] - \mathrm{Log}_R[Q] \right\|_R \\ &\leq \left\| \mathrm{Log}_R[P] - \mathrm{Log}_R[S] \right\|_R + \left\| \mathrm{Log}_R[S] - \mathrm{Log}_R[Q] \right\|_R \\ &= \mathrm{b}_R(P,S) + \mathrm{b}_R(Q,S), \end{aligned} \tag{4.36}$$

where the inequality comes from the fact that the tangent space is an inner-product space, and therefore, triangle inequality holds.

4.2 Geometric properties of generalized fidelity

We now arrive at the central geometric results of this chapter. We discuss the geometric properties of the generalized fidelity $\mathrm{F}_R(P,Q)$ for fixed $P, Q \in \mathbb{P}_d^+$ and variable R. In particular, we study the change (or lack thereof) in generalized fidelity $\mathrm{F}_R(P,Q)$ as R varies along certain curves. Each curve discussed is visualized in Figure 4.2. In all the below cases, we choose $t \in [0,1]$. The curves we discuss in $\mathbb{P}_d^+$ are related to three Riemannian metrics: Bures–Wasserstein, Affine-invariant, and Euclidean. We concisely list important aspects of each metric in Table 4.1.

Each path we study falls into two categories: geodesic paths (represented as straight lines in Figure 4.2) and inverse of geodesic paths (represented as curved

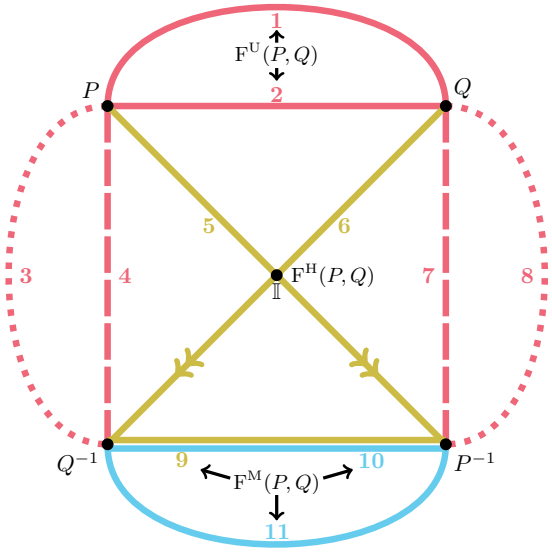

Figure 4.2: Generalized fidelity $\mathrm{F}_R(P,Q)$ at different base points $R \in \mathbb{P}_d$ for fixed $P, Q \in \mathbb{P}_d$. Red, yellow, and blue curves indicate that the associated metric is Bures–Wasserstein (BW), Affine-invariant (AI), and Euclidean, respectively. Straight and curved paths indicate that R and R^{-1} are elements of geodesics in their respective metrics. Solid lines indicate that the generalized fidelity is real-valued along these paths, while non-solid lines indicate complex values. Arrows on paths 5 and 6 indicate a conjectured monotonic decrease in $\mathrm{F}_R(P,Q)$ as R moves in the direction of the arrow. Details discussed in Section 4.2.

lines in Figure 4.2). By 'inverse of geodesic paths' we mean that the inverse R^{-1} of the base R moves along some geodesic path. We now discuss the properties of generalized fidelity as the base R moves along each specified path, grouped by the metric to which each path is related. For better readability, we defer all proofs of this section to Appendix C.3.

We represent a geodesic path with respect to the Riemannian metric 'RM' between points $A, B \in \mathbb{P}_d^+$ as

$$\gamma_{AB}^{\mathrm{RM}} : [0,1] \to \mathbb{P}_d^+, \tag{4.37}$$

with $\gamma_{AB}^{\mathrm{RM}}(0).2 = A$ and $\gamma_{AB}^{\mathrm{RM}}(1) = B$.

4.2.1 Bures–Wasserstein (red) paths

We first discuss the paths that are related to the BW metric, which are colored red in Figure 4.2. We consider six paths related to this metric. One defining property of the base R as it travels along these paths is that either R or R^{-1} is an element of a particular BW geodesic. Recall that the BW geodesic between

Object	**BW metric**	**AI metric**	**Euc. metric**
Inner product: $\langle X, Y\rangle_P$	$\mathrm{Re}\,\mathrm{Tr}[\mathcal{L}_P(X)P\mathcal{L}_P(Y)]$	$\mathrm{Tr}\left[P^{-1}XP^{-1}Y\right]$	$\mathrm{Tr}[XY]$
Exp. map: $\mathrm{Exp}_P[X]$	$P \star (\mathbb{I} + \mathcal{L}_P(X))^2$	$P^{\frac{1}{2}} \exp\left(\frac{X}{P}\right) P^{\frac{1}{2}}$	$P + X$
Log map: $\mathrm{Log}_P[Q]$	$\mathcal{L}_P^{-1}((P^{-1}\#Q) - \mathbb{I})$	$P^{\frac{1}{2}} \log\left(\frac{Q}{P}\right) P^{\frac{1}{2}}$	$Q - P$
Geodesic: $\gamma_{PQ}(t)$	$P \star [(1-t)\mathbb{I} + tS]^2$	$P^{\frac{1}{2}} \left(\frac{Q}{P}\right)^t P^{\frac{1}{2}}$	$(1-t)P + tQ$
Squared distance: $\mathrm{dist}^2(P, Q)$	$\mathrm{Tr}[P+Q] - 2\,\mathrm{F}^{\mathrm{U}}(P, Q)$	$\left\|\log\left(\frac{P}{Q}\right)\right\|_2^2$	$\|P - Q\|_2^2$

Table 4.1: Different Riemannian metrics on $\mathbb{P}_d^+$ and associated objects. Here $S := P^{-1}\#Q$, $\mathcal{L}_P(X)$ is the unique solution to the equation $\mathcal{L}_P(X)P + P\mathcal{L}_P(X) = X$ and $\mathcal{L}_P^{-1}(Y) := YP + PY$. The star product is defined as $A \star B := B^{\frac{1}{2}}AB^{\frac{1}{2}}$ for $A, B \geq 0$.

Path	**Path definition**	**Property of generalized fidelity** $\mathrm{F}_R(P, Q)$
1	$R = \left[\gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t)\right]^{-1}$	Invariant. $\mathrm{F}_R(P, Q) = \mathrm{F}^{\mathrm{U}}(P, Q)$ for all $t \in [0, 1]$.
2	$R = \gamma^{\mathrm{BW}}_{PQ}(t)$	Invariant. $\mathrm{F}_R(P, Q) = \mathrm{F}^{\mathrm{U}}(P, Q)$ for all $t \in [0, 1]$.
3	$R = \left[\gamma^{\mathrm{BW}}_{P^{-1}Q}(t)\right]^{-1}$	Complex-valued and covaries with Path 8.
4	$R = \gamma^{\mathrm{BW}}_{PQ^{-1}}(t)$	Complex-valued and covaries with Path 7.
5	$R = \gamma^{\mathrm{AI}}_{PP^{-1}}(t)$	Recovers Uhlmann-, Holevo-, and Matsumoto-fidelity.
6	$R = \gamma^{\mathrm{AI}}_{QQ^{-1}}(t)$	Recovers Uhlmann-, Holevo-, and Matsumoto-fidelity.
7	$R = \gamma^{\mathrm{BW}}_{QP^{-1}}(t)$	Complex-valued and covaries with Path 4.
8	$R = \left[\gamma^{\mathrm{BW}}_{Q^{-1}P}(t)\right]^{-1}$	Complex-valued and covaries with Path 3.
9	$R = \gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t)$	Invariant. $\mathrm{F}_R(P, Q) = \mathrm{F}^{\mathrm{M}}(P, Q)$ for all $t \in [0, 1]$.
10	$R = \gamma^{\mathrm{Euc}}_{P^{-1}Q^{-1}}(t)$	Invariant. $\mathrm{F}_R(P, Q) = \mathrm{F}^{\mathrm{M}}(P, Q)$ for all $t \in [0, 1]$.
11	$R = \left[\gamma^{\mathrm{Euc}}_{PQ}(t)\right]^{-1}$	Invariant. $\mathrm{F}_R(P, Q) = \mathrm{F}^{\mathrm{M}}(P, Q)$ for all $t \in [0, 1]$.

Figure 4.3: Definitions and properties of the paths discussed in Figure 4.2.

$A, B \in \mathbb{P}_d^+$ has the form

$$C \equiv \gamma^{\mathrm{BW}}_{AB}(t) = [(1-t)\mathbb{I} + tA^{-1}\#B]A[(1-t)\mathbb{I} + tA^{-1}\#B]. \tag{4.38}$$

Moreover, being an element of the BW geodesic (at time t) is equivalent to being the BW barycenter of a 2-point distribution with weights $(1-t, t)$. As such, it uniquely satisfies the following fixed-point equation.

$$C = (1-t)\sqrt{C^{\frac{1}{2}} A C^{\frac{1}{2}}} + t\sqrt{B^{\frac{1}{2}} A B^{\frac{1}{2}}}. \tag{4.39}$$

These two properties will be exploited in the proofs.

Path 2: $R = \gamma^{\mathrm{BW}}_{PQ}(t)$

Here the base R moves along the BW geodesic path between P and Q:

$$R = \gamma^{\mathrm{BW}}_{PQ}(t) = [(1-t)\mathbb{I} + tS]\, P\, [(1-t)\mathbb{I} + tS]\,, \quad \text{for } t \in [0,1], \tag{4.40}$$

where $S = P^{-1}\#Q$. As shown in the following theorem, $\mathrm{F}_R(P,Q)$ is constant and equal to Uhlmann fidelity along this curve.

Theorem 4.2.1. *Let $P, Q \in \mathbb{P}_d^+$ be fixed. Let the base $R = \gamma^{\mathrm{BW}}_{PQ}(t)$ for any $t \in [0,1]$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{U}}(P,Q). \tag{4.41}$$

Proof. See Theorem C.3.1. □

Thus, as the base R varies along the BW geodesic $\gamma^{\mathrm{BW}}_{PQ}$ between P and Q, the generalized fidelity is invariant and equal to the Uhlmann fidelity between P and Q. This is the first example where the generalized fidelity is real-valued for a non-trivial base ($R \notin \{P^x, Q^y\}$). We next look at another path that exhibits the same property.

Path 1: $R = [\gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t)]^{-1}$

Each point on this path is the inverse of a point on the BW geodesic between P^{-1} and Q^{-1}:

$$R = [\gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t)]^{-1} = [(1-t)\mathbb{I} + tS]^{-1}\, P\, [(1-t)\mathbb{I} + tS]^{-1}\,, \quad \text{for any } t \in [0,1], \tag{4.42}$$

where $S = Q^{-1}\#P$. As elaborated in the following theorem, the generalized fidelity $\mathrm{F}_R(P,Q)$ is constant and equal to Uhlmann fidelity along this curve.

Theorem 4.2.2. *Let $P, Q \in \mathbb{P}_d^+$ be fixed. Let the base R be any point along the curve $[\gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t)]^{-1}$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{U}}(P,Q). \tag{4.43}$$

Proof. See Theorem C.3.2. □

Paths 4 and 7: $R_1 = \gamma^{\mathrm{BW}}_{PQ^{-1}}(t)$ and $R_2 = \gamma^{\mathrm{BW}}_{QP^{-1}}(t)$

These are the BW geodesic paths from P to Q^{-1} and from Q to P^{-1}:

$$R_1 = \gamma^{\mathrm{BW}}_{PQ^{-1}}(t) = \left[(1-t)\mathbb{I} + tP^{-1}\#Q^{-1}\right] P \left[(1-t)\mathbb{I} + tP^{-1}\#Q^{-1}\right], \tag{4.44}$$

and

$$R_2 = \gamma^{\mathrm{BW}}_{QP^{-1}}(t) = \left[(1-t)\mathbb{I} + tP^{-1}\#Q^{-1}\right] Q \left[(1-t)\mathbb{I} + tP^{-1}\#Q^{-1}\right], \tag{4.45}$$

The generalized fidelity along these paths is complex-valued. We currently do not know any special closed-form for $\mathrm{F}_R(P,Q)$ when the base moves along either of the paths. However, we can show that the generalized fidelity varies in a *covariant* manner across the two paths.

Theorem 4.2.3. *Let $P, Q \in \mathbb{P}_d$ be fixed. For any fixed $t \in [0,1]$, let*

$$R_1 := \gamma^{\mathrm{BW}}_{PQ^{-1}}(t) \quad \textit{and} \quad R_2 := \gamma^{\mathrm{BW}}_{QP^{-1}}(t). \tag{4.46}$$

Then $\mathrm{F}_{R_1}(P,Q) = \mathrm{F}_{R_2}(P,Q)$.

Proof. See Theorem C.3.3 for proof. □

Paths 3 and 8: $R_1 = [\gamma^{\mathrm{BW}}_{P^{-1}Q}(t)]^{-1}$ and $R_2 = [\gamma^{\mathrm{BW}}_{Q^{-1}P}(t)]^{-1}$

These paths correspond to the inverse of the BW geodesics from P^{-1} to Q and Q^{-1} to P:

$$R_1 = [\gamma^{\mathrm{BW}}_{P^{-1}Q}(t)]^{-1} \quad \text{and} \quad R_2 = [\gamma^{\mathrm{BW}}_{Q^{-1}P}(t)]^{-1}. \tag{4.47}$$

Similar to the previous result, generalized fidelity covaries along these paths. This is formalized in the following theorem.

Theorem 4.2.4. *Let $P, Q \in \mathbb{P}_d^+$ be fixed. For any fixed $t \in [0,1]$, let*

$$R_1 := [\gamma^{\mathrm{BW}}_{P^{-1}Q}(t)]^{-1} \quad \textit{and} \quad R_2 := [\gamma^{\mathrm{BW}}_{Q^{-1}P}(t)]^{-1}. \tag{4.48}$$

Then $\mathrm{F}_{R_1}(P,Q) = \mathrm{F}_{R_2}(P,Q)$.

Proof. See Theorem C.3.4 for proof. □

4.2.2 Affine-invariant (yellow) paths

We now discuss the geometric properties of generalized fidelity as the base moves along geodesic paths related to the Affine-invariant metric. Recall that the AI geodesic path between $A, B \in \mathbb{P}_d^+$ is defined as

$$\gamma^{\mathrm{AI}}_{AB}(t) = A^{\frac{1}{2}} \left(\frac{B}{A}\right)^t A^{\frac{1}{2}}. \tag{4.49}$$

We now discuss the geodesic properties of generalized fidelity $\mathrm{F}_R(P,Q)$ as the base R moves along three geodesic paths related to this metric.

Path 9: $R = \gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t) = \left[\gamma^{\mathrm{AI}}_{PQ}(t)\right]^{-1}$**.**

Here the base R moves along the AI-geodesic path between P^{-1} and Q^{-1}. A point R on this curve has the form

$$R = \gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t) := P^{-\frac{1}{2}} \left(P^{\frac{1}{2}} Q^{-1} P^{\frac{1}{2}}\right)^t P^{-\frac{1}{2}}. \tag{4.50}$$

Observe that for $t \in [0,1]$, we have$\gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t) = [\gamma^{\mathrm{AI}}_{PQ}(t)]^{-1}$. For $t = \frac{1}{2}$, we have $R = P^{-1}\#Q^{-1} = (P\#Q)^{-1}$. We first provide a short proof for the claim that when R is the midpoint ($t = \frac{1}{2}$) of the geodesic, the generalized fidelity equals the Matsumoto fidelity $\mathrm{F}^{\mathrm{M}}(P,Q)$. Subsequently, we also show that for any point on the AI geodesic between P^{-1} and Q^{-1}, the generalized fidelity is constantly equal to the Matsumoto fidelity.

Theorem 4.2.5. *Let $P, Q \in \mathbb{P}_d^+$ and choose $R = P^{-1}\#Q^{-1}$. Then* $\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q)$.

Proof. See Theorem C.3.5 for proof. □

We now show that the generalized fidelity constantly equals the Matsumoto fidelity along any point in the path $\gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t)$.

Theorem 4.2.6. *Let $P, Q \in \mathbb{P}_d^+$ and let $R = \gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t)$ for any $t \in [0,1]$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q). \tag{4.51}$$

Proof. See Theorem C.3.6 for proof. □

Paths 5 and 6: $R_1\gamma^{\mathrm{AI}}_{PP^{-1}}(t)$ and $R_2 = \gamma^{\mathrm{AI}}_{QQ^{-1}}(t)$.

These are the geodesic paths from P to P^{-1} and Q to Q^{-1} with respect to the AI metric.

$$\begin{aligned} R_1 = \gamma^{\mathrm{AI}}_{PP^{-1}}(t) &= P^{\frac{1}{2}}\left(P^{-\frac{1}{2}}P^{-1}P^{-\frac{1}{2}}\right)^t P^{\frac{1}{2}} = P^{1-2t} \equiv P^x, \\ R_2 = \gamma^{\mathrm{AI}}_{QQ^{-1}}(t) &= Q^{\frac{1}{2}}\left(Q^{-\frac{1}{2}}Q^{-1}Q^{-\frac{1}{2}}\right)^t Q^{\frac{1}{2}} = Q^{1-2t} \equiv Q^x, \end{aligned} \tag{4.52}$$

where we have denoted $x \equiv 1-2t$ for convenience. As R_1 and R_2 move along these geodesics, the generalized fidelity takes the form

$$\begin{aligned} \mathrm{F}_{R_1}(P,Q) &= \mathrm{Tr}\left[\sqrt{P^{\frac{x}{2}}PP^{\frac{x}{2}}}P^{-x}\sqrt{P^{\frac{x}{2}}QP^{\frac{x}{2}}}\right] \\ &= \mathrm{Tr}\left[P^{\frac{1-x}{2}}\sqrt{P^{\frac{x}{2}}QP^{\frac{x}{2}}}\right] = \mathrm{Tr}\left[P^{\frac{1-x}{2}}\#P^{\frac{1+x}{4}}QP^{\frac{1+x}{4}}\right]. \end{aligned} \tag{4.53}$$

A similar calculation shows

$$\mathrm{F}_{R_2}(P,Q) = \mathrm{Tr}\left[Q^{\frac{1-x}{2}}\sqrt{Q^{\frac{x}{2}}PQ^{\frac{x}{2}}}\right] = \mathrm{Tr}\left[Q^{\frac{1-x}{2}}\#Q^{\frac{1+x}{4}}PQ^{\frac{1+x}{4}}\right]. \tag{4.54}$$

These two paths are special because we recover all the three *named fidelities* along these paths:

$$\begin{aligned} t = 0 &\iff x = 1 \quad \implies \quad \mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{U}}(P,Q), \\ t = \tfrac{1}{2} &\iff x = 0 \quad \implies \quad \mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{H}}(P,Q), \\ t = 1 &\iff x = -1 \quad \implies \quad \mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q). \end{aligned} \tag{4.55}$$

Moreover, these are paths along which the generalized fidelity varies while being real-valued throughout. Thus, we have a *one parameter* family of fidelities that recover all the three named fidelities. The one-parameter nature becomes even more apparent with the alternative form of generalized fidelity along these paths, as described in the following theorem.

Theorem 4.2.7. *Let $P, Q \in \mathbb{P}_d^+$ and let $x \in \mathbb{R}$. Then*

$$\mathrm{F}_{P^x}(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}} U_x Q^{\frac{1}{2}}\right] \quad \textit{and} \quad \mathrm{F}_{Q^x}(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}} V_x Q^{\frac{1}{2}}\right], \tag{4.56}$$

where

$$U_x := \mathrm{Pol}\left(P^{\frac{x}{2}} Q^{\frac{1}{2}}\right) \quad \textit{and} \quad V_x := \mathrm{Pol}\left(P^{\frac{1}{2}} Q^{\frac{x}{2}}\right). \tag{4.57}$$

Proof. Let $U_x := \mathrm{Pol}\left(P^{\frac{x}{2}} Q^{\frac{1}{2}}\right)$, which implies $U_x^* = \mathrm{Pol}\left(Q^{\frac{1}{2}} P^{\frac{x}{2}}\right)$. Thus we have

$$Q^{\frac{1}{2}} P^{\frac{x}{2}} = U_x^* \sqrt{P^{\frac{x}{2}} Q P^{\frac{x}{2}}} \iff U_x = \sqrt{P^{\frac{x}{2}} Q P^{\frac{x}{2}}} P^{-\frac{x}{2}} Q^{-\frac{1}{2}}. \tag{4.58}$$

It then follows that

$$\begin{aligned} \mathrm{Tr}\left[P^{\frac{1}{2}} U_x Q^{\frac{1}{2}}\right] &= \mathrm{Tr}\left[P^{\frac{1}{2}} \sqrt{P^{\frac{x}{2}} Q P^{\frac{x}{2}}} P^{-\frac{x}{2}} Q^{-\frac{1}{2}} Q^{\frac{1}{2}}\right] \\ &= \mathrm{Tr}\left[P^{\frac{1-x}{2}} \sqrt{P^{\frac{x}{2}} Q P^{\frac{x}{2}}}\right] = \mathrm{F}_{P^x}(P,Q), \end{aligned} \tag{4.59}$$

which proves the first claim. Now let $V_x := \mathrm{Pol}\left(P^{\frac{1}{2}} Q^{\frac{x}{2}}\right)$, which implies $V_x^* = \mathrm{Pol}\left(Q^{\frac{x}{2}} P^{\frac{1}{2}}\right)$. Using a similar calculation as above, we have

$$V_x = P^{-\frac{1}{2}} Q^{-\frac{x}{2}} \sqrt{Q^{\frac{x}{2}} P Q^{\frac{x}{2}}}, \tag{4.60}$$

which implies

$$\begin{aligned} \mathrm{Tr}\left[P^{\frac{1}{2}} V_x Q^{\frac{1}{2}}\right] &= \mathrm{Tr}\left[P^{\frac{1}{2}} P^{-\frac{1}{2}} Q^{-\frac{x}{2}} \sqrt{Q^{\frac{x}{2}} P Q^{\frac{x}{2}}} Q^{\frac{1}{2}}\right] \\ &= \mathrm{Tr}\left[\sqrt{Q^{\frac{x}{2}} P Q^{\frac{x}{2}}} Q^{\frac{1-x}{2}}\right] = \mathrm{F}_{Q^x}(P,Q), \end{aligned} \tag{4.61}$$

as claimed. This completes the proof. □

Note that $\mathrm{F}_{P^x}(P,Q) \neq \mathrm{F}_{Q^x}(P,Q)$ for general values of x, except when $x \in \{1, 0, -1\}$, where they are equal. In terms of the geodesic paths $\gamma^{\mathrm{AI}}_{PP^{-1}}(t)$ and $\gamma^{\mathrm{AI}}_{QQ^{-1}}(t)$, this corresponds to the values of $t \in \{0, 1/2, 1\}$ respectively. Moreover, we see that generalized fidelity along these paths is not symmetric under the swap of P and Q, as the bases themselves depend on P and Q. However, we can construct a symmetrized version by taking their average for the same value of x:

$$\overline{\mathrm{F}}_x(P,Q) := \frac{\mathrm{F}_{P^x}(P,Q) + \mathrm{F}_{Q^x}(P,Q)}{2}. \tag{4.62}$$

We call this parametrized family of fidelities the *x-Polar fidelities* and remark that it is a one-parameter family of fidelities that recover Uhlmann-, Holevo-, and Matsumoto-fidelity at $x = 1, 0, -1$ respectively. To our knowledge, this is the first such generalization. Further properties, including some remarkable connections to the spectral norm unit ball and the Lie group of $\mathrm{SU}(d)$, are discussed in Section 4.3.

An interesting numerical observation regarding generalized fidelity along these paths is as follows. For $x, y \in [0, 1]$ we observe that

$$x \geq y \implies \begin{cases} \mathrm{F}_{P^x}(P, Q) \geq \mathrm{F}_{P^y}(P, Q) \\ \mathrm{F}_{Q^x}(P, Q) \geq \mathrm{F}_{Q^y}(P, Q). \end{cases} \tag{4.63}$$

This observation is consistent with the known relations [Mat10; CS20] regarding the named fidelities:

$$\mathrm{F}^{\mathrm{M}}(P, Q) \leq \mathrm{F}^{\mathrm{H}}(P, Q) \leq \mathrm{F}^{\mathrm{U}}(P, Q), \tag{4.64}$$

for any $P, Q \geq 0$.

If this observation can be proven mathematically, it would imply that x-Polar fidelity is monotonic in $x \in [-1, 1]$, which would give us a monotonic family of fidelities that recover the three named fidelities.

4.2.3 Euclidean (blue) paths

Finally, we discuss paths related to the Euclidean metric, where a geodesic is obtained via convex combinations of the endpoints:

$$\gamma^{\mathrm{Euc}}_{AB}(t) = (1 - t)A + tB. \tag{4.65}$$

These paths are colored blue in Figure 4.2. We now discuss generalized fidelity along 2 paths related to this metric.

Path 10: $R = \gamma^{\mathbf{Euc}}_{P^{-1}Q^{-1}}(t)$**.**

We first study the Euclidean geodesic between P^{-1} and Q^{-1}, which is simply the convex combination of the end-points: $\gamma^{\mathrm{Euc}}_{P^{-1}Q^{-1}}(t) = (1 - t)P^{-1} + tQ^{-1}$. As the base R moves along this path, the generalized fidelity is constant and equal to the Matsumoto fidelity.

Theorem 4.2.8. *Let $P, Q \in \mathbb{P}_d^+$ and let $R = \gamma^{\mathrm{Euc}}_{P^{-1}Q^{-1}}(t)$ for any $t \in [0,1]$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q). \tag{4.66}$$

Proof. See Appendix C.3.7 for proof. □

Path 11: $R = [\gamma^{\mathbf{Euc}}_{PQ}(t)]^{-1}$

Conversely, we now consider the case where the base moves along the path defined by the inverse of the Euclidean geodesic (straight line) between P and Q:

$$R = \left[\gamma^{\mathrm{Euc}}_{PQ}(t)\right]^{-1} = \left[(1-t)P + tQ\right]^{-1}. \tag{4.67}$$

Like in the previous case, the generalized fidelity is constant along this path and equal to the Matsumoto fidelity.

Theorem 4.2.9. *Let $P, Q \in \mathbb{P}_d^+$ and let $R = [\gamma^{\mathrm{Euc}}_{PQ}(t)]^{-1}$ for any $t \in [0,1]$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q). \tag{4.68}$$

Proof. See Appendix C.3.8 for proof. □

4.3 Polar fidelities and Interior fidelities

In this section, we introduce a new family of fidelities and show how generalized fidelities can be thought of as elements of the extreme points of this family of fidelities. We begin with recalling one of the equivalent characterizations of generalized fidelity. For any $P, Q, R \in \mathbb{P}_d^+$, we have

$$\mathrm{F}_R(P,Q) = \mathrm{Tr}\left[Q^{\frac{1}{2}} U_Q U_P^* P^{\frac{1}{2}}\right], \tag{4.69}$$

where $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}}R^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}}R^{\frac{1}{2}}\right)$. We call the unitary $U_Q U_P^*$ the *unitary factor* of the generalized fidelity $\mathrm{F}_R(P,Q)$.

We first show that the unitary factor of any generalized fidelity has $+1$ determinant.

Proposition 4.3.1. *Let $P, Q, R \in \mathbb{P}_d^+$. Then the unitary factor $U_Q U_P^*$ of $\mathrm{F}_R(P,Q)$ has determinant $+1$. Or equivalently, $U_Q U_P^* \in \mathrm{SU}(d)$.*

Proof. The proof will be done in two steps. We first show that the polar factor of any matrix that is the product of two positive definite matrices is a special unitary

matrix. To this end, let $A, B \in \mathbb{P}_d^+$. We aim to show that $U := \mathrm{Pol}\,(AB) \in \mathrm{SU}(d)$. By definition, we have

$$AB = U|AB| = U\sqrt{BA^2B}. \tag{4.70}$$

Now take determinant across to get

$$\det(AB) = \det(U|AB|) = \det(U)\det(|AB|). \tag{4.71}$$

The LHS equals $\det(A)\det(B)$. By properties of determinant, we have

$$\det(|AB|) = \det\left(\sqrt{AB^2A}\right) = \sqrt{\det(AB^2A)} = \sqrt{\det(A)^2\det(B)^2} = \det(A)\det(B). \tag{4.72}$$

Thus we have $\det(U) = +1$, or $\mathrm{Pol}\,(AB) \in \mathrm{SU}(d)$ for any $A, B \in \mathbb{P}_d^+$. Thus we have $U_P, U_Q \in \mathrm{SU}(d)$, and by the group structure of special unitaries, we have $U_P U_Q^* \in \mathrm{SU}(d)$. □

Thus, the unitary factor of *any* generalized fidelity is a special unitary. Let us now recall the definition of x-Polar fidelity. For $P, Q \in \mathbb{P}_d^+$, we defined

$$\overline{\mathrm{F}}_x(P,Q) := \frac{\mathrm{F}_{P^x}(P,Q) + \mathrm{F}_{Q^x}(P,Q)}{2} = \mathrm{Tr}\left[P^{\frac{1}{2}}\left(\frac{U_x + V_x}{2}\right)Q^{\frac{1}{2}}\right], \tag{4.73}$$

where $U_x := \mathrm{Pol}\left(P^{\frac{x}{2}}Q^{\frac{1}{2}}\right)$ and $V_x := \mathrm{Pol}\left(P^{\frac{1}{2}}Q^{\frac{x}{2}}\right)$. Observe that Polar fidelities are *not* generalized fidelities (except in the cases of $x \in \{1, 0, -1\}$). In particular, convex combinations, over bases, of generalized fidelities are not, in general, generalized fidelities. We call such convex combinations of generalized fidelities over a distribution of bases as *interior fidelities*. We first formally define the notion of interior fidelities and then show why the name is appropriate.

Definition 4.3.2. *Let $\mu \in {}_n$ be a probability vector, $\mathcal{R} = \{R_1, \ldots, R_n\} \subset \mathbb{P}_d^+$ be a* collection of bases*, and let $P, Q \in \mathbb{P}_d^+$. Define the* interior fidelity *between P* and *Q* over *$(\mathcal{R}, \mu)$ as*

$$\mathrm{F}_{(\mathcal{R},\mu)}(P,Q) := \sum_{i=1}^{n} \mu_i \mathrm{F}_{R_i}(P,Q) = \sum_{i=1}^{n} \mu_i \mathrm{Tr}\left[Q^{\frac{1}{2}}V_iP^{\frac{1}{2}}\right] = \mathrm{Tr}\left[Q^{\frac{1}{2}}\overline{V}P^{\frac{1}{2}}\right], \tag{4.74}$$

where V_i is the unitary factor of $\mathrm{F}_{R_i}(P,Q)$ and $\overline{V} := \sum_{i=1}^n \mu_i V_i$ is the mean of the unitary factors.

One can also construct the corresponding *interior Bures distance*, and geometrically this would correspond to linearizing the manifold at different points

and then taking the average distance over these different linearizations. We also note that since any generalized fidelity reduces to classical fidelity in the commuting case, so do interior fidelities. Thus, interior fidelities are valid quantizations of classical fidelity.

We now note a different kind of geometric property of generalized fidelities and interior fidelities. Specifically, generalized fidelities can be seen to be the *extreme points* of a convex family of fidelities, whose interior points are constituted by interior fidelities. This correspondence is easier to see once we recall that the set of all unitaries from extreme points of the convex and compact *spectral-norm unit ball* [Wat18, Theorem 1.10]:

$$\mathcal{B}_\infty := \{K : \|K\|_\infty \leq 1\}. \tag{4.75}$$

It then follows from the definition of interior fidelities that the 'non-commutative part' is an interior point of $\mathcal{B}_\infty$. However, we note that the correspondence is not bijective as $\mathcal{B}_\infty$ also contains unitaries that are not special unitaries, whereas the unitary factor of any generalized fidelity is necessarily a special unitary.

Thus, we see that the x-Polar fidelity is an interior fidelity. We showed that the unitary factor of any generalized fidelity is a special unitary. An open question we pose is whether the converse is true. That is, for any $U \in \mathrm{SU}(d)$, does there exist a triple $P, Q, R \in \mathbb{P}_d^+$ such that $U = U_Q U_P^*$? In [LL08, Remark 5.2], it was shown that $\mathrm{SU}(d)$ can be generated by the set of polar factors of the product of two positive definite matrices. However, for our open question to be answered affirmatively, we would require that any special unitary can be written by the product of at most two polar factors of the above kind. It is unclear whether this is true. If the question can be answered affirmatively, then it would shed light on a new connection between polar decomposition, the Bures–Wasserstein manifold of positive definite matrices, and the Lie group $\mathrm{SU}(d)$.

4.4 Block matrix characterization of generalized fidelity

We now show that the generalized fidelity, and thereby the generalized Bures distance, has a Block-matrix representation. This representation is intimately related to the semidefinite program (SDP) for Uhlmann fidelity [Wat18] and its generalization to SDP for optimal average fidelity [AKF22]. We will first discuss the SDP for average fidelity (which, in this context, is an SDP for *total fidelity*) and then show how one can recover generalized fidelity and generalized Bures distance from it.

Proposition 4.4.1. *Let $P, Q, R \in \mathbb{P}_d^+$ be arbitrarily chosen. Define the matrices A, B as*

$$A := \frac{1}{2}\begin{pmatrix} 0 & 0 & \mathbb{I} \\ 0 & 0 & \mathbb{I} \\ \mathbb{I} & \mathbb{I} & 0 \end{pmatrix} \in \mathbb{H}_{3d}, \quad B := \begin{pmatrix} P & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{pmatrix} \in \mathbb{P}_{3d}, \tag{4.76}$$

and $\Phi : \mathbb{H}_{3d} \to \mathbb{H}_{3d}$ to be the Hermitian preserving map which acts as

$$\begin{pmatrix} M_{11} & \cdot & \cdot \\ \cdot & M_{22} & \cdot \\ \cdot & \cdot & M_{33} \end{pmatrix} \overset{\Phi}{\mapsto} \begin{pmatrix} M_{11} & 0 & 0 \\ 0 & M_{22} & 0 \\ 0 & 0 & M_{33} \end{pmatrix}. \tag{4.77}$$

Consider the SDP (Φ, A, B) whose primal problem is

$$\begin{aligned} \text{maximize}: \quad & \langle A, \mathrm{X} \rangle, \\ \text{subject to}: \quad & \mathrm{X} \geq 0, \quad \Phi(\mathrm{X}) = B. \end{aligned} \tag{4.78}$$

The optimal value of this SDP is $\mathrm{F}^{\mathrm{U}}(P, R) + \mathrm{F}^{\mathrm{U}}(Q, R)$, and the unique optimal feasible point that attains this value is of the form

$$\mathrm{X}_\star := \begin{pmatrix} P & P^{\frac{1}{2}} U_P U_Q^* Q^{\frac{1}{2}} & P^{\frac{1}{2}} U_P R^{\frac{1}{2}} \\ Q^{\frac{1}{2}} U_Q U_P^* P^{\frac{1}{2}} & Q & Q^{\frac{1}{2}} U_Q R^{\frac{1}{2}} \\ R^{\frac{1}{2}} U_P^* P^{\frac{1}{2}} & R^{\frac{1}{2}} U_Q^* Q^{\frac{1}{2}} & R \end{pmatrix} \tag{4.79}$$

where $U_P := \mathrm{Pol}\left(R^{\frac{1}{2}} P^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(R^{\frac{1}{2}} Q^{\frac{1}{2}}\right)$.

Proof. The proof is a straightforward application of results in [Wat18] and Chapter 3. The fact that the value of the objective function cannot exceed $\mathrm{F}^{\mathrm{U}}(P, R) + \mathrm{F}^{\mathrm{U}}(Q, R)$ is shown in Lemma 3.3.2. That is, for any feasible point X,

$$\langle A, \mathrm{X} \rangle \leq \mathrm{F}^{\mathrm{U}}(P, R) + \mathrm{F}^{\mathrm{U}}(Q, R)\,. \tag{4.80}$$

To show that the inequality is saturated, we construct a feasible point that attains the optimal value $\mathrm{F}^{\mathrm{U}}(P, R) + \mathrm{F}^{\mathrm{U}}(Q, R)$. Later, we will show that it is also the unique optimal feasible point. For $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}} R^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}} R^{\frac{1}{2}}\right)$, define the $3d \times d$ block matrix T as follows.

$$T := \begin{pmatrix} P^{\frac{1}{2}} U_P \\ Q^{\frac{1}{2}} U_Q \\ R^{\frac{1}{2}} \end{pmatrix} \tag{4.81}$$

Now consider $\mathrm{X}_\star := TT^*$:

$$\begin{aligned}\mathrm{X}_\star := TT^* &= \begin{pmatrix} P^{\frac{1}{2}}U_P \\ Q^{\frac{1}{2}}U_Q \\ R^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} U_P^* P^{\frac{1}{2}} & U_Q^* Q^{\frac{1}{2}} & R^{\frac{1}{2}} \end{pmatrix} \\ &= \begin{pmatrix} P & P^{\frac{1}{2}}U_P U_Q^* Q^{\frac{1}{2}} & P^{\frac{1}{2}}U_P R^{\frac{1}{2}} \\ Q^{\frac{1}{2}}U_Q U_P^* P^{\frac{1}{2}} & Q & Q^{\frac{1}{2}}U_Q R^{\frac{1}{2}} \\ R^{\frac{1}{2}}U_P^* P^{\frac{1}{2}} & R^{\frac{1}{2}}U_Q^* Q^{\frac{1}{2}} & R \end{pmatrix} \equiv \begin{pmatrix} P & Z & Y_P \\ Z^* & Q & Y_Q \\ Y_P^* & Y_Q^* & R \end{pmatrix} \geq 0.\end{aligned} \tag{4.82}$$

In the last equality, we renamed some quantities for brevity. We then see that

$$\langle A, \mathrm{X}_\star \rangle = \mathrm{Re}\,\mathrm{Tr}[Y_P] + \mathrm{Re}\,\mathrm{Tr}[Y_Q] = \mathrm{F}^{\mathrm{U}}(P,R) + \mathrm{F}^{\mathrm{U}}(Q,R)\,, \tag{4.83}$$

which implies that the SDP achieves optimality.

We now prove the uniqueness of the optimal feasible point. We first note that the positivity of $\mathrm{X}_\star$ necessarily implies the positivity of the principal submatrices:

$$M_P := \begin{pmatrix} P & Y_P \\ Y_P^* & R \end{pmatrix} \geq 0 \quad \text{and} \quad M_Q := \begin{pmatrix} Q & Y_Q \\ Y_Q^* & R \end{pmatrix} \geq 0. \tag{4.84}$$

By [Wat18, Lemma 3.18], we have $M_P \geq 0$ if and only if $Y_P = P^{\frac{1}{2}} K R^{\frac{1}{2}}$ for some contraction $K : \|K\|_\infty \leq 1$. Note that $\mathrm{Re}\,\mathrm{Tr}[Y_P]$ attains the maximum necessarily for some unitary K, which is an extreme point of the set of contractions. To see this, observe that

$$\mathrm{Re}\,\mathrm{Tr}[Y_P] = \mathrm{Re}\,\mathrm{Tr}\left[P^{\frac{1}{2}} K R^{\frac{1}{2}}\right] = \mathrm{Re}\,\mathrm{Tr}\left[K R^{\frac{1}{2}} P^{\frac{1}{2}}\right] \leq \left\| K R^{\frac{1}{2}} P^{\frac{1}{2}} \right\|_1, \tag{4.85}$$

where the last equality comes from the variational characterization of trace norm. Now note that for any contraction K we have $K^*K \leq \mathbb{I}$ and thus

$$P^{\frac{1}{2}} R^{\frac{1}{2}} K^* K R^{\frac{1}{2}} P^{\frac{1}{2}} \leq P^{\frac{1}{2}} R P^{\frac{1}{2}}, \tag{4.86}$$

with equality if and only if K is a unitary. Since the operator square root function over positive semidefinite matrices is monotonic, we have

$$\sqrt{P^{\frac{1}{2}} R^{\frac{1}{2}} K^* K R^{\frac{1}{2}} P^{\frac{1}{2}}} \leq \sqrt{P^{\frac{1}{2}} R P^{\frac{1}{2}}}. \tag{4.87}$$

Take the trace across to obtain

$$\left\| K R^{\frac{1}{2}} P^{\frac{1}{2}} \right\|_1 = \mathrm{Tr}\left[\sqrt{P^{\frac{1}{2}} R^{\frac{1}{2}} K^* K R^{\frac{1}{2}} P^{\frac{1}{2}}}\right] \leq \mathrm{Tr}\left[\sqrt{P^{\frac{1}{2}} R P^{\frac{1}{2}}}\right] = \mathrm{F}^{\mathrm{U}}(P,R). \tag{4.88}$$

Thus we see that the inequality is saturated necessarily at some unitary. The linearity of the objective function $\mathrm{Tr}[Y_P]$ implies that the optimum is attained at a unique unitary. This is because if there are two distinct optimal unitaries, by linearity of the objective function, any convex combination of these unitaries would also attain the optimum. However, such a (non-trivial) convex combination will not be a unitary, which would contradict the previous statement. Thus the optimal unitary is unique.

Indeed, U_P is the unique optimal feasible point that achieves this value for M_P, and similar reasoning shows that U_Q is the unique optimal unitary for M_Q. Thus, $\mathrm{X}_\star$ is the unique optimal feasible point of the SDP (Φ, A, B). This concludes the proof. □

Given the above SDP for a triple $P, Q, R \in \mathbb{P}_d^+$ one can *extract* the generalized fidelity $\mathrm{F}_R(P,Q)$ and the squared generalized Bures distance $\mathrm{B}_R(P,Q)$ from the *optimal primal feasible* of SDP. This is formalized in the following theorem.

Theorem 4.4.2. *Define the SDP (Φ, A, B) as above for an arbitrary triple $P, Q, R \in \mathbb{P}_d^+$. Let $\mathrm{X}_\star$ be the optimal primal feasible:*

$$\langle \mathrm{X}_\star, A\rangle = \mathrm{F}^{\mathrm{U}}(P,R) + \mathrm{F}^{\mathrm{U}}(Q,R)\,. \tag{4.89}$$

Then,

$$\mathrm{F}_R(P,Q) = \langle K, \mathrm{X}_\star\rangle, \quad \mathrm{Re}\,\mathrm{F}_R(P,Q) = \left\langle \frac{K+K^*}{2}, \mathrm{X}_\star \right\rangle, \quad \mathrm{B}_R(P,Q) = \langle J, \mathrm{X}_\star\rangle, \tag{4.90}$$

where

$$K := \begin{pmatrix} 0 & \mathbb{I} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad J := \begin{pmatrix} \mathbb{I} & -\mathbb{I} & 0 \\ -\mathbb{I} & \mathbb{I} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.91}$$

Proof. From Proposition 4.4.1, we have that the optimal feasible points $\mathrm{X}_\star$ is of the form

$$\mathrm{X}_\star = \begin{pmatrix} P & Z & Y_P \\ Z^* & Q & Y_Q \\ Y_P^* & Y_Q^* & R \end{pmatrix} = \begin{pmatrix} P & P^{\frac{1}{2}}U_P U_Q^* Q^{\frac{1}{2}} & P^{\frac{1}{2}}U_P R^{\frac{1}{2}} \\ Q^{\frac{1}{2}}U_Q U_P^* P^{\frac{1}{2}} & Q & Q^{\frac{1}{2}}U_Q R^{\frac{1}{2}} \\ R^{\frac{1}{2}}U_P^* P^{\frac{1}{2}} & R^{\frac{1}{2}}U_Q^* Q^{\frac{1}{2}} & R \end{pmatrix} \geq 0. \tag{4.92}$$

Taking trace-inner product of $\langle K, X_\star\rangle = \mathrm{Tr}[K^*\mathrm{X}_\star]$ then gives

$$\mathrm{Tr}\left[Q^{\frac{1}{2}}U_Q U_P^* P^{\frac{1}{2}}\right] = \mathrm{F}_R(P,Q). \tag{4.93}$$

A similar calculation yields

$$\operatorname{Re} \mathrm{F}_R(P,Q) = \left\langle \frac{K+K^*}{2}, \mathrm{X}_\star \right\rangle. \tag{4.94}$$

Finally, we have

$$\langle J, \mathrm{X}_\star \rangle = \operatorname{Tr}[J\mathrm{X}_\star] = \operatorname{Tr}[P+Q] - 2\operatorname{Re} \mathrm{F}_R(P,Q) =: \mathrm{B}_R(P,Q). \tag{4.95}$$

This concludes the proof. □

Thus, we have shown a block-matrix characterization of generalized fidelity and squared generalized Bures distance. The SDP used in the characterization is closely related to the SDP for *optimal average fidelity* [AKF22], which finds a problem that is equivalent to finding the Bures–Wasserstein barycenter of a finite collection of states. Further implications of this relation are explored in Chapter 6.

4.5 Uhlmann-(like) theorem for generalized fidelity

Recall that Uhlmann's theorem states that the Uhlmann fidelity can be written as the largest absolute overlap between purifications. We now show that generalized fidelity can be written as the overlap of particular purifications, with the purifications dependent on the base. See Definition 2.3.2 for the definition of purifications of a positive semidefinite matrix.

Let $P \in \mathbb{P}_d^+$. The *canonical purification* of P is given by

$$|P\rangle = (P^{\frac{1}{2}} \otimes \mathbb{I})|\omega\rangle, \tag{4.96}$$

where $|\omega\rangle := \sum_{i=1}^d |i,i\rangle$ is the unnormalized canonical Bell state. Since there is a unitary degree of freedom in the auxiliary space, we may write arbitrary purifications of $P, Q \in \mathbb{P}_d^+$ as

$$|P_U\rangle \equiv \left(P^{\frac{1}{2}} \otimes U\right)|\omega\rangle \quad \text{and} \quad |Q_V\rangle \equiv \left(Q^{\frac{1}{2}} \otimes V\right)|\omega\rangle. \tag{4.97}$$

Their overlap is given by

$$\langle Q_V, P_U\rangle = \left\langle \omega, \left(Q^{\frac{1}{2}}P^{\frac{1}{2}} \otimes V^*U\right)\omega \right\rangle = \operatorname{Tr}\left[Q^{\frac{1}{2}}P^{\frac{1}{2}}(V^*U)^\intercal\right] = \operatorname{Tr}\left[Q^{\frac{1}{2}}P^{\frac{1}{2}}U^\intercal\overline{V}\right], \tag{4.98}$$

where $\overline{V}$ denotes the complex conjugate of V. Uhlmann's theorem states that

the maximum (absolute-valued) overlap is achieved when $U^\intercal \overline{V} = \mathrm{Pol}\left(P^{\frac{1}{2}} Q^{\frac{1}{2}}\right)$. This result also allows us to state an *Uhlmann's theorem* for generalized fidelity.

Theorem 4.5.1. *Let $P, Q, R \in \mathbb{P}_d^+$. Consider the purifications of P and Q defined as*

$$|P_{U_P^\intercal}\rangle := \left(P^{\frac{1}{2}} \otimes U_P^\intercal\right)|\omega\rangle \quad \textit{and} \quad |Q_{U_Q^\intercal}\rangle := \left(Q^{\frac{1}{2}} \otimes U_Q^\intercal\right)|\omega\rangle, \tag{4.99}$$

where $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}} R^{\frac{1}{2}}\right)$ and $U_Q = \mathrm{Pol}\left(Q^{\frac{1}{2}} R^{\frac{1}{2}}\right)$. Then

$$\mathrm{F}_R(P, Q) = \left\langle P_{U_P^\intercal}, Q_{U_Q^\intercal} \right\rangle. \tag{4.100}$$

Proof. Let $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}} R^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}} R^{\frac{1}{2}}\right)$. Consider the purifications

$$|P_{U_P^\intercal}\rangle = \left(P^{\frac{1}{2}} \otimes U_P^\intercal\right)|\omega\rangle \quad \text{and} \quad |Q_{U_Q^\intercal}\rangle = \left(Q^{\frac{1}{2}} \otimes U_Q^\intercal\right)|\omega\rangle. \tag{4.101}$$

Now consider the overlap

$$\begin{aligned} \langle P_{U_P^\intercal}, Q_{U_Q^\intercal}\rangle &= \langle \omega, \left(P^{\frac{1}{2}} Q^{\frac{1}{2}} \otimes \overline{U}_P U_Q^\intercal\right)\omega\rangle = \mathrm{Tr}\left[P^{\frac{1}{2}} Q^{\frac{1}{2}} \otimes \left(\overline{U}_P U_Q^\intercal\right)^\intercal\right] \\ &= \mathrm{Tr}\left[P^{\frac{1}{2}} Q^{\frac{1}{2}} U_Q U_P^*\right] = \mathrm{Tr}\left[Q^{\frac{1}{2}} U_Q U_P^* P^{\frac{1}{2}}\right] = \mathrm{F}_R(P, Q). \end{aligned} \tag{4.102}$$

This completes the proof. □

Thus, the generalized fidelity $\mathrm{F}_R(P, Q)$ can be seen as the overlap of a pair of specific purifications of P and Q, with the choice of purification depending on R.

4.6 An analogous generalization of some Rényi divergences

We now briefly extend the formalism of generalized fidelities to quantum Rényi divergences. We refer to [AD13; Tom15; Mül+13] for a detailed treatment. We restrict our treatment to normalized states (probability vectors and density matrices), keeping in mind that generalization to non-normalized states is straightforward and can be seen in the previously mentioned references.

Quantum Rényi divergences are *quantizations* of *classical* Rényi divergences [Rén61;

VH14; Tom15], which, for probability vectors $p, q \in \Delta_n$, are defined as

$$D_\alpha(p\|q) := \frac{1}{\alpha - 1} \log \sum_{i=1}^{d} p_i^\alpha q_i^{1-\alpha}, \tag{4.103}$$

for $\alpha \in (0,1) \cup (1,\infty)$. In subsequent discussions, we also restrict our attention to the case supp $P \subset$ supp Q (and thus conveniently ignore the cases where the divergence diverges to ∞).

As with classical fidelity, there is no unique way of generalizing this to positive semidefinite matrices. For an axiomatic approach see [AD13; Tom15; Mül+13]. Here, we will mention the definitions of various well-studied quantum Rényi divergences and show how some of them can be unified by defining a quantity inspired by generalized fidelity. Every quantum Rényi divergence we study is of the form

$$D_\alpha(P\|Q) := \frac{1}{\alpha - 1} \log f_\alpha(P, Q), \tag{4.104}$$

where $f_\alpha(\rho, \sigma)$ is the trace functional which can be thought of as the (asymmetric) *fidelity part* of the divergence. Since the relation between $D_\alpha(P\|Q)$ and $f_\alpha(P, Q)$ follows directly, we will restrict our attention to the trace functional $f_\alpha(P, Q)$.

The first family of divergences we discuss is the family of α-z divergences [AD13], which is defined as

$$D_{\alpha,z}(P\|Q) = \frac{1}{\alpha - 1} \log \operatorname{Tr}\left[\left(P^{\frac{\alpha}{2z}} Q^{\frac{1-\alpha}{z}} P^{\frac{\alpha}{2z}}\right)^z\right], \tag{4.105}$$

for $P \ll Q$, $\alpha \in \mathbb{R}\backslash\{1\}$ (with the limit being taken for $\alpha \to 1$) and $z \in \mathbb{R}_+$ (with limit being taken for $z \to 0$). This family unifies various various quantum Rényi divergences. In particular, set $\alpha = z$ to obtain the *sandwiched relative entropy* of order α:

$$D_{\alpha,\alpha}(P\|Q) = D_\alpha^{\mathrm{S}} := \frac{1}{\alpha - 1} \log \operatorname{Tr}\left[\left(Q^{\frac{1-\alpha}{2\alpha}} P Q^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]. \tag{4.106}$$

Setting $z = 1$ recovers the *Petz-Rényi divergence of order* α:

$$D_{\alpha,1}(P\|Q) = D_\alpha^{\mathrm{PR}}(P\|Q) := \frac{1}{\alpha - 1} \log \operatorname{Tr}[P^\alpha Q^{1-\alpha}]. \tag{4.107}$$

Setting $z = 1 - \alpha$ recovers the *reverse sandwiched relative entropy* of order α:

$$D_{\alpha,1-\alpha}(P\|Q) = D_\alpha^{\mathrm{RS}}(P\|Q) := \frac{1}{\alpha - 1} \log \operatorname{Tr}\left[\left(P^{\frac{\alpha}{2(1-\alpha)}} Q P^{\frac{\alpha}{2(1-\alpha)}}\right)\right]. \tag{4.108}$$

From the sandwiched relative entropy, one can recover other divergences such as the *min-relative entropy*, *Umegaki quantum relative entropy*, and the *max-relative*

entropy:

$$\begin{aligned} D_{\frac{1}{2},\frac{1}{2}}(P\|Q) = D^{\min}(P\|Q) &:= -2\log \mathrm{F}^{\mathrm{U}}(P,Q)\,, \\ \lim_{\alpha\to 1} D_{\alpha,\alpha}(P\|Q) = D^{\mathrm{Um}}(P\|Q) &:= \mathrm{Tr}\left[P\left(\log P - \log Q\right)\right], \\ \lim_{\alpha\to\infty} D_{\alpha,\alpha}(P\|Q) = D^{\max}(P\|Q) &:= \inf\{\lambda : P \le 2^{\lambda} Q\}. \end{aligned} \tag{4.109}$$

Another quantum generalization of Rényi divergence is based on the AI geodesic between P and Q. Specifically, the *α-geometric Rényi divergence* [Mat15; FF21; KW21] between P and Q is defined as

$$D_{\alpha}^{\mathrm{G}}(P\|Q) = \frac{1}{\alpha-1}\log \mathrm{Tr}[Q\#_{\alpha}P] = \frac{1}{\alpha-1}\log \mathrm{Tr}\left[Q^{\frac{1}{2}}\left(Q^{-\frac{1}{2}}PQ^{-\frac{1}{2}}\right)^{\alpha}Q^{\frac{1}{2}}\right]. \tag{4.110}$$

The quantity was first introduced in [Mat15] and is the largest quantum Rényi divergence that satisfies data processing inequality. It is also known [Mat15; Tom15; KW21] that for $P \in \mathbb{D}_d$ and $Q \in \mathbb{P}_d^+$

$$\lim_{\alpha\to 1} D_{\alpha}^{\mathrm{G}}(P\|Q) = D^{\mathrm{BS}}(P\|Q) := \mathrm{Tr}\left[P\log\left(P^{\frac{1}{2}}Q^{-1}P^{\frac{1}{2}}\right)\right], \tag{4.111}$$

where $D^{\mathrm{BS}}(P\|Q)$ is the *Belavkin-Staszewski* relative entropy [BS82].

We will now define a quantity inspired by the definition of generalized fidelity, which can recover many of the above-mentioned divergences. In particular, the quantity we define will recover the following divergences

$$D_{\alpha}^{\mathrm{S}}, D_{\alpha}^{\mathrm{PR}}, D_{\alpha}^{\mathrm{RS}}, \text{ and } D_{\alpha}^{\mathrm{G}}. \tag{4.112}$$

Since the sandwiched relative entropy, in turn, can recover the min-relative entropy $D^{\min}$, Umegaki relative entropy D^{Um}, and max-relative entropy $D^{\max}$ and the geometric Rényi divergence has the Belavkin-Staszewski D^{BS} as its $\alpha \to 1$ limit, the quantity we define would recover these divergences too. We note that this section is meant for introductory and illustrative purposes alone, and further information-theoretic properties are deferred to a future article.

The key idea is to define a *base-dependent* quantity that generalizes the trace functional term. Similar to generalized fidelity, this term is complex in general, and therefore, we will only use its real part.

Definition 4.6.1. *Let $P, Q, R \in \mathbb{P}_d^+$ and $\alpha \in (0,1) \cup (1,\infty)$. Define*

$$\hat{D}_{\alpha,R}(P\|Q) = \frac{1}{\alpha - 1} \log \mathrm{Re}\, \mathrm{F}_R^\alpha(P,Q), \tag{4.113}$$

where the trace functional F_R^α is defined as

$$\mathrm{F}_R^\alpha(P,Q) := \mathrm{Tr}\left[\left(R^{\frac{1}{2}} P R^{\frac{1}{2}}\right)^\alpha R^{-1} \left(R^{\frac{1}{2}} Q R^{\frac{1}{2}}\right)^{1-\alpha}\right]. \tag{4.114}$$

We firstly note that in the *classical scenario*, where P, Q and R mutually commute, the above quantity reduces to the (classical) Rényi divergence.

We now show how the above quantity recovers the previously mentioned divergences. This is formalized in the following theorem.

Theorem 4.6.2. *For $P \in \mathbb{D}_d$ and $Q \in \mathbb{P}_d^+$, define $\hat{D}_{\alpha,R}(P\|Q)$ as above. Then we have*

$$R = \mathbb{I} \quad \Longrightarrow \quad \hat{D}_{\alpha,R}(P\|Q) = D_\alpha^{\mathrm{PR}}(P\|Q) \qquad \textit{(Petz-Rényi)} \tag{4.115}$$

$$R = Q^{\frac{1-\alpha}{\alpha}} \quad \Longrightarrow \quad \hat{D}_{\alpha,R}(P\|Q) = D_\alpha^{\mathrm{S}}(P\|Q) \qquad \textit{(Sandwiched)} \tag{4.116}$$

$$R = P^{\frac{\alpha}{1-\alpha}} \quad \Longrightarrow \quad \hat{D}_{\alpha,R}(P\|Q) = D_\alpha^{\mathrm{RS}}(P\|Q) \qquad \textit{(Rev. sandwiched)} \tag{4.117}$$

$$R = Q^{-1} \quad \Longrightarrow \quad \hat{D}_{\alpha,R}(P\|Q) = D_\alpha^{\mathrm{G}}(P\|Q) \qquad \textit{(Geometric)} \tag{4.118}$$

Proof. We only need to work with the trace functional $\mathrm{F}_R^\alpha(P,Q)$ as its relation to $\hat{D}_{\alpha,R}$ follows directly. Recall the form of the trace functional:

$$\mathrm{F}_R^\alpha(P,Q) := \mathrm{Tr}\left[\left(R^{\frac{1}{2}} P R^{\frac{1}{2}}\right)^\alpha R^{-1} \left(R^{\frac{1}{2}} Q R^{\frac{1}{2}}\right)^{1-\alpha}\right]. \tag{4.119}$$

Choosing $R = \mathbb{I}$, we easily see that

$$\mathrm{F}_{\mathbb{I}}^\alpha(P,Q) = \mathrm{Tr}[P^\alpha Q^{1-\alpha}], \tag{4.120}$$

which leads to the Petz-Rényi divergence. For the Sandwiched Rényi relative entropy, we choose $R = Q^{\frac{1-\alpha}{\alpha}}$. Substituting in the trace functional, we have

$$
\begin{aligned}
\mathrm{F}_R^{\alpha}\left(P,Q\right) :&= \operatorname{Tr}\left[\left(R^{\frac{1}{2}}PR^{\frac{1}{2}}\right)^{\alpha}R^{-1}\left(R^{\frac{1}{2}}QR^{\frac{1}{2}}\right)^{1-\alpha}\right] \\
&= \operatorname{Tr}\left[\left(Q^{\frac{1-\alpha}{2\alpha}}PQ^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}Q^{\frac{\alpha-1}{\alpha}}\left(Q^{\frac{1-\alpha}{2\alpha}}QQ^{\frac{1-\alpha}{2\alpha}}\right)^{1-\alpha}\right] \\
&= \operatorname{Tr}\left[\left(Q^{\frac{1-\alpha}{2\alpha}}PQ^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}Q^{\frac{\alpha-1}{\alpha}}Q^{\frac{1-\alpha}{\alpha}}\right] = \operatorname{Tr}\left[\left(Q^{\frac{1-\alpha}{2\alpha}}PQ^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right],
\end{aligned}
\tag{4.121}
$$

which corresponds to the sandwiched Rényi relative entropy. For the reverse sandwiched Rényi relative entropy, set $R = P^{\frac{\alpha}{1-\alpha}}$, which leads to

$$
\begin{aligned}
\mathrm{F}_R^{\alpha}\left(P,Q\right) :&= \operatorname{Tr}\left[\left(R^{\frac{1}{2}}PR^{\frac{1}{2}}\right)^{\alpha}R^{-1}\left(R^{\frac{1}{2}}QR^{\frac{1}{2}}\right)^{1-\alpha}\right] \\
&= \operatorname{Tr}\left[\left(P^{\frac{\alpha}{2(1-\alpha)}}PP^{\frac{\alpha}{2(1-\alpha)}}\right)^{\alpha}P^{\frac{-\alpha}{1-\alpha}}\left(P^{\frac{\alpha}{2(1-\alpha)}}QP^{\frac{\alpha}{2(1-\alpha)}}\right)^{1-\alpha}\right] \\
&= \operatorname{Tr}\left[P^{\frac{\alpha}{1-\alpha}}P^{\frac{-\alpha}{1-\alpha}}\left(P^{\frac{\alpha}{2(1-\alpha)}}QP^{\frac{\alpha}{2(1-\alpha)}}\right)^{1-\alpha}\right] = \operatorname{Tr}\left[\left(P^{\frac{\alpha}{2(1-\alpha)}}QP^{\frac{\alpha}{2(1-\alpha)}}\right)^{1-\alpha}\right],
\end{aligned}
\tag{4.122}
$$

which is the trace functional that defines the reverse sandwiched Rényi relative entropy. Finally, choose $R = Q^{-1}$ to obtain the α-geometric Rényi divergence:

$$
\begin{aligned}
\mathrm{F}_{Q^{-1}}^{\alpha}\left(P,Q\right) :&= \operatorname{Tr}\left[\left(Q^{-\frac{1}{2}}PQ^{-\frac{1}{2}}\right)^{\alpha}Q\left(Q^{-\frac{1}{2}}QQ^{-\frac{1}{2}}\right)^{1-\alpha}\right] \\
&= \operatorname{Tr}\left[\left(Q^{-\frac{1}{2}}PQ^{-\frac{1}{2}}\right)^{\alpha}Q\cdot\mathbb{I}^{1-\alpha}\right] \\
&= \operatorname{Tr}\left[Q^{\frac{1}{2}}\left(Q^{-\frac{1}{2}}PQ^{-\frac{1}{2}}\right)^{\alpha}Q^{\frac{1}{2}}\right] = \operatorname{Tr}\left[Q\#_{\alpha}P\right],
\end{aligned}
\tag{4.123}
$$

which is the trace functional in the definition of α-geometric Rényi divergence. This concludes the proof. □

4.7 Open problems

We now discuss some related open problems.

1. **Data Processing Inequality.** For a given pair $P, Q \in \mathbb{P}_d^+$, for what values of the base R does the squared generalized Bures distance satisfy the data processing inequality (DPI)? Equivalently, for what values of R does the following inequality hold for any quantum channel Φ?

$$
\mathrm{B}_R(P,Q) \overset{?}{\geq} \mathrm{B}_{\Phi(R)}\left(\left(\Phi(P),\Phi(Q)\right)\right). \tag{4.124}
$$

   Preliminary numerical experiments have not identified any instances that violate DPI.

2. **Convexity.** A related open question is regarding (joint) convexity of generalized Bures distance (and/or its squared version) in P, Q, and R. These appear to be difficult problems, and perhaps it would be easier to tackle these questions for x-Polar fidelities (for $x \in [-1, 1]$).

3. **Recovering z-fidelities from generalized fidelities.** Can the z-fidelities and/or the Log-Euclidean [Nur+24; BGJ19] fidelity be written as the generalized fidelity (or interior fidelity) for some choice of base (or a distribution over bases)? An affirmative answer to the above problem might also lead to the analogous generalization for more members of the α-z divergences.

4. **Monotonicity of polar fidelities.** In the discussion of Paths 5 and 6, we remarked on the numerical observation that the generalized fidelity was monotonic along these curves. Does this observation always hold? That is, for any $P, Q \in \mathbb{P}_d^+$ and any pair $x, y \in [-1, 1]$, does the following statement hold true?

$$x \geq y \implies \begin{cases} \mathrm{F}_{P^x}(P, Q) \geq \mathrm{F}_{P^y}(P, Q)\,, \\ \mathrm{F}_{Q^x}(P, Q) \geq \mathrm{F}_{Q^y}(P, Q)\,. \end{cases} \tag{4.125}$$

   An affirmative answer would also imply that the x-Polar fidelity is monotonic in x in the range $[-1, 1]$. This numerically observed monotonic is also in line with the known relation regarding the named fidelities:

$$\mathrm{F}^{\mathrm{M}}(P, Q) \leq \mathrm{F}^{\mathrm{H}}(P, Q) \leq \mathrm{F}^{\mathrm{U}}(P, Q), \tag{4.126}$$

   for any $P, Q \geq 0$. Moreover, it is also known F^{M} is the smallest and F^{U} is the largest quantization of classical fidelity that satisfies DPI [Mat10]. Thus, the family of x-Polar fidelities (in the range $x \in [-1, 1]$) might be helpful in studying fidelities that satisfy the data processing inequality.

5. **Other bases for Holevo fidelity.** Another open problem we pose is whether other non-trivial bases exist where generalized fidelity recovers Holevo fidelity. Recall that we showed that $R = \mathbb{I}$ implies $\mathrm{F}_R(P, Q) = \mathrm{F}^{\mathrm{H}}(P, Q)$. In fact, this is the only choice of R where we showed this equality for a general (non-commuting) pair of states P, Q. Are there other bases on which the generalized fidelity reduces to Holevo fidelity? To find such a non-trivial base, it would suffice to find $R \in \mathbb{P}_d^+$ such that

$$\mathrm{Pol}\left(R^{\frac{1}{2}} P^{\frac{1}{2}}\right) = \mathrm{Pol}\left(R^{\frac{1}{2}} Q^{\frac{1}{2}}\right). \tag{4.127}$$

   We believe that results from [LL08] would be helpful in this endeavor.

6. **SDP representation.** The next question we ask is whether (the real part of) generalized fidelity or squared generalized Bures distance has a true semidefinite program representation. Though we have presented a block-matrix characterization of generalized fidelity, it is not an SDP. An SDP formulation could be vastly beneficial in optimization problems involving generalized Bures distance, such as in a potential formulation of metric learning [YJ06; ZHS16].

7. **Unitary factor of generalized fidelity and** $\mathrm{SU}(d)$**.** The final open problem we pose is asking whether the set of unitary factors of generalized fidelity and the set of special unitaries are the same. We have shown that the set of unitary factors forms a subset of $\mathrm{SU}(d)$, but it is unclear whether the reverse inclusion holds. To prove this, one would have to show that any special unitary $U \in \mathrm{SU}(d)$ has the form

$$U = U_Q U_P^* \tag{4.128}$$

for a triple $P, Q, R \in \mathbb{P}_d^+$, where $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}} R^{\frac{1}{2}}\right)$ and $U_Q = \mathrm{Pol}\left(Q^{\frac{1}{2}} R^{\frac{1}{2}}\right)$. An affirmative answer would illuminate a connection between the Bures manifold, generalized fidelity, and the Lie group $\mathrm{SU}(d)$. We refer the reader to [Mod16; Uhl10] for detailed expositions of the connection between the unitary factor, the polar decomposition, and the geometry of the Bures manifold.

4.8 Conclusion

In this paper, we introduce a family of fidelities between positive definite matrices that generalize and unify various existing quantum fidelities. The definition is motivated by the Riemannian geometry of the Bures–Wasserstein manifold, and it endows the existing fidelities with novel geometric interpretations. We also define and study the generalization of the closely related Bures–Wasserstein distance.

After studying the basic properties of these objects, we proved several remarkable geometric properties of generalized fidelity including invariance and covariance properties of generalized fidelity along geodesic-related paths related to the Bures–Wasserstein, Affine-invariant, and Euclidean Riemannian metrics on the manifold of positive definite matrices.

We then showed how convex combinations of generalized fidelity define a new family of fidelities. One such family, the *Polar fidelity*, is shown to be a family of fidelities parametrized by a single real number that recovers Uhlmann-, Holevo-, and Matsumoto fidelity. We then derived a block-matrix characterization of gen-

eralized fidelity and squared generalized Bures–Wasserstein distance and showed an interesting relation between generalized fidelity, multivariate fidelities, and Bures–Wasserstein barycenters.

We also discussed an *Uhlmann-like* theorem for generalized fidelity, and finally, we demonstrated how our formalism can also be extended to generalized certain quantum Rényi divergences. We then discussed various open problems.

We conclude this section by discussing potential applications of generalized Bures distance in quantum and classical machine learning. Distances are crucial in machine learning as we often embed data points in high-dimensional metric spaces, and often, we find that different distance metrics (and divergences) are suitable for different tasks. In particular, consider the problem of *Metric learning* [YJ06; ZHS16; Kul+13; Dav+07]. The problem can be succinctly stated as follows: find the distance metric that is *best* suited to represent the data. For example, one could consider the case where we are given pairwise distances between many data points (represented by vectors in $\mathbb{R}^d$), and we are interested in finding the positive definite matrix that defines the Mahalanobis distance which would be most consistent with the given distances [ZHS16]. Metric learning has applications in various machine learning fields, including classification algorithms like the K Nearest Neighbors algorithm [Pet09] and dimensionality reduction problems [WS15; HSH17].

Generalized Bures distance could extend the problem to the setting where the data points are positive definite matrices or quantum states. Specifically, given a collection of states and associated data (such as pairwise distances or labels for each state), find a suitable base for the given task. For example, this could be a classification problem where the states are given labels indicating membership in disjoint classes. Then, the task would be to find a base such that the generalized Bures distance at this base increases *interclass* distance while decreasing *intraclass* distance. Thus, metric learning based on generalized Bures distance could benefit various classical ML problems involving the BW manifold or even quantum variants of the aforementioned (classical) ML problems [WKS15; BAG20; SSP14; Dua+19; Lia+20].

Chapter 5

Bures projection

5.1 Introduction

This chapter studies the final set of results of this thesis. We are interested in the *projection problem*, which can be defined as finding the closest point in a particular set to a given point outside the set. To elaborate, we are given a set $\mathcal{C}$ and a point $x \notin \mathcal{C}$. The problem is to find the closest point $y \in \mathcal{C}$ to x. Such a y, if it exists, is called the *projection* of x to $\mathcal{C}$.

Perhaps the earliest projection problem, and indeed something that is taught in high school mathematics[1], is the projection of a point to a line with respect to Euclidean distance. The solution, as Pythagoras theorem would tell us, is obtained by dropping the perpendicular from the point to the line.

More generally, we are given a set $\mathcal{V}$ which has a squared distance (or a divergence) D defined on it. Then the *projection of* $x \in \mathcal{V}$ to a subset $\mathcal{C} \subset \mathcal{V}$ is the (a) solution to the optimization problem

$$\Pi_{\mathcal{C}}^{\mathrm{D}}[x] := \operatorname*{argmin}_{y \in \mathcal{C}} \mathrm{D}(y; x), \tag{5.1}$$

provided it exists. As one expects, the projection varies as the choice of the distance (or divergence) D changes. Typically the set $\mathcal{C}$ is convex, and the distance is chosen to be the Euclidean distance. For the Euclidean setting, efficient methods exist for the solution to the projection problem in certain scenarios.

In this chapter, in line with the general affinity to fidelity, we have in the thesis, we are interested in projection with respect to fidelity, or more precisely, Bures distance. That is, we are interested in projecting a given positive matrix to certain convex and compact subsets of the positive semidefinite cone. For the purposes of this thesis, these sets are the preimages of a given *output state* under

[1] The author still remembers learning this in high school.

a given quantum channel.

Let us formalize these notions. Let $\mathcal{H}$ and $\mathcal{K}$ be finite-dimensional complex Euclidean spaces and let $P \in \mathbb{P}_{\mathcal{H}}$, $C \in \mathbb{P}_{\mathcal{K}}$ be states and $\Lambda \in \mathrm{CPT}(\mathcal{H}, \mathcal{K})$ be a quantum channel. For the *constraint pair* (Λ, C), define its *feasible set* as

$$\Lambda^{-1}[C] := \{Q \in \mathbb{P}_{\mathcal{H}} : \Lambda(Q) = C\}. \tag{5.2}$$

Note that we are interested only in positive semidefinite feasible points and thus defined the set accordingly. This set is the intersection of an affine subspace of $\mathbb{H}_{\mathcal{H}}$ (as defined by the constraint $\Lambda(X) = C$) and the closed set of positive semidefinite cone $\mathbb{P}_{\mathcal{H}}$, which implies $\Lambda^{-1}[C]$ is a compact set. Such a set is called a *spectrahedron* and forms the feasible region of semidefinite programs [RG95; Vin14; Chi23].

Observe that $|\Lambda^{-1}[C]|$ equals $0, 1$, or ∞. The first case is vacuous and thus ignored. The second case can be solved via linear inversion. Thus we only consider the third case, where the feasible set is a spectrahedron containing an infinite number of elements.

The projection problem can be then formally stated as follows:

$$\Pi_{\Lambda,C}[P] := \operatorname*{argmin}_{Q \in \Lambda^{-1}[C]} \mathrm{B}(P, Q) = \operatorname*{argmax}_{Q \in \Lambda^{-1}[C]} \mathrm{F}(P, Q), \tag{5.3}$$

where, again, we have taken to Uhlmann fidelity to be the *unlabeled* fidelity. We first remark that the projection problem is a textbook case of a semidefinite program, and thus numerical solutions are easily obtained. However, in this work, we are interested in closed-form solutions. Here $\Pi_{\Lambda,C}[P]$ is taken to be the set of all feasible elements that attain the optimal value. Since we are minimizing a continuous function over a compact set, the set of solutions $\Pi_{\Lambda,C}[P]$ is non-empty. If $\Pi_{\Lambda,C}[P]$ is a singleton set then we use, via a slight abuse of notation, $\Pi_{\Lambda,C}[P]$ to denote this unique projection.

The equivalence between minimizing Bures distance and maximizing fidelity becomes clear by expanding the squared Bures distance and observing that the trace terms do not play a role in the optimization, as every feasible point has the same trace (equal to $\mathrm{Tr}[C]$), owing to the fact that Λ is trace-preserving.

Remark 5.1.1 (Uniqueness of projection)**.** We have shown that a projection always exists. The question of uniqueness is slightly trickier. A sufficient condition for the uniqueness is if the P is full-rank and there exists a full-rank optimal feasible $Q \in \Pi_{\Lambda,C}[P]$. To see this, recall that fidelity $\mathrm{F}_P(Q) \equiv \mathrm{F}(P, Q)$ is a strictly concave function of $Q > 0$ whenever $P > 0$, and thus any full-rank optimum is necessarily unique.

The rest of the chapter is structured as follows. In Section 5.2 we introduce a function that serves as a closed-form for the projection for certain channels such as partial trace (including trace), pinching channels (including completely dephasing channel), and projective measurements. This closed-form is closely related to the saturation of the data processing inequality for fidelity. Moreover, we show that for these channels, the input and output states (of the projection problem) and the constraint channel saturate the Data Processing Inequality for fidelity and elaborate upon the explicit closed forms for the above-mentioned channels. In Section 5.3 we discuss an extension of channel-state duality based on the ideas of Chruściński and Matsuoka [CM20], Leifer [Lei06], and Leifer and Spekkens [LS13] which turns out to be quite useful while discussing Bures projections with respect to the partial-trace channel. We then discuss a plethora of applications in Chapter 5.4 including:

1. A *Bures* projected least squares channel tomography technique.
2. A Bayesian channel tomography technique.
3. Application in generic optimization problems over certain compact sets.
4. A unified method to generate various random states and ensembles of interest in quantum information.

We then show that Bures projection has manifested in various existing results and protocols of importance in quantum information. In particular, we show that

1. Certain *normalizations* [AS08; Bru+09; Kuk+21; CŻ24] used to construct CPT and CPU maps from CP maps (at the level of Choi matrices) are actually Bures projections.
2. The pretty good measurement [Bel75; HW94; EF01] can be given novel geometric and operational interpretations in terms of Bures projection of ensembles.
3. A novel geometric interpretation for the Petz recovery map [Pet86b; Pet88] in terms of Bures projection. We show that (the transpose of the Choi matrix of) the Petz recovery map is the Bures projection of a certain joint channel-state matrix defined by the original channel and the reference state.
4. A geometric interpretation and shorter proof for the recent result by Bai, Buscemi, and Scarani [BBS24] on *quantum minimal change principle*.

Finally, we conclude with Section 5.6 where we discuss future directions and open problems. We begin with the derivation of the closed form for projections.

(Λ, C)	$\Gamma_{\Lambda,C}[P]$	Remarks
(Tr, c)	$c\dfrac{P}{\mathrm{Tr}[P]}$	$c \in \mathbb{R}_+$. $c = 1$ yields trace normalization.
$(\mathrm{Tr}_{\mathcal{Y}}, C)$	$P \star_{\mathcal{X}} (P_{\mathcal{X}}^{-1} \# C)^2$	$P \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$. $C \in \mathbb{P}_{\mathcal{X}}$. $C = \mathbb{I}_{\mathcal{X}}$ gives projection to Choi matrices of CPT maps from $\mathcal{X}$ to $\mathcal{Y}$.
$(\mathrm{Tr}_{\mathcal{X}}, C)$	$P \star_{\mathcal{Y}} (P_{\mathcal{Y}}^{-1} \# C)^2$	$P \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$. $C \in \mathbb{P}_{\mathcal{Y}}$. $C = \mathbb{I}_{\mathcal{Y}}$ gives projection to Choi matrices of CPU from $\mathcal{X}$ to $\mathcal{Y}$.
(Δ, C)	$\sum_{i,j=1}^{d} \sqrt{\dfrac{C_i C_j}{P_{ii} P_{jj}}} P_{ij} \lvert i\rangle\langle j\rvert$	Δ is the completely dephasing map. C is a diagonal positive definite matrix.
(Λ, C)	$P \star \left(\bigoplus_{i=1}^{n} [P]_i^{-1} \# [C]_i\right)^2$	Λ is the pinching channel. $[C]_i$ is a block diagonal matrix compatible with Λ. $[P]_i$ are the corresponding blocks of P (the form described on the left assumes block-diagonality in computational basis).
$(\mathrm{M}_{\mathcal{E}}, c)$	$P \star \left(\sum_{i=1}^{n} \sqrt{\dfrac{c_i}{\langle E_i, P\rangle}} E_i\right)^2$	M is a projective measurement with orthogonal projectors $\mathcal{E} := \{E_i\}$. $c \in \mathbb{R}_+^n$ is a positive vector.

Table 5.1: Bures projection closed-form for specific channels.

5.2 Closed-form for projections

The aim of this section is to show that for certain channels, the projection can be given by a specific closed form. Essentially, the closed-form can be *defined* for any channel, but whether it yields the projection depends on the channel and, interestingly, is related to the saturation of the data processing inequality (DPI) for fidelity.

We begin with two useful lemmas. The first is based on the main result of Leditzky, Rouzé, and Datta [LRD17] (also see Cree and Sorce [CS22b]) necessary and sufficient conditions for the saturation of DPI for fidelity. The second lemma shows that saturation of DPI implies projection.

Lemma 5.2.1 (Leditzky, Rouzé, and Datta [LRD17])**.** *Let* $\Lambda \in \mathrm{CPT}(\mathcal{H}, \mathcal{K})$ *and* $P, Q \in \mathbb{P}_{\mathcal{H}}$ *such that* $P \not\perp Q$*. Then* $\mathrm{F}(P, Q) = \mathrm{F}(\Lambda(P), \Lambda(Q))$ *if and only if*

$$Q^{\frac{1}{2}} \sqrt{Q^{-\frac{1}{2}} P^{-1} Q^{-\frac{1}{2}}} Q^{\frac{1}{2}} = \Lambda^* \left(\Lambda(Q)^{\frac{1}{2}} \sqrt{\Lambda(Q)^{-\frac{1}{2}} \Lambda(P)^{-1} \Lambda(Q)^{-\frac{1}{2}}} \Lambda(Q)^{\frac{1}{2}} \right). \tag{5.4}$$

If $P, Q, \Lambda(P)$*, and* $\Lambda(Q)$ *are invertible, this above equation is the same as*

$$P^{-1} \# Q = \Lambda^* \left(\Lambda(P)^{-1} \# \Lambda(Q) \right). \tag{5.5}$$

By the symmetry of fidelity, one can swap the operators P *and* Q *to obtain equivalent relations.*

Remark. The original result from Leditzky, Rouzé, and Datta [LRD17] derives the condition for saturation of DPI for the more general sandwiched Rényi divergence of order α. We only require the case of $\alpha = 1/2$, which is equivalent to saturation of DPI for fidelity.

Some comments are due. Observe that the LHS and RHS of Eq. (5.4) resembles the forms of the geometric mean $P^{-1} \# Q$ and $\Lambda(P)^{-1} \# \Lambda(Q)$ respectively. However, the equivalence typically holds true only if they are invertible. Properties of geometric mean mentioned in Proposition 2.1.3 need not hold true in the case P or Q are not invertible. Thus, henceforth, we will put the minimal assumption that relevant quantities are invertible.

The next lemma states that if a triple (Λ, P, Q) saturates DPI for fidelity for positive definite P and Q, then P is the nearest point to Q over all points P' with $\Lambda(P') = \Lambda(P)$ and conversely Q is the nearest point to P over all points Q' with $\Lambda(Q') = \Lambda(Q)$.

Lemma 5.2.2. *Let* $P, Q \in \mathbb{P}^+_{\mathcal{H}}$ *and* $\Lambda \in \mathrm{CPT}(\mathcal{H}, \mathcal{K})$*. If the triple* (Λ, P, Q) *saturates DPI for fidelity, then* Q *is the unique projection of* P *to the set* $\Lambda^{-1}[\Lambda(Q)]$ *and* P *is the unique projection of* Q *to the set* $\Lambda^{-1}[\Lambda(P)]$*:*

$$Q = \Pi_{\Lambda, \Lambda(Q)}[P] \quad \textit{and} \quad P = \Pi_{\Lambda, \Lambda(P)}[Q]. \tag{5.6}$$

Proof. We first show that if (Λ, P, Q) saturates DPI, then $Q \in \Pi_{\Lambda, \Lambda(Q)}[P]$. Since P and Q are full-rank by assumption, it would then follow that Q is the unique projection.

Suppose (Λ, P, Q) saturates DPI: $\mathrm{F}(P, Q) = \mathrm{F}(\Lambda(P), \Lambda(Q))$. We want to show that for any $Q' \in \Lambda^{-1}[\Lambda(Q)]$, it holds that $\mathrm{F}(P, Q') \leq \mathrm{F}(P, Q)$. Towards

contradiction, assume there exists a feasible point Q' with strictly higher fidelity with P. Then we have

$$\mathrm{F}(P, Q') > \mathrm{F}(P, Q) = \mathrm{F}(\Lambda(P), \Lambda(Q)) = \mathrm{F}(\Lambda(P), \Lambda(Q')). \tag{5.7}$$

Here the first equality is by assumption and the last equality follows by the fact that $\Lambda(Q) = \Lambda(Q')$ as Q' is a feasible point. However, this implies the violation of DPI for fidelity for the triple (Λ, P, Q'), which is impossible. Hence such a feasible point Q' cannot exist.

This implies $\mathrm{F}(P, Q) \geq \mathrm{F}(P, Q')$ for all $Q' \in \Lambda^{-1}[\Lambda(Q)]$, which implies Q is a projection of P to this set. Since P and Q are assumed to be full-rank, it follows from strict concavity of fidelity that it uniquely maximizes fidelity over the convex and compact feasible set, and thus Q is the unique projection. The second equation of Eq. (5.6) follows by swapping P and Q in the above arguments. □

Having these lemmas in place, we define the *Gamma map* which, for certain channels, yields the solution of Bures projection. After the definition, we study the condition under which the Gamma map yields the projection.

Definition 5.2.3. *Let (Λ, C) denote a constraint pair for $C \in \mathbb{P}_{\mathcal{K}}$, and $\Lambda \in \mathrm{CPT}(\mathcal{H}, \mathcal{K})$. The* Gamma map $\Gamma_{\Lambda,C} : \mathbb{P}_{\mathcal{H}} \to \mathbb{P}_{\mathcal{K}}$ *is defined as*

$$\begin{aligned} \Gamma_{\Lambda,C}[P] &:= P \star [\Lambda^*(\Lambda(P)^{-1} \# C)]^2 \\ &= [\Lambda^*(\Lambda(P)^{-1} \# C)] P [\Lambda^*(\Lambda(P)^{-1} \# C)], \end{aligned} \tag{5.8}$$

for any $P \in \mathbb{P}_{\mathcal{H}}$.

A sufficient condition for $\Lambda^*(\Lambda(P)^{-1} \# C)$ to be invertible is $\Lambda(P) > 0$ and $C > 0$. Observe that $\Gamma_{\Lambda,C}$ map can be defined for any projection problem (P, Λ, C). For certain channels—such as partial trace, pinching maps, and projective measurements with orthogonal projectors—the Γ map is a closed form for the Bures projection. See Table 5.1 for the specific forms the Γ map takes for these channels.

We now show that a sufficient condition for the Gamma map to yield the projection is if it yields a feasible point. For this theorem, we assume $P, \Lambda(P)$, and C are invertible.

Theorem 5.2.4. *Let (P, Λ, C) denote a projection problem such that $P, \Lambda(P)$, and C are positive definite matrices. Then the Gamma map yields the unique projection if it yields a feasible point:*

$$\Gamma_{\Lambda,C}[P] \in \Lambda^{-1}[C] \quad \implies \quad \Gamma_{\Lambda,C}[P] = \Pi_{\Lambda,C}[P]. \tag{5.9}$$

Proof. Observe that

$$\Lambda(P), C > 0 \implies \Lambda(P)^{-1} \# C > 0 \implies \Lambda^*(\Lambda(P)^{-1} \# C) > 0, \tag{5.10}$$

where the last implication follows from the fact that unital maps are strictly positive maps. We thus have

$$Q \equiv \Gamma_{\Lambda,C}[P] := P \star (\Lambda^*(\Lambda(P)^{-1} \# C))^2 > 0, \tag{5.11}$$

since we have assumed $P > 0$. This allows us to write

$$P^{-1} \# Q = \Lambda^*(\Lambda(P)^{-1} \# C). \tag{5.12}$$

Now let us prove the main statement. To this end, assume $Q \equiv \Gamma_{\Lambda,C}[P] \in \Lambda^{-1}[C]$, or equivalently $\Lambda(Q) = C$. Eq. (5.12) then takes the form

$$P^{-1} \# Q = \Lambda^*(\Lambda(P)^{-1} \# \Lambda(Q)), \tag{5.13}$$

which, by Lemma 5.2.1, is a necessary and sufficient condition for (Λ, P, Q) to saturate DPI for fidelity. This further implies (by Lemma 5.2.2) $Q \in \Pi_{\Lambda,C}[P]$. Since Q is full-rank, it is also the unique projection. This concludes the proof. □

The above theorem has the following implications.

1. Closed-form for Bures projection for certain channels.
2. A way to verify if the closed-form *works* for an arbitrary channel—just check if the function $\Gamma_{\Lambda,C}$ yields a feasible point.
3. Explicit examples for the saturation of DPI for fidelity—the triple $(\Lambda, P, \Gamma_{\Lambda,C}[P])$ saturate DPI for fidelity (and thus also for sandwiched Rényi divergence of order $\alpha = 1/2$) for any channel where the map Γ yields a feasible point.

We now show that for the channels previously mentioned (partial trace, pinching channels, and projective measurements), the Gamma map yields the projection. This is done by explicitly showing that $\Gamma_{\Lambda,C}[P]$ is an element of $\Lambda^{-1}[C]$. We begin with Λ being the partial trace channel.

5.2.1 Partial trace projection

We now show that when the projection problem is defined by the partial trace channel, the Gamma map indeed yields the closed-form solution. Due to the ubiquity of the Bures projection problem with respect to partial channel in the rest of the paper, we reserve a separate name and call this the *marginal projection* problem. Let us identify the relevant objects.

Let $\mathcal{H} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ and $\mathcal{K} = \mathcal{X}_k$ for some $k \in [n]$. Let $\Lambda = \mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}} \in \mathrm{CPT}(\mathcal{H}, \mathcal{K})$ be the partial trace channel that discards every subsystem except $\mathcal{K} = \mathcal{X}_k$. Let $C \in \mathbb{P}^+_{\mathcal{X}_k}$ and denote $\mathcal{J} \equiv \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_{k-1}$ and $\mathcal{L} \equiv \mathcal{X}_{k+1} \otimes \cdots \otimes \mathcal{X}_n$ such that $\mathcal{H} = \mathcal{J} \otimes \mathcal{K} \otimes \mathcal{L}$. We then have the following result.

Theorem 5.2.5 (Closed-form for marginal projection)**.** *Let $\mathcal{H}, \mathcal{K}$, and $\Lambda := \mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}}$ be defined as above. Let $C \in \mathbb{P}^+_{\mathcal{K}}$ be chosen arbitrarily. For any $P \in \mathbb{P}^+_{\mathcal{H}}$, we have*

$$\begin{aligned} \Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P] &= P \star (\mathbb{I}_{\mathcal{J}} \otimes [P]_{\mathcal{K}}^{-1} \# C \otimes \mathbb{I}_{\mathcal{L}})^2 \\ &= P \star_{\mathcal{K}} ([P]_{\mathcal{K}}^{-1} \# C)^2 \qquad \in \mathrm{Tr}^{-1}_{\mathcal{H}\setminus\mathcal{K}}[C]. \end{aligned} \tag{5.14}$$

where $[P]_{\mathcal{K}} = [P]_{\mathcal{X}_k} \equiv \mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}}[P]$ is the $\mathcal{X}_k$-marginal of P.

Proof. By definition, we have

$$\Gamma_{\Lambda,C}[P] := P \star (\Lambda^*(\Lambda(P)^{-1} \# C))^2. \tag{5.15}$$

Make the identifications $\Lambda(P) = P_{\mathcal{K}}$ and $\Lambda^*(P_{\mathcal{K}}^{-1} \# C) = \mathbb{I}_{\mathcal{J}} \otimes P_{\mathcal{K}}^{-1} \# C \otimes \mathbb{I}_{\mathcal{L}}$, whence it follows

$$\begin{aligned} \Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P] &= P \star (\mathbb{I}_{\mathcal{J}} \otimes P_{\mathcal{K}}^{-1} \# C \otimes \mathbb{I}_{\mathcal{L}})^2 \\ &= P \star_{\mathcal{K}} (P_{\mathcal{K}}^{-1} \# C)^2. \end{aligned} \tag{5.16}$$

Now we must show that $\Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P]$ is a feasible point, or equivalently show that

$$\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}} \left[\Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P] \right] = \left[\Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P] \right]_{\mathcal{K}} \stackrel{?}{=} C. \tag{5.17}$$

To see this, observe that

$$\left[\Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P] \right]_{\mathcal{K}} = [P \star_{\mathcal{K}} (P_{\mathcal{K}}^{-1} \# C)^2]_{\mathcal{K}} = P_{\mathcal{K}} \star (P_{\mathcal{K}}^{-1} \# C)^2 = C, \tag{5.18}$$

where, in the second equality, we used the property of the star product $[P \star_{\mathcal{K}} K]_{\mathcal{K}} = [P]_{\mathcal{K}} \star K$ for $K \in \mathbb{P}_{\mathcal{K}}$ (Eq. (2.21)) and in the final equality we used the fact that $A \star (A^{-1} \# B)^2 = B$ for any $A, B > 0$. We have proven that $\Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P]$ is a feasible point. By Theorem 5.2.4 we have that it is also the projection:

$\Gamma_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P] = \Pi_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C}[P]$. This concludes the proof. □

We thus have a closed form for projection with respect to the partial trace channel (marginal projection). Due to the subsequent ubiquity of this projection, we reserve a separate notation for it:

$$\Pi_{\mathrm{Tr}_{\mathcal{H}\setminus\mathcal{K}},C} \equiv \hat{\Pi}_{\mathcal{K},C}. \tag{5.19}$$

Essentially $\hat{\Pi}_{\mathcal{K},C}$ denotes the projection to the compact set of positive semidefinite matrices whose $\mathcal{K}$-marginal is $C \in \mathbb{P}_{\mathcal{K}}$. We now look at some special cases of marginal projection, with the simplest one being the trace map.

Corollary 5.2.6 (Trace-projection)**.** *Let $\mathcal{K} = \mathbb{C}, \Lambda = \mathrm{Tr}$, and $c \in \mathbb{R}_+$. Then for any $P \in \mathbb{P}_{\mathcal{H}}$ the projection with respect to the trace channel is given by*

$$\Pi_{\mathrm{Tr},c}[P] = c\frac{P}{\mathrm{Tr}[P]}. \tag{5.20}$$

On choosing $c = 1$, we get the Bures projection to the set of density matrices, which is given by trace-normalization: $\Pi_{\mathrm{Tr},1}[P] = \frac{P}{\mathrm{Tr}[P]}$

Proof. Proof directly follows from Theorem 5.2.5 on using the fact that $\mathrm{Tr}^*[c] = c\mathbb{I}_{\mathcal{H}}$. □

Of particular importance would be marginal projection over bipartite systems, whose closed forms are given as follows.

Proposition 5.2.7 (Marginal projection for bipartite systems)**.** *Let $\mathcal{H} = \mathcal{X} \otimes \mathcal{Y}$, $C \in \mathbb{P}^+_{\mathcal{X}}$, and $D \in \mathbb{P}^+_{\mathcal{Y}}$. For any $P \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$, the $\mathcal{X}$- and $\mathcal{Y}$-marginal projections are given by*

$$\hat{\Pi}_{\mathcal{X},C}[P] = P \star_{\mathcal{X}} (P_{\mathcal{X}}^{-1} \# C)^2 \quad \text{and} \quad \hat{\Pi}_{\mathcal{Y},D}[P] = P \star_{\mathcal{Y}} (P_{\mathcal{Y}}^{-1} \# D)^2. \tag{5.21}$$

In particular, choosing $C = \mathbb{I}_{\mathcal{X}}$ and $D = \mathbb{I}_{\mathcal{Y}}$ yields

$$\hat{\Pi}_{\mathcal{X},C}[P] = P \star_{\mathcal{X}} [P]_{\mathcal{X}}^{-1} \quad \text{and} \quad \hat{\Pi}_{\mathcal{Y},D}[P] = P \star_{\mathcal{Y}} [P]_{\mathcal{Y}}^{-1}, \tag{5.22}$$

where we use the fact that $A\#\mathbb{I} = \sqrt{A}$ for any $A \geq 0$.

Proof. The proof follows from relevant substitutions in Theorem 5.2.5. □

In later sections, we will use these marginal projections over bipartite systems to perform projections related to quantum channels (or CP maps in more gen-

eral). Let us thus quickly remind ourselves of the relation between CP maps and bipartite PSD matrices. See Section 2.3.2 for further details.

Recall that, by Choi-Jamiołkowski isomorphism, the set of bipartite PSD matrices $\mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$ is isomorphic to the set of completely positive maps $\mathrm{CP}(\mathcal{X},\mathcal{Y})$ from $\mathbb{M}_{\mathcal{X}}$ to $\mathbb{M}_{\mathcal{Y}}$. Indeed it is also isomorphic to the set of completely positive maps $\mathrm{CP}(\mathcal{Y},\mathcal{X})$ from $\mathbb{M}_{\mathcal{Y}}$ to $\mathbb{M}_{\mathcal{X}}$. That is, for any $P \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$, one can define the CP maps $\Phi_P \in \mathrm{CP}(\mathcal{X},\mathcal{Y})$ and $\Psi_P \in \mathrm{CP}(\mathcal{Y},\mathcal{X})$ with actions defined as

$$\Phi_P(X) := [P \cdot (X^\intercal \otimes \mathbb{I}_{\mathcal{Y}})]_{\mathcal{Y}} \quad \text{and} \quad \Psi_P(Y) := [P \cdot (\mathbb{I}_{\mathcal{X}} \otimes Y^\intercal)]_{\mathcal{X}}, \tag{5.23}$$

for any $X \in \mathbb{M}_{\mathcal{X}}$ and $Y \in \mathbb{M}_{\mathcal{Y}}$. We have the following relations as well:

$$\begin{aligned} [P]_{\mathcal{X}} = \mathbb{I}_{\mathcal{X}} \quad &\iff \Phi_P \in \mathrm{CPT}(\mathcal{X},\mathcal{Y}) \text{ and } \Psi_P \in \mathrm{CPU}(\mathcal{Y},\mathcal{X}), \\ [P]_{\mathcal{Y}} = \mathbb{I}_{\mathcal{Y}} \quad &\iff \Phi_P \in \mathrm{CPU}(\mathcal{X},\mathcal{Y}) \text{ and } \Psi_P \in \mathrm{CPT}(\mathcal{Y},\mathcal{X}). \end{aligned} \tag{5.24}$$

That is, if the marginal on the *input space* is identity, then the map is CP and trace-preserving and if the marginal on the *output space* is identity, then the map is CP and unital.

This brief overview of the Choi-Jamiołkowski isomorphism is to remind us of an important application of marginal projection. Essentially Proposition 5.2.7 gives us a way to find the closest (at the level of Bures distance between Choi matrices) trace-preserving and (separately) unital completely positive maps to a given completely positive map. As we will discuss in Section 5.4, this result has applications in practical quantum process tomography and also in the geometric interpretations of various results in quantum information.

Owing to the fact that we will use these projections frequently, we will reserve special notations for them. In the following equations, we assume the positive matrix (to be projected) is an element of $\mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$.

$$\hat{\Pi}_{\mathcal{X},\mathbb{I}_{\mathcal{X}}} \equiv \Pi_{\mathrm{CPT}(\mathcal{X},\mathcal{Y})} \equiv \Pi_{\mathrm{CPU}(\mathcal{Y},\mathcal{X})}, \tag{5.25}$$

where the second equivalence signifies the equivalence between the sets $\mathrm{CPT}(\mathcal{X},\mathcal{Y})$ and $\mathrm{CPU}(\mathcal{Y},\mathcal{X})$. Similarly, we denote

$$\hat{\Pi}_{\mathcal{Y},\mathbb{I}_{\mathcal{Y}}} \equiv \Pi_{\mathrm{CPU}(\mathcal{X},\mathcal{Y})} \equiv \Pi_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}, \tag{5.26}$$

Remark. This is not the first time these operations have appeared in the literature. They have been used in Audenaert and Scheel [AS08] and Bruzda, Cappellini, Sommers, and Życzkowski [Bru+09] (where they are considered as *normalizations*) to generate random Choi matrices of quantum channels and CPU

maps. However, to the best of our knowledge, hitherto it has not been shown that these constructions are bona fide projections with respect to Bures distance. By recognizing these *normalizations* as bona fide projections, one can now rigorously analyze protocols involving these operations. Such protocols involving these projections, both new and existing, are discussed in Section 5.4.

Next, we discuss the projection closed-form for a different class of channels, namely, pinching channels [Wat18; Tom15] which also contain the completely dephasing channel as a special case.

5.2.2 Pinching channel projection

We next discuss the closed form for projection with respect to pinching channels. Let $\mathcal{E} := \{E_i\}_{i\in[n]}$ be a collection of mutually orthogonal projectors summing to unity. Then the pinching channel with respect $\mathcal{E}$ is given by

$$\Lambda_{\mathcal{E}}(P) := \sum_{i=1}^{n} E_i P E_i. \tag{5.27}$$

Observe that $\Lambda_{\mathcal{E}}$ is a self-adjoint map: $\Lambda_{\mathcal{E}}^* = \Lambda_{\mathcal{E}}$. We now show that the Gamma map yields a feasible point and thus constitutes the projection. We first prove the case where E_i are diagonal in the computational basis as this is the more illustrative scenario. The proof for the general case is deferred to Appendix D.1.

Theorem 5.2.8. *Let $\mathcal{E} = \{E_i\}_{i\in[n]} \subset \mathbb{P}_{\mathcal{H}}$ be a collection of orthogonal projectors summing to $\mathbb{I}_{\mathcal{H}}$ which are diagonal in the computational basis. Let $\Lambda_{\mathcal{E}}$ be the corresponding pinching channel. Let $C \in \mathbb{P}_{\mathcal{H}}$ be in the image of $\Lambda_{\mathcal{E}}$. We then have*

$$\begin{aligned} \Gamma_{\Lambda_{\mathcal{E}},C}[P] &= P \star (\Lambda_{\mathcal{E}}(\Lambda_{\mathcal{E}}(P)^{-1} \# C))^2 \\ &= P \star \left(\bigoplus_{i=1}^{n} [P]_i^{-1} \# [C]_i \right)^2 \qquad \in \Lambda_{\mathcal{E}}^{-1}[C], \end{aligned} \tag{5.28}$$

where $[P]_i$ denotes the non-zero block of $E_i P E_i$.

Remark. In the above statement and for the rest of the proof, we assume we are in a basis such that $E_i P E_i$ can be written a $d \times d$ matrix with every element being zero except an $r_i \times r_i$ principal submatrix, which we denote by $[P]_i$, with $r_i := \mathrm{rank}(E_i)$. Essentially we choose the basis which *visually illustrates* action

of the pinching map. That is, the basis such that

$$\bigoplus_{i=1}^{n}[P]_i = \sum_{i=1}^{n} E_i P E_i, \tag{5.29}$$

holds for any P, where $[P]_i$ is an $r_i \times r_i$ block.

Proof. By the action of the pinching channel, we have both $\Lambda_{\mathcal{E}}(P) = \bigoplus_{i=1}^{n}[P]_i$ and $C = \Lambda_{\mathcal{E}}[C] \equiv \bigoplus_{i=1}^{n}[C]_i$ to be block-diagonal in the computational basis, where $[P]_i$ and $[C]_i$ indicate the non-zero blocks of $E_i P E_i$ and $E_i C E_i$. Indeed it follows that

$$\Lambda_{\mathcal{E}}(P)^{-1} = \bigoplus_{i=1}^{n}[P]_i^{-1}, \tag{5.30}$$

whence it follows

$$M \equiv \Lambda_{\mathcal{E}}(P)^{-1} \# C = \bigoplus_{i=1}^{n}[P]_i^{-1} \# [C]_i. \tag{5.31}$$

Observe that $\Lambda_{\mathcal{E}}^{*}(M) = \Lambda_{\mathcal{E}}(M) = M$ as M is already block-diagonal (with respect to $\mathcal{E}$). We thus have

$$\Gamma_{\Lambda,C}[P] = P \star \left(\bigoplus_{i=1}^{n}[P]_i^{-1} \# [C]_i \right)^2 \tag{5.32}$$

as claimed. To see that it is a feasible element, we apply $\Lambda_{\mathcal{E}}$ to this matrix, which removes all the non-diagonal submatrices, which gives us

$$\bigoplus_{i=1}^{n}[P]_i^{-1} \# [C]_i \cdot P_i \cdot [P]_i^{-1} \# [C]_i = \bigoplus_{i=1}^{n}[C]_i = C. \tag{5.33}$$

This completes the proof. □

By choosing $\mathcal{E} = \{|i\rangle\langle i|\}_{i\in[d]}$ to be the rank-one projectors corresponding to the computational basis vectors, one gets the closed-form for projection with respect to the completely dephasing channels.

Corollary 5.2.9 (Projection with respect to completely dephasing map)**.** *Let $\mathcal{H} = \mathbb{C}^d$ and $\Delta : \mathbb{P}_{\mathcal{H}} \to \mathbb{P}_{\mathcal{H}}$ be the completely dephasing map. For a diagonal positive matrix $C := \sum_{i=1}^{d} C_i |i\rangle\langle i| \in \mathbb{P}_{\mathcal{H}}$, the Gamma map is given by*

$$\Gamma_{\Delta,C}[P] = \sum_{i,j=1}^{d} \sqrt{\frac{C_i C_j}{P_{ii} P_{jj}}} P_{ij} |i\rangle\langle j| \qquad \in \Delta^{-1}[C], \tag{5.34}$$

for any $P^{+} \in \mathbb{P}_{\mathcal{H}}$.

Proof. Choose $E_i = |i\rangle\langle i|$ for $i \in [d]$ in Theorem 5.2.8. □

This corresponds to the Bures projection onto the set of positive matrices with a specified diagonal.

5.2.3 Projective measurement projection

The final type of channels we consider is projective measurements (viewed as channels) (see Watrous [Wat18, Section 2.3]). Let $\mathcal{E} = \{E_i\}_{i\in[n]}$ be a collection of mutually orthogonal projectors that sum to identity. The associated measurement channel $\mathrm{M}_\mathcal{E} : \mathbb{H}_d \to \mathbb{R}^n$ and its adjoint $\mathrm{M}_\mathcal{E}^* : \mathbb{R}^n \to \mathbb{H}_d$ are given by

$$\mathrm{M}_\mathcal{E}(H) := \sum_{i=1}^n \langle E_i, H\rangle\, |i\rangle\langle i| \quad \text{and} \quad \mathrm{M}_\mathcal{E}^*(v) = \sum_{i=1}^n v_i E_i, \tag{5.35}$$

for $H \in \mathbb{H}_d$ and $v \in \mathbb{R}^n$. We now show that the corresponding Gamma map yields a feasible state and thus also the closed-form for Bures projection.

Theorem 5.2.10. *Let $\mathrm{M}_\mathcal{E}$ be the measurement channel associated with the projectors $\mathcal{E} := \{E_i\}_{i\in[n]} \subset \mathbb{P}_\mathcal{H}$. Let $c \in \mathbb{R}^n_+$ be a positive vector. Then*

$$\Gamma_{\mathrm{M}_\mathcal{E},c}[P] = P \star \left(\sum_{i=1}^n \sqrt{\frac{c_i}{\langle E_i, P\rangle}} E_i\right)^2 \qquad \in \mathrm{M}_\mathcal{E}^{-1}[c] \tag{5.36}$$

for any $P \in \mathbb{P}^+_\mathcal{H}$.

Proof. We first show the form of $\Gamma_{\mathrm{M}_\mathcal{E},c}[P]$ by unpacking the definition. To this end, observe that

$$\mathrm{M}_\mathcal{E}(P)^{-1} = \sum_{i=1}^n \langle P, E_i\rangle^{-1}\, |i\rangle\langle i| \quad \Longrightarrow \quad \mathrm{M}_\mathcal{E}(P)^{-1} \# c = \sum_{i=1}^n \sqrt{\frac{c_i}{\langle E_i, P\rangle}}\, |i\rangle\langle i|\,. \tag{5.37}$$

The image under the adjoint is then obtained as $\mathrm{M}_\mathcal{E}^*\left(\mathrm{M}_\mathcal{E}(P)^{-1}\#c\right) = \sum_{i=1}^n \sqrt{\frac{c_i}{\langle E_i,P\rangle}} E_i$, which leads to the form

$$\Gamma_{\mathrm{M}_\mathcal{E},c}[P] = P \star \left(\sum_{i=1}^n \sqrt{\frac{c_i}{\langle E_i, P\rangle}} E_i\right)^2 \tag{5.38}$$

as claimed. To show that $\Gamma_{\mathrm{M}_\mathcal{E},c}[P]$ is a feasible point, compute

$$\mathrm{M}_\mathcal{E}\left(\Gamma_{\mathrm{M}_\mathcal{E},c}[P]\right) = \sum_{i=1}^n \langle E_i, \Gamma_{\mathrm{M}_\mathcal{E},c}[P]\rangle\, |i\rangle\langle i| = \sum_{i=1}^n c_i\, |i\rangle\langle i| = c. \tag{5.39}$$

The second equality follows from

$$\langle E_i, \Gamma_{\mathrm{M}_{\mathcal{E}},c}[P]\rangle = \sum_{j,k=1}^{n} \sqrt{\frac{c_j c_k}{\langle E_j, P\rangle\langle E_k, P\rangle}}\,\mathrm{Tr}\left[E_i E_j P E_k\right] = \frac{c_i}{\langle E_i, P\rangle}\langle E_i, P\rangle = c_i, \tag{5.40}$$

where we have used $E_i E_j = \delta_{ij} E_i$ as $\mathcal{E} := \{E_i\}_{i\in[n]}$ form a collection of mutually orthogonal projectors. We have thus shown $\Gamma_{\Lambda_{\mathcal{E}},C}[P] \in \mathrm{M}_{\mathcal{E}}^{-1}[c]$ which, by Theorem 5.2.4, implies it is also the projection $\Pi_{\mathrm{M}_{\mathcal{E}},c}[P]$. This concludes the proof. □

5.2.4 Ensemble projection

The final type of projection we discuss is the projection of *ensembles*. Though this is a special case of the partial-trace projection, it is useful to study it separately. The setting is as follows.

Let $\mathcal{P} = (P_1, \ldots, P_n)$ be a tuple of positive definite matrices in $\mathbb{P}_d^+$ with $P := \sum_{i=1}^n P_i$. Let $Q \in \mathbb{P}_d^+$ and define

$$\mathrm{Dec}_n(Q) := \left\{ \mathcal{Q} := (Q_1, \ldots, Q_n) \in {\mathbb{P}_d}^{\times n} : \sum_{i=1}^{n} Q_i = Q \right\}. \tag{5.41}$$

That is, $\mathrm{Dec}_n(Q)$ is the set of all *n-decompositions* of Q. The problem of interest is then to *project* $\mathcal{P}$ to this compact and convex set, where the projection is measured in terms of the sum of fidelity (or equivalently sum of Bures distance) between respective elements. That is solve,

$$\operatorname*{argmin}_{\mathcal{Q}\in\mathrm{Dec}_n(Q)} \sum_{i=1}^{n} \mathrm{B}(P_i, Q_i) = \operatorname*{argmax}_{\mathcal{Q}\in\mathrm{Dec}_n(Q)} \sum_{i=1}^{n} \mathrm{F}(P_i, Q_i). \tag{5.42}$$

Note that choosing $Q = \mathbb{I}_d$ gives the projection to the set of n-outcome POVMs. Using a direct sum construction and the closed-form for partial trace projection, one can show that the optimal ensemble is given by

$$Q_i := P_i \star (P^{-1}\#Q)^2 \quad \text{for each } i \in [n]. \tag{5.43}$$

Before we prove this, we first state a useful lemma.

Lemma 5.2.11. *Let* $\mathcal{X} = \mathbb{C}^n$ *and* $\mathcal{Y} = \mathbb{C}^d$ *and* $\{P_i\}_{i\in[n]}, \{Q_i\}_{i\in[n]} \subset \mathbb{P}_{\mathcal{Y}}$. *Define* $\mathrm{P}, \mathrm{Q} \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$ *to be the block-diagonals* $\mathrm{P} := \sum_{i=1}^n |i\rangle\langle i| \otimes P_i$ *and* $\mathrm{Q} := \sum_{i=1}^n |i\rangle\langle i| \otimes Q_i$. *Then*

$$\mathrm{F}(\mathrm{P}, \mathrm{Q}) = \sum_{i=1}^n \mathrm{F}(P_i, Q_i) \quad \textit{and} \quad \mathrm{B}(\mathrm{P}, \mathrm{Q}) = \sum_{i=1}^n \mathrm{B}(P_i, Q_i) \tag{5.44}$$

Proof. Proof directly follows from additivity of fidelity under direct sum [Wat18; Wil13]. □

We will solve the problem by showing that finding the projection of the ensemble $\mathcal{P}$ is equivalent to projecting the direct-sum matrix P to the feasible set $\mathrm{Tr}_{\mathcal{X}}^{-1}[Q]$ (note that $\mathrm{Tr}_{\mathcal{X}} : \mathbb{M}_{\mathcal{X}\otimes\mathcal{Y}} \to \mathbb{M}_{\mathcal{Y}}$) of states of the form:

$$\mathrm{Tr}_{\mathcal{X}}^{-1}[Q] := \{\mathrm{Q} \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}} : \mathrm{Tr}_{\mathcal{X}}[\mathrm{Q}] = Q\}. \tag{5.45}$$

Observe that we showed the equivalence of fidelity (and Bures distance) between the ensemble and direct-sum matrix only for block-diagonal matrices. Although the feasible set defined in Eq. (5.45) also contains elements that are not block-diagonal, data processing inequality for fidelity dictates that it suffices to consider block-diagonal matrices. To see this choose $\mathrm{Q} := \sum_{i,j=1}^n |i\rangle\langle j| \otimes Q_{ij} \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$ arbitrarily and $P := \sum_{i=1}^n |i\rangle\langle i| \otimes P_i$ to be block-diagonal. Choose a pinching channel with elements $\{|i\rangle\langle i| \otimes \mathbb{I}_{\mathcal{Y}}\}_{i\in[n]}$. By DPI for fidelity, we have

$$\mathrm{F}\left(\sum_{i=1}^n |i\rangle\langle i| \otimes P_i, \sum_{i,j=1}^n |i\rangle\langle j| \otimes Q_{ij}\right) \leq \mathrm{F}\left(\sum_{i=1}^n |i\rangle\langle i| \otimes P_i, \sum_{i=1}^n |i\rangle\langle i| \otimes Q_{ii}\right). \tag{5.46}$$

Thus, although $\mathrm{Tr}_{\mathcal{X}}^{-1}[Q]$ has elements that are not block-diagonal (in the sense of P), the optimal value will be achieved at a block-diagonal matrix of compatible structure. If P is full-rank (which is equivalent to every $P_i \in \mathcal{P}$ being full-rank), then the optimal block-diagonal is unique (and full-rank) as well. The main result can be formulated as follows.

Theorem 5.2.12 (Ensemble projection)**.** *Let $\mathcal{P} = (P_1, \ldots, P_n) \subset \mathbb{P}_{\mathcal{Y}}^+$ be a fixed tuple of states and let $Q \in \mathbb{P}_{\mathcal{Y}}^+$ be a fixed matrix. Then the tuple $\mathcal{Q} = (Q_1, \ldots, Q_n) \subset \mathbb{P}_{\mathcal{Y}}^+$ summing to Q that maximizes the total fidelity $\sum_{i=1}^n \mathrm{F}(P_i, Q_i)$ with $\mathcal{P}$ is given by*

$$Q_i = P_i \star (P^{-1}\#Q)^2 = (P^{-1}\#Q)P_i(P^{-1}\#Q), \tag{5.47}$$

for each $i \in [n]$, where $P := \sum_{i=1}^n P_i$.

Proof. Let $\mathrm{P} := \sum_{i=1}^n |i\rangle\langle i| \otimes P_i \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$. Projecting P to the compact set $\mathrm{Tr}_{\mathcal{X}}^{-1}[Q]$, we get

$$\begin{aligned}
\hat{\Pi}_{\mathcal{Y},Q}[\mathrm{P}] &= \mathrm{P} \star_{\mathcal{Y}} ([\mathrm{P}]_{\mathcal{Y}}^{-1}\#Q)^2 \\
&= \mathrm{P} \star (\mathbb{I}_{\mathcal{X}} \otimes (P^{-1}\#Q))^2 \\
&= (\mathbb{I}_{\mathcal{X}} \otimes P^{-1}\#Q) \cdot \left(\sum_{i=1}^n |i\rangle\langle i| \otimes P_i \right) \cdot (\mathbb{I}_{\mathcal{X}} \otimes P^{-1}\#Q) \\
&= \sum_{i=1}^n |i\rangle\langle i| \otimes \left[P_i \star (P^{-1}\#Q)^2 \right],
\end{aligned} \tag{5.48}$$

where we used the fact that $[\mathrm{P}]_{\mathcal{Y}} = \mathrm{Tr}_{\mathcal{X}}[P] = \sum_{i=1}^n P_i = P$. Observe that even though the feasible set contains matrices that are not block-diagonal, the optimal feasible element is indeed block-diagonal as implied by data processing inequality. Picking the ensemble elements as the corresponding diagonal submatrices, we complete the proof. □

The above closed-form should find applications in optimization problems involving fidelity. For example, if we choose $\mathcal{P} \equiv (\rho_1, \ldots, \rho_n)$ to be a (possibly weighted) ensemble of states and $Q = \sigma$ to be a fixed density matrix, then the closed-form gives the closest ensemble to $\mathcal{P}$ which is a decomposition of Q. By the bijection between positive semidefinite matrices and weighted quantum states, this optimal ensemble is equivalent to an ensemble of weighted states.

Remark. We now note an interesting observation regarding the form of ensemble projection. Recall that the projection of the ensemble $\mathcal{P} := (P_i)_{i\in[n]}$ to the set of ensembles summing to Q is given by $\mathcal{Q} := (Q_i)_{i\in[n]}$ where

$$Q_i = P_i \star (P^{-1}\#Q)^2 \equiv \frac{P_i}{(Q^{-1}\#P)^2}. \tag{5.49}$$

Suppose we are in the commuting scenario and all the matrices involved commute. Then the relation reduces to

$$Q_i = Q \cdot \frac{P_i}{P}, \tag{5.50}$$

which is reminiscent of the *scaling* one does in the case of vectors, where to obtain a vector that sums to $\bar{q} \in \mathbb{R}_+$ from a vector $p \in \mathbb{R}^d_+$, one performs the scaling $p \mapsto \frac{\bar{q}}{\bar{p}} p$ for $\bar{p} := \sum_{i=1}^d p_i$.

Let us now return to the non-commutative setting and discuss a particularly interesting choice of Q—namely $Q = \mathbb{I}_y$. For this choice of Q, the ensemble is projected to the compact set of n-outcome measurements (POVMs). The closed-form for ensemble projection then yields the familiar *pretty good measurement* (PGM) as the following corollary reveals.

Corollary 5.2.13. *Let $\mathcal{P} = (P_1, \ldots, P_n) \subset \mathbb{P}_d^+$ be an ensemble and $P = \sum_{i=1}^n P_i$. The ensemble projection of $\mathcal{P}$ to the compact set of n-outcome POVMs (which is equivalent to $\mathrm{Dec}_n(\mathbb{I}_d)$) is given by*

$$Q_i = P_i \star P^{-1} = P^{-\frac{1}{2}} P_i P^{-\frac{1}{2}}, \tag{5.51}$$

which is the 'pretty good measurement' associated with the ensemble $\mathcal{P}$.

Proof. The proof directly follows from Theorem 5.2.12 on setting $Q = \mathbb{I}_d$ and recalling that $P \# \mathbb{I} = \sqrt{P}$ for any $P \geq 0$. □

As we will discuss in Section 5.4.4, this result endows the PGM with new operational and geometric interpretations. We now proceed to discuss the various applications and manifestations of the closed-form for the Bures projection. All our examples pertain to the partial trace channel. Before diving into these examples, it is helpful to understand a certain isomorphism between bipartite states and pairs of channels and input states, which can be viewed as an extension of the Choi–Jamiołkowski isomorphism.

5.3 A note on Channel-state duality

Recall that the Choi–Jamiołkowski isomorphism associates to every bipartite positive semidefinite matrix $P \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$ a completely positive map in $\mathrm{CP}(\mathcal{X}, \mathcal{Y})$. If the *input marginal* of the bipartite PSD matrix equals the identity matrix ($P_{\mathcal{X}} = \mathbb{I}_{\mathcal{X}}$), then the associated CP map is trace-preserving as well. The aim of this section is to demonstrate a variant of this isomorphism by associating to every bipartite positive definite matrix[2] $P \in \mathbb{P}^+_{\mathcal{X}\otimes\mathcal{Y}}$ not a CP map, but instead a pair (Φ_P, ρ_P) of a quantum channel (CPT) map $\Phi_P \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y})$ and a (typically unnormalized) input state $\rho_P \in \mathbb{P}_{\mathcal{X}}$.

[2]The association can be extended to positive semidefinite matrices as well by appropriate restrictions to support.

We first note that such an association [CM20] and slight variants [Lei06; LS13] can be derived from results in the literature. Let us now define this association.

Theorem 5.3.1. *To each $P \in \mathbb{P}^+_{\mathcal{X}\otimes\mathcal{Y}}$, one can bijectively associate a channel $\Phi_P \in \mathrm{CPT}(\mathcal{X},\mathcal{Y})$ and a (possibly unnormalized) input state $\rho_P \in \mathbb{P}^+_{\mathcal{X}}$ such that the following relations hold:*

$$\rho_P = [P]_{\mathcal{X}}, \quad J(\Phi_P) = \hat{\Pi}_{\mathcal{X},\mathbb{I}_{\mathcal{X}}}[P], \quad \text{and} \quad [P]_{\mathcal{Y}} = \Phi_P(\rho_P^{\intercal}). \tag{5.52}$$

where the transpose is taken in the basis used to define the Choi matrix. Moreover, for any pair $P, Q \in \mathbb{P}^+_{\mathcal{X}\otimes\mathcal{Y}}$, we have

$$P = Q \quad \text{if and only if} \quad \rho_P = \rho_Q \text{ and } \Phi_P = \Phi_Q, \tag{5.53}$$

where $\rho_Q \in \mathbb{P}^+_{\mathcal{Y}}$ and $\Phi_Q \in \mathrm{CPT}(\mathcal{X},\mathcal{Y})$ are defined analogously to Eq. (5.52).

Proof. First, observe that

$$\rho_P := \mathrm{Tr}_{\mathcal{Y}}[P] \in \mathbb{P}^+_{\mathcal{X}} \quad \text{and} \quad J(\Phi_P) := \Pi_{\mathrm{CPT}(\mathcal{X},\mathcal{Y})}[P] = P \star_{\mathcal{X}} [P]_{\mathcal{X}}^{-1} \in \mathrm{J}_{\mathrm{CPT}}(\mathcal{X},\mathcal{Y}). \tag{5.54}$$

Thus (Φ_P, ρ_P) constitute a valid pair of channel and input state. The fact that $P_{\mathcal{Y}} = \Phi_P(\rho_P^{\intercal})$ immediately follows from properties of Choi states. Indeed we have

$$\Phi_P(\rho_P^{\intercal}) = [J(\Phi_P) \star_{\mathcal{X}} (\rho_P^{\intercal})^{\intercal}]_{\mathcal{Y}} = [(P \star_{\mathcal{X}} P_{\mathcal{X}}^{-1}) \star_{\mathcal{X}} P_{\mathcal{X}}]_{\mathcal{Y}} = [P]_{\mathcal{Y}}, \tag{5.55}$$

where the first equality is by Eq. (2.42). In the second equality, we used the definitions of $J(\Phi)_P$ and ρ_P.

Let us now prove the bijection. For the forward direction, we first prove the equality of the input states. Indeed we have

$$P = Q \quad \Longrightarrow \quad \rho_P := [P]_{\mathcal{X}} = [Q]_{\mathcal{X}} =: \rho_Q. \tag{5.56}$$

For the equality of channels, observe that

$$J(\Phi_P) := P \star_{\mathcal{X}} P_{\mathcal{X}}^{-1} = Q \star_{\mathcal{X}} Q_{\mathcal{X}}^{-1} =: J(\Phi_Q). \tag{5.57}$$

Here the equality follows from the equality of corresponding terms. The reverse direction is also straightforward as $\rho_P = \rho_Q$ and $\Phi_P = \Phi_Q$ implies

$$P = J(\Phi_P) \star_{\mathcal{X}} \rho_P = J(\Phi_Q) \star_{\mathcal{X}} \rho_Q = Q, \tag{5.58}$$

where the first and last equalities are obtained by inverting the form of Choi

projection and the middle equality follows by assumption. □

Essentially one can uniquely decompose any bipartite positive definite matrix as the joint state of a channel and an input state, and when performing a projection of an arbitrary PSD matrix onto the set of CPT maps, one is actually 'dividing out' the input state.

Chruściński and Matsuoka [CM20], Leifer [Lei06], and Leifer and Spekkens [LS13] all have remarked upon (up to slight variations) such decompositions of bipartite matrices into channels and input states as a quantum version of the application of a classical channel (which is a conditional distribution) on an input probability distribution.

Our result would also provide an information-geometric underpinning to this formalism as the operations correspond to appropriate Bures projection. The following corollary is essentially a restatement of the previous theorem which we write for emphasis.

Corollary 5.3.2. *Any $P \in \mathbb{P}^{+}_{\mathcal{X}\otimes\mathcal{Y}}$ can be uniquely written as*

$$P = J(\Phi_P) \star_{\mathcal{X}} \rho_P, \tag{5.59}$$

for a channel $\Phi_P \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y})$ and an input state $\rho_P \in \mathbb{P}_{\mathcal{X}}$ such that

$$J(\Phi_P) = \Pi_{\mathrm{CPT}(\mathcal{X},\mathcal{Y})}[P], \quad \rho_P = [P]_{\mathcal{X}}, \quad \text{and} \quad \Phi_P[\rho_P^{\intercal}] = [P]_{\mathcal{Y}}. \tag{5.60}$$

Conversely, for a channel $\Phi \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y})$ and an input state $\rho \in \mathbb{P}_{\mathcal{X}}$, we will call $J(\Phi) \star_{\mathcal{X}} \rho$ the joint state of the channel-state pair (Φ, ρ).

We now discuss applications of Bures projection.

5.4 Applications and manifestations of Bures projection

We now look at various applications and manifestations of the closed-form for Bures projection. Most of the applications we discuss here pertain to marginal projection (Bures projection with respect to the partial trace channel).

5.4.1 Projection for optimization problems

We first discuss a general application for closed forms for projections. An important subfield of optimization is constrained optimization, where one is interested

in finding the optimal point satisfying some constraint [BV04]. A generic problem of this kind can be written as

$$\begin{aligned} \text{minimize}: \quad & f(x) \\ \text{subject to}: \quad & x \in \mathcal{C}. \end{aligned} \tag{5.61}$$

where $\mathcal{C}$ is the feasible set of the constraint.

Semidefinite programs are constrained optimizations with a linear objective and a spectrahedral feasible set. Quantum information often involves optimization problems over the spectrahedral sets of quantum states or channels. While the method of Lagrange multipliers handles certain cases, more complex problems demand other approaches. For unconstrained problems, gradient descent iteratively updates points until a stopping condition is met, such as reaching a maximum iteration count or a small objective difference between steps [BV04].

A key issue with vanilla gradient descent in constrained settings is that it may generate infeasible points. For some problems, one can ensure iterates remain within the feasible set by modifying the geometry of the problem. For example, the RGD method discussed in Chapter 3 ensures that every iterate is positive semidefinite, which corresponds to the feasible set. Another popular method is mirror descent [NY83; HSF24; YCL22; YFT19], which has found success in the quantum setting as well.

A more common fix is *projected gradient descent* [Bub15], where after each step we *project* back to the feasible set:

$$x_{t+1} = \Pi_{\mathcal{C}}[x_t - \eta_t \nabla f(x_t)], \tag{5.62}$$

where the projection operation is defined as

$$\Pi_{\mathcal{C}}(x) = \operatorname*{argmin}_{y \in \mathcal{C}} \mathrm{D}(y, x), \tag{5.63}$$

where D is typically a divergence or a squared distance, with squared Euclidean distance (which is both) being a standard choice. For projected GD to be efficient, we must be able to compute the projection efficiently, which is possible for certain sets and distances. Of course, closed forms for such projections are of immense use in both theory and practice.

Our results provide closed forms for Bures projections to various sets of interest. The closed-form for marginal projection gives the projection to the sets of Choi matrices of CPT maps and CPU maps, the set of multipartite states with a given marginal, measurement ensembles, or ensembles of matrices that sum to a fixed matrix. The closed form for pinching channels give projection to matrices

with a given image under the pinching channel. For the completely dephasing map, this gives the projection to the set of positive semidefinite matrices with a given diagonal. Similarly, the closed-form for projective measurements gives projections to specific sets of interest defined by expected measurement outcomes.

As an example of marginal projection consider the following optimization problem over Choi matrices of quantum channels [CGW21].

$$\begin{aligned} \text{minimize}: \quad & f(Q) \\ \text{subject to}: \quad & Q \in J_{\mathrm{CPT}}(\mathcal{X}, \mathcal{Y}), \end{aligned} \tag{5.64}$$

for some function f. If one chooses a projected gradient descent approach to this problem, one can use our closed form for the projection after every step. We now turn our attention to more concrete applications.

5.4.2 Choi projection for channel tomography

The first concrete application we present is to use the closed-form for the projection of CP maps to CPT maps (at the level of Choi matrices) in the context of tomography of quantum channels [MRL08; Sur+22; GCC16]. We now detail two specific examples of channel tomography where Choi projection finds application.

Bures Projected Least Squares Tomography. Surawy-Stepney, Kahn, Kueng, and Guta [Sur+22] introduce and analyze the *Projected least squares* (PLS) process tomography protocol, where they estimate Choi matrix of a channel (from experimental data) as follows.

Let $\mathcal{X} = \mathbb{C}^d$ and $\mathcal{Y} = \mathbb{C}^{d'}$ and suppose $\Phi \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y})$ is the channel one wants to perform tomography over. The idea is to first construct a state that faithfully represents the channel, and one way to obtain such a state is to act $\mathrm{Id}_{\mathcal{X}} \otimes \Phi$ on the Bell state $\frac{1}{d}\Omega = \frac{1}{d}|\omega\rangle\langle\omega| \in \mathbb{D}_{\mathcal{X}\otimes\mathcal{X}}$, to get a rescaled version of the Choi matrix:

$$\frac{1}{d}(\mathrm{Id}_{\mathcal{X}} \otimes \Phi)(\Omega) = \frac{1}{d} J(\Phi) \in \mathbb{D}_{\mathcal{X}\otimes\mathcal{Y}}. \tag{5.65}$$

Once we have such a state, we can perform standard state tomography to compute a least-square estimate of this state. This constitutes the first step of the protocol.

The least-square estimate would be a Hermitian matrix $\hat{\rho}_{\mathrm{LS}} \in \mathbb{H}_{\mathcal{X}\otimes\mathcal{Y}}$, with the corresponding linear map Φ_{LS} being a Hermitian-preserving map. The idea is to then *classically* obtain the 'nearest' Choi matrix of a CPT map from $\hat{\rho}_{\mathrm{LS}}$.

1. Encode the channel Φ to a state ρ and perform tomography to obtain measurement data. Construct the Hermitian least-squares estimate $\hat{\rho}_{\mathrm{LS}} \in \mathbb{H}_{\mathcal{X}\otimes\mathcal{Y}}$. The corresponding linear map $\hat{\Phi}_{\mathrm{LS}} = J^{-1}(\hat{\rho}_{\mathrm{LS}})$ is typically neither completely positive nor trace-preserving and is only Hermitian-preserving.

2. Euclidean-project $J(\hat{\Phi}_{\mathrm{LS}})$ to the set of positive semidefinite matrices. This is done by a simple optimization process involving the eigenvalues of $J(\hat{\Phi}_{\mathrm{LS}})$. We now have a Choi matrix $J(\hat{\Phi}_{\mathrm{CP}}) \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$ corresponding to a CP map, but not trace-preserving.

3. Euclidean-project $J(\hat{\Phi}_{\mathrm{CP}})$ to the affine subspace of matrices with marginal $\frac{1}{d}\mathbb{I}_{\mathcal{X}}$ on the input space. We now have a (rescaled) trace-preserving map, but this second projection can take the estimate out of the PSD cone, thereby losing the CP property.

4. Repeat Steps 2 and 3 iteratively until either convergence is achieved or a maximum step count is reached.

Step 1 is essentially a state tomography protocol [Guţ+20]. The tricky parts are Steps 2 and 3, where we need to project onto the intersection of two convex sets. This is done by projecting, in an alternating manner, onto the PSD cone and the affine subspace of matrices whose input marginal equals $\frac{1}{d}\mathbb{I}_{\mathcal{X}}$. The CP projection is done via an optimization over the eigenvalues which requires diagonalization [SGS12] of a Hermitian matrix of size $dd' \times dd'$, which has cost $O(d^3d'^3)$. The second projection, which corresponds to trace-preserving property, can be done via a closed form of lesser complexity as it involves only partial trace and tensoring with identity (apart from matrix sum).

However, one will have to do multiple iterations of the two steps, and thus multiple eigendecompositions, each with cost $O(d^3d'^3) \approx O(d^6)$, with $d = 2^n$ for n-qubit systems. Thus the cost can get prohibitive quite fast even for moderate system sizes.

Instead, we propose the *Bures Projected Least Squares* where do steps 1 and 2 as before to get a PSD estimate for the Choi matrix, which corresponds to the Choi matrix of a CP map. The cost until now is $O(d^6)$, which corresponds to the cost of eigendecomposition. However, we now perform a single Bures projection Π_{CPT} to get a Choi state corresponding to a valid quantum channel, thereby completing the protocol. This step involves computing the marginal, its inverse-square root ($O(d^3)$), followed by two matrix multiplications ($O(d^2d'^2)$). Since we only have to perform each of these steps exactly once, this protocol would be much faster than the (Euclidean) PLS method.

Moreover, the protocol is geometrically well-motivated. We first construct a least-square estimate, which is Hermitian. We then project it to the set of PSD matrices with respect to Euclidean distance, which is a natural (and Riemannian) distance on Hermitian matrices. Once we have entered the PSD cone, we now project with respect to Bures distance to the set of Choi matrices corresponding to

CPT maps. This projection is also natural as we have seen that Bures distance is often a better choice over Euclidean distance over positive semidefinite matrices.

Bayesian tomography of quantum channels. We now discuss how Choi projection can be used for Bayesian tomography of quantum channels. This would be an extension of results in Afham, Kueng, and Ferrie [AKF22] (the results discussed in Chapter 3), where we studied methods to perform Bayesian tomography of quantum states.

The problem can be formulated as follows. Let $\mathcal{Q} = \{Q_1, \ldots, Q_n\}$ be a finite collection of Choi matrices (corresponding to CPT maps) and $w = (w_1, \ldots, w_n)$ be (probability) weights associated with the ensemble $\mathcal{Q}$. We are interested in finding

$$Q_\sharp := \operatorname*{argmax}_{Q \in \mathrm{J}_{\mathrm{CPT}}(\mathcal{X},\mathcal{Y})} \sum_{i=1}^{n} w_i \, \mathrm{F}(Q_i, Q) \tag{5.66}$$

One can formulate this problem as a *Projected Riemannian Gradient Descent* where the projection is onto the set of (Choi matrices of) quantum channels. Indeed this can be seen as a generalization of results from Chapter 3, where we generalize the Bures projection onto density matrices to the Bures projection onto Choi matrices of quantum channels. Numerical experiments suggest a fixed-point algorithm similar to the one in Chapter 3 (with an appropriate replacement of the projection operation) would work.

5.4.3 Random state and ensemble generation

The next application we discuss is the random generation of various states and ensembles of interest in quantum information. We first provide a short primer on Ginibre [Gin65; AGZ10] and Wishart [Wis28] distribution.

Let $\mathcal{N}(0, 1)$ denote the standard normal distribution (zero mean and unit variance). The *Ginibre distribution* $\mathcal{G}(d, r)$ is the random distribution of $(d \times r)$ complex matrices whose entries are independently and identically sampled (i.i.d.) from the complex standard normal distribution. That is,

$$G \sim \mathcal{G}(d \times r) \implies G_{j,k} \sim \mathcal{N}(0, 1) + i\mathcal{N}(0, 1), \tag{5.67}$$

for each $j \in [d]$ and $k \in [r]$. The *Wishart distribution* $\mathcal{W}(d, r)$ is the distribution of d-dimensional and r-rank positive semidefinite matrices $W = GG^*$ where $G \sim \mathcal{G}(d, r)$.

We now describe some methods of generating random matrices and ensembles of interest in quantum information.

Random quantum state generation. In Zyczkowski and Sommers [ZS01] (also see [Ben98]), a protocol to generate d-dimensional and r-rank quantum

states is presented as follows.

First sample a positive semidefinite matrix W from the Wishart distribution $\mathcal{W}(d, r)$. Then perform the trace-normalization $W \mapsto W/\operatorname{Tr}[W]$, which yields a random density matrix of required characteristics. Al Osipov, Sommers, and Życzkowski [ASŻ10] discusses a protocol to generate full-rank random states according to the Bures measure by first sampling $W \sim \mathcal{W}(d, d)$ and sampling a Haar unitary $U \in \mathrm{U}_d$, and then computing

$$P = (\mathbb{I}_d + U)W(\mathbb{I}_d + U^*) \in \mathbb{P}_d^+. \tag{5.68}$$

Trace normalize $P \mapsto P/\operatorname{Tr}[P]$ to obtain a full-rank density matrix distributed according to the Bures measure.

Observe that both of these methods can be seen as the construction of a random positive matrix followed by trace-normalization, which, as we have discussed, can be seen as the Bures projection to the set of density matrices.

Random quantum processes generation. Bruzda, Cappellini, Sommers, and Życzkowski [Bru+09] study methods of generating Choi matrices corresponding to random quantum channels. To generate a Choi matrix corresponding to a random quantum channel, one samples $W \sim \mathcal{W}(d, r)$ and *normalizes it* by the operation

$$W \mapsto W \star_x [W]_{\mathrm{x}}^{-1}. \tag{5.69}$$

This operation has been used independently in [AS08] for numerical generation of Choi matrices of random CPT maps and CPU maps, where it was termed as a *projection*, although it was not proven to be a projection in the sense we use it. Also see [Kuk+21], where the authors study multiple ways of generating random quantum processes at the levels of Choi matrices, Kraus operators, and the Stinespring representation and equivalence between these methods. Nechita and Park [NP24] uses the same 'normalization' as an intermediate state to generate random *covariant quantum channels*.

Knowing the form of the marginal projection, one can now identify these methods as first generating a random positive semidefinite matrix followed by marginal projection to the set of CPT maps and CPU maps.

Random POVM generation. Heinosaari, Jivulescu, and Nechita [HJN20] define the *Wishart-random POVM* of parameters $(d, n, r_1, \ldots, r_n)$ by independently sampling $P_i \sim \mathcal{W}(d, r_i)$ for $i \in [n]$ and *normalizing* it as $P_i \mapsto P_i \star P^{-1}$, where $P := \sum_{i=1}^n P_i$. As we have seen in Corollary 5.2.13, this is equivalent to computing the PGM associated with the ensemble $\mathcal{P} = (P_1, \ldots, P_n)$, which is

equivalent to the ensemble projection to the set of n-outcome POVMs.

Having studied the different closed forms for Bures projection in Section 5.2, we can now identify all the above methods as a two-step '*generate and project*' protocol defined as follows.

1. Generate a random PSD matrix or an ensemble of PSD matrices.
2. Perform the appropriate Bures projection.

The above protocols correspond to trace projection, marginal projection to Choi matrices of CPT and CPU maps, and ensemble projection. One can now study the analogous projections for the other closed forms we have described. In particular, we can randomly generate the following objects.

1. Random states with an arbitrary fixed-marginal (using marginal projection).
2. Random ensembles that sum to a given positive matrix.
3. Random states with fixed (block) diagonals, or more generally, random states that produce a given output under a given pinching channel.
4. Random states that with a fixed expected outcome for a projective measurement.

We conclude this section with the following remark. By showing that these 'natural' normalizations have a valid operational interpretation as Bures projections onto the corresponding sets, our results also exemplify the importance of Bures geometry in quantum information.

5.4.4 Pretty good measurement as Bures projection

The pretty good measurement (PGM), introduced in [Bel75] (also see [Hol78]) and named so in [HW94], is a way of constructing a POVM from an ensemble of quantum states. It has implications in the task of *quantum state discrimination* (QSD) [Hel69], where it performs near-optimally [BK02; Wat18]. For certain structured instances of QSD, the PGM is optimal. This includes variants of the hidden subgroup problem [BCD05; MR05; HKK06], port-based teleportation [Led22], and state discrimination problems with a high degree of symmetry [EMV04; Zho+25; DP15]. For a detailed understanding of the problem of state discrimination, we refer to [BK15; BC09]. We also note the existence of

variants of PGM, which can be seen as the PGM associated with the ensemble $\mathcal{P}^\alpha \equiv (P^\alpha)_{i\in[n]}$, for $\alpha \geq 1$ [Tys09a; Tys09b].

Other applications of PGM include proofs of certain PAC-learning problems [AD18] and as a decoder in quantum coding protocols [Che23; BG14]. See [Gil+22] for a quantum algorithm for PGM (and Petz recovery map). Thus it is reasonable to say that PGM is a versatile object finding applications in various subfields of quantum information.

We now give novel geometric and operational interpretations for the pretty good measurement. Recall that in Corollary 5.2.13, we showed that the Bures projection of an ensemble to the set POVMs is the PGM associated with the ensemble. That is, if $\mathcal{P} = (P_1, \ldots, P_n) \subset \mathbb{P}_\mathfrak{X}$ is an ensemble with $P := \sum_{i=1}^n P_i$, then the closest n-outcome POVM to $\mathcal{P}$ is given by $\mathcal{E} := (E_1, \ldots, E_n)$ where $E_i = P_i \star P^{-1}$. Indeed $\mathcal{E}$ is the pretty good measurement associated with the ensemble $\mathcal{P}$. We now discuss some interpretations and implications of this result.

Firstly, as the corollary directly implies, the pretty good measurement is the closest POVM to the given ensemble in terms of fidelity/squared Bures distance.

A second implication is for the quantum state discrimination problem, whose setting we now describe. We are given a ensemble of states $\mathcal{R} := (\rho_1, \ldots, \rho_n) \subset \mathbb{D}_\mathfrak{X}$ with associated probability weights $w := (w_1, \ldots, w_n) \in \Delta_n$. Indeed this can be seen as a weighted ensemble $\mathcal{P} := (P_1, \ldots, P_n)$ with $P_i := w_i\rho_i$ for each $i \in [n]$. The state discrimination problem can be stated as follows:

$$\begin{aligned} \text{maximize :} \quad & \sum_{i=1}^n \langle P_i, E_i \rangle \\ \text{subject to :} \quad & \mathcal{E} := (E_1, \ldots, E_n) \text{ is a POVM.} \end{aligned} \tag{5.70}$$

That is, find the POVM $\mathcal{E}$ that *aligns* the most with the ensemble $\mathcal{P}$, where the alignment is measured in terms of the HS inner product. By Born's rule, this is equivalent to maximizing the success probability of the task where we want to identify the label of state from the measurement outcome.

There is no known closed-form for the optimal POVM $\mathcal{E}_{\text{opt}}$ that solves the above optimization problem except in the case of $n = 2$ or highly structured ensembles [EMV04; Zho+25; DP15]. However, it can be phrased as an SDP and thus solved numerically [YKL75; EMV03; Wat18]. Although there is no closed-form for the *optimal* POVM, there is a closed-form solution for a POVM that performs *pretty good*. Indeed this is the pretty good measurement $\mathcal{E}_{\text{PGM}}$ associated with the ensemble $\mathcal{P}$ whose performance is no-worse than the square of the optimal POVM [BK02; Wat18]:

$$\eta_{\text{opt}}^2 \leq \eta_{\text{PGM}} \leq \eta_{\text{opt}}, \tag{5.71}$$

where η_{opt} is the optimal success probability (the solution to Eq. (5.70)) and $\eta_{\text{PGM}} = \sum_{i=1}^{n} \langle P_i \star P^{-1}, P_i \rangle$ is the success probability of the pretty good measurement $\mathcal{E}_{\text{PGM}}$.

Now consider a variant of the problem where we replace the HS inner product with fidelity:

$$\begin{aligned} \text{maximize :} \quad & \sum_{i=1}^{n} \mathrm{F}(P_i, E_i) \\ \text{subject to :} \quad & \mathcal{E} := (E_1, \ldots, E_n) \text{ is a POVM.} \end{aligned} \tag{5.72}$$

Indeed this is equivalent to the problem of ensemble projection, and the solution is given by the pretty good measurement $\mathcal{E}_{\text{PGM}}$. Thus, for this *fidelity-variant* of the state discrimination problem, we have the closed form for the optimal POVM. Hence for the fidelity-based state discrimination problem, the pretty good measurement is not just pretty good, but as good as it gets![3]

Now consider the scenario where all the states in the ensemble $\mathcal{P}$ are pure (rank-one). That is, $P_i = |p_i\rangle\langle p_i|$ for some complex vector $|p_i\rangle \in \mathcal{X}$ with the associated weight given by $w_i = \langle p_i, p_i \rangle$. In this case, we have $\langle P_i, E_i \rangle = \langle p_i, E_i p_i \rangle = \mathrm{F}(E_i, P_i)^2$, and thus the state discrimination task corresponds to maximizing total square fidelity, and also that fidelity corresponds to 'square-root' probability.[4]

Hence the fidelity-based state discrimination problem (Eq. (5.72)), is optimizing for the sum of the square root of probabilities, which indicates that the PGM, which is also called *square-root measurement*, is optimal for the sum of the square root of probabilities. Indeed, by invoking Uhlmann's theorem [Uhl76; Joz94] one can extend the above interpretation to states of arbitrary rank as well. Uhlmann's theorem states that for a pair of positive semidefinite matrices $P, Q \in \mathbb{P}_{\mathcal{X}}$, we have

$$\mathrm{F}(P, Q) = \max_{\substack{|p\rangle \in \mathrm{Pur}(P) \\ |q\rangle \in \mathrm{Pur}(Q)}} |\langle p, q \rangle|, \tag{5.73}$$

where $\mathrm{Pur}(P)$ denotes the set of all purifications of $P \in \mathbb{P}_{\mathcal{X}}$. The fidelity-based state discrimination can then be interpreted as

$$\begin{aligned} \text{maximize :} \quad & \sum_{i=1}^{n} |\langle u_i, v_i \rangle| \\ \text{subject to :} \quad & |u_i\rangle \in \mathrm{Pur}(P_i), \quad |v_i\rangle \in \mathrm{Pur}(E_i), \\ & \sum_{i=1}^{n} E_i = \mathbb{I}. \end{aligned} \tag{5.74}$$

[3] the words *as good as it gets* is borrowed from the title of [MR05] which shows that PGM is optimal for distinguishing conjugate hidden subgroups.

[4] Uhlmann originally defined fidelity as the square of the definition used in this thesis. He identified this quantity as *transition probability*.

Essentially now we are maximizing the sum of the square root of transition probabilities (overlap) between the purifications of the ensemble states and measurements. As we have seen, this is optimized by the PGM $\mathcal{E}_{\text{PGM}}$ associated with the ensemble $\mathcal{P}$.

To conclude, we have shown that the pretty good measurement is the Bures projection of the given ensemble of states. This also implies PGM is optimal for a variant of state discrimination where the objective function is total fidelity. Since fidelity corresponds to the square root of probabilities, the PGM actually optimizes for the sum of the square root of success probabilities, whereas in the standard state discrimination problem we are interested in maximizing the sum of probabilities. This points to the near-yet-sub-optimality of PGM.

5.4.5 Geometric interpretation of the Petz recovery map

The Petz recovery map [Pet86b; Pet88] (of which pretty-good measurement is a special case) is a central object in quantum information. It describes the saturation of DPI for relative entropy (quantum sufficiency) [Pet86b; Pet88; Hay+04], is seen as a quantum analog for Bayes theorem [LS13; CM20; PB23; BBS24], and has applications in error correction [BK02; NM10], channel capacity [BDL16; Hau+96] among others. Also see [CS22a] for results at the intersection of geometry and approximate Petz recovery map.

We now give a geometric interpretation to the Petz recovery map in terms of Bures projection. Let $\Phi \in \text{CPT}(\mathcal{X}, \mathcal{Y})$ be a channel and $\rho \in \mathbb{P}_{\mathcal{X}}$ be an input state. The Petz recovery map associated with this pair is $\tilde{\Phi}_\rho \in \text{CPT}(\mathcal{Y}, \mathcal{X})$ whose action on $Y \in \mathbb{M}_{\mathcal{Y}}$ is defined as

$$\tilde{\Phi}_\rho(Y) := \rho^{\frac{1}{2}} \left(\Phi^* \left(\Phi(\rho)^{-\frac{1}{2}} Y \Phi(\rho)^{-\frac{1}{2}} \right) \right) \rho^{\frac{1}{2}} = \left(\Phi^* \left(Y \star \Phi(\rho)^{-1} \right) \right) \star \rho. \tag{5.75}$$

The transpose of the Choi matrix of the Petz recovery map can be derived as [BBS24]

$$J(\tilde{\Phi}_\rho)^\intercal = J(\Phi) \star \left(\rho^\intercal \otimes \Phi(\rho)^{-1} \right) = \left(\sqrt{\rho^\intercal} \otimes \Phi(\rho)^{-\frac{1}{2}} \right) J(\Phi) \left(\sqrt{\rho^\intercal} \otimes \Phi(\rho)^{-\frac{1}{2}} \right). \tag{5.76}$$

By the transpose property $(A \star B)^\intercal = A^\intercal \star B^\intercal$ of the star product (Proposition 2.1.5) we have the Choi matrix of the Petz recovery map $\tilde{\Phi}_\rho$ as

$$J(\tilde{\Phi}_\rho) = J(\Phi)^\intercal \star \left(\rho \otimes (\Phi(\rho)^{-1})^\intercal \right). \tag{5.77}$$

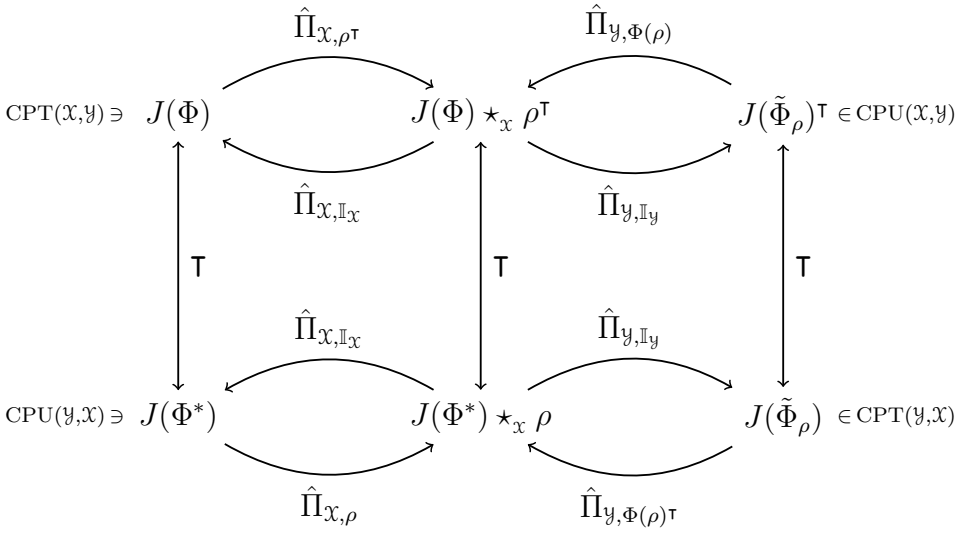

Figure 5.1: $J(\Phi) \in \mathbb{P}_{\mathcal{X}\otimes\mathcal{Y}}$ is the Choi matrix of the channel $\Phi \in \mathrm{CPT}(\mathcal{X},\mathcal{Y})$. For a reference state $\rho \in \mathbb{P}_\mathcal{X}$, $\tilde{\Phi}_\rho \in \mathrm{CPT}(\mathcal{Y},\mathcal{X})$ is the Petz recovery channel associated with the pair (Φ,ρ). By Proposition 2.3.7, we have $J(\Psi)^\intercal = J(\Psi^*) = J(\Psi^\intercal)$ for any completely positive map Ψ. The vertical bidirectional arrows indicate this relation under transpose. Thus, the Petz recovery map $\tilde{\Phi}_\rho$ can be reached from Φ via two projections and a transpose (through multiple *paths*). Alternatively, it can be reached from the joint state $J(\Phi) \star \rho^\intercal$ via a projection and a transpose.

One may readily verify the action as follows:

$$\begin{aligned}
\tilde{\Phi}_\rho(Y) &= [J(\tilde{\Phi}_\rho) \star_{\mathcal{Y}} Y^\intercal]_\mathcal{X} \\
&= [J(\Phi)^\intercal \star (\rho \otimes \Phi(\rho)^{-1\intercal}) \star_{\mathcal{Y}} Y^\intercal]_\mathcal{X} && \text{By Eq. (5.77)} \\
&= \left[J(\Phi)^\intercal \star_{\mathcal{Y}} \Phi(\rho)^{-1\intercal} \star_{\mathcal{Y}} Y^\intercal \star_{\mathcal{X}} \rho\right]_\mathcal{X} \\
&= [J(\Phi)^\intercal \star_{\mathcal{Y}} \Phi(\rho)^{-1\intercal} \star_{\mathcal{Y}} Y^\intercal]_\mathcal{X} \star \rho && [A \star_{\mathcal{X}} X]_\mathcal{X} = [A]_\mathcal{X} \star X \\
&= \left[J(\Phi^*) \star_{\mathcal{Y}} \left(\frac{Y}{\Phi(\rho)}\right)^\intercal\right]_\mathcal{X} \star \rho && \text{Cyclic property of partial trace} \\
&= \Phi^*\left(\frac{Y}{\Phi(\rho)}\right) \star \rho \\
&= \rho^{\frac{1}{2}} \Phi^*\left(\Phi(\rho)^{-\frac{1}{2}} Y \Phi(\rho)^{-\frac{1}{2}}\right) \rho^{\frac{1}{2}},
\end{aligned} \tag{5.78}$$

which is exactly the action of the Petz recovery map. In the above equation, for convenience (while taking star product), we have assumed $Y \in \mathbb{P}_\mathcal{Y}$. However, the verification can be easily extended to arbitrary $Y \in \mathbb{M}_\mathcal{Y}$.

One can now give a geometric interpretation to (the Choi matrix of) the Petz recovery map—it is the transpose marginal projection (to $J(\mathcal{Y},\mathcal{X})$) of the joint channel-state matrix $J(\Phi) \star_{\mathcal{X}} \rho^\intercal$.

Theorem 5.4.1. *Let $\Phi \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y})$ and $P \in \mathbb{D}_{\mathcal{X}}$. Then*

$$\Pi_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}[J(\Phi) \star_{\mathcal{X}} \rho^{\intercal}] = J(\tilde{\Phi}_{\rho})^{\intercal}, \tag{5.79}$$

where $\tilde{\Phi}_{\rho}$ is the Petz transpose map associated with the pair (Φ, ρ).

Proof. For $P \in \mathbb{P}^{+}_{\mathcal{X}\otimes\mathcal{Y}}$, the projection to $\mathrm{CPT}(\mathcal{Y}, \mathcal{X})$ (which corresponds to the set $\mathrm{J}(\mathcal{Y}, \mathcal{X})$) is defined as

$$\hat{\Pi}_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}[P] = P \star_{\mathcal{Y}} P_{\mathcal{Y}}^{-1}. \tag{5.80}$$

Choose $P = J(\Phi) \star_{\mathcal{X}} \rho^{\intercal}$. Observe that $[P]_{\mathcal{Y}} = \mathrm{Tr}_{\mathcal{X}}[J(\Phi) \star_{\mathcal{X}} \rho^{\intercal}] = \Phi(\rho)$. Then the above equation takes the form

$$\begin{aligned} \hat{\Pi}_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}[J(\Phi) \star_{\mathcal{X}} \rho^{\intercal}] = (J(\Phi) \star_{\mathcal{X}} \rho^{\intercal}) \star_{\mathcal{Y}} \Phi(\rho)^{-1} &= J(\Phi) \star \left(\rho^{\intercal} \otimes \Phi(\rho)^{-1}\right) \\ &= J(\tilde{\Phi}_{\rho})^{\intercal}. \end{aligned} \tag{5.81}$$

This concludes the proof. □

We now discuss the geometric interpretations of the above result. As Figure 5.1 indicates, there are multiple equivalent ways of phrasing the geometric interpretation of the Petz recovery map, all of which involve two Bures projection and a transpose.

1. First project $J(\Phi)$ to the joint state $J(\Phi) \star_{\mathcal{X}} \rho^{\intercal}$. Then project it to the set of CPT maps from $\mathbb{M}_{\mathcal{Y}}$ to $\mathbb{M}_{\mathcal{X}}$. Take the transpose of the Choi matrix to obtain (the Choi matrix of) the Petz recovery map $\tilde{\Phi}_{\rho}$.

2. Equivalently, begin with the Choi matrix of the adjoint Φ^{*} of Φ. Project it to the spectrahedron defined by $\mathcal{X}$-marginal being ρ to get $J(\Phi^{*}) \star_{\mathcal{X}} \rho$. Project it again to the set of quantum channels from $\mathbb{M}_{\mathcal{Y}}$ to $\mathbb{M}_{\mathcal{X}}$. The result is the (Choi matrix) of the Petz recovery map $\tilde{\Phi}_{\rho}$.

Essentially, up to a couple of transposes, the Petz recovery map $\tilde{\Phi}_{\rho}$ is the nearest (in terms of fidelity) 'reverse' channel to the extended joint state formed by (Φ, ρ).

Petz Recovery maps for transpose-preserving channels

Let us consider the above results for transpose-preserving channels. Recall that we say a channel Φ is transpose-preserving if $\Phi(X^{\intercal}) = \Phi(X)^{\intercal}$ for any $X \in \mathbb{P}_{\mathcal{X}}$. This is equivalent to $\Phi = \Phi^{\intercal} := \intercal \circ \Phi \circ \intercal$, which in turn is equivalent to $J(\Phi) = J(\Phi)^{\intercal}$, or that the Choi matrix is symmetric. See Section 2.3.2 for proofs and further details regarding these statements.

We now show that, if Φ is transpose-preserving, the Bures projection can be written as a composition of two projections (sans any transpose).

Proposition 5.4.2. *Let $\Phi \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y})$ be a transpose-preserving channel and $\rho \in \mathbb{P}_{\mathcal{X}}$ be an input reference state. It then holds*

$$J(\tilde{\Phi}_\rho) = \hat{\Pi}_{\mathcal{Y},\mathbb{I}_{\mathcal{Y}}} \circ \hat{\Pi}_{\mathcal{X},\rho}[J(\Phi)] = \Pi_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}[J(\Phi) \star_{\mathcal{X}} \rho]. \tag{5.82}$$

Proof. Let us evaluate the above expression. We have

$$\hat{\Pi}_{\mathcal{X},\rho}[J(\Phi)] = J(\Phi) \star_{\mathcal{X}} \rho. \tag{5.83}$$

The $\mathcal{Y}$ marginal of this joint state is $[J(\Phi) \star_{\mathcal{X}} \rho]_{\mathcal{Y}} = \Phi(\rho^\intercal) = \Phi(\rho)^\intercal$ where the second equality follows from the transpose-preserving property. Now compute the second projection.

$$\hat{\Pi}_{\mathcal{Y},\mathbb{I}_{\mathcal{Y}}}[J(\Phi) \star_{\mathcal{X}} \rho] = J(\Phi) \star_{\mathcal{X}} \rho \star_{\mathcal{Y}} \Phi(\rho)^{\intercal^{-1}} = J(\tilde{\Phi}_\rho). \tag{5.84}$$

This completes the proof. □

Thus, for transpose-preserving channels, the Petz recovery map is equivalent to two Bures projections of the original channel (at the level of Choi matrices). Or equivalently, it is the nearest 'reverse channel' to the extended joint-state $J(\Phi)\star_{\mathcal{X}}\rho$ of the pair (Φ, ρ). Some important channels, such as partial trace, depolarizing channel, and pinching channels (whose projectors have real representation in basis used to take transpose) are transpose-preserving.

5.4.6 Shorter proof for and geometric interpretation of quantum minimal change principle

The final application we discuss is a shorter proof and geometric interpretation for the central result of [BBS24]. Let $\Phi \in \mathrm{CPT}(\mathcal{X}, \mathcal{Y})$ be a channel, $\rho \in \mathbb{P}_{\mathcal{X}}$ be a *reference input*, and $\sigma \in \mathbb{P}_{\mathcal{Y}}$ be a *reference output*. The main problem of interest in [BBS24] is the following optimization problem.

$$\begin{aligned} &\text{maximize} : \mathrm{F}(J(\Phi) \star_{\mathcal{X}} \rho^\intercal, (J(\Psi) \star_{\mathcal{Y}} \sigma^\intercal)^\intercal) \\ &\text{subject to} : \Psi \in \mathrm{CPT}(\mathcal{Y}, \mathcal{X}). \end{aligned} \tag{5.85}$$

They obtain a closed-form solution for the optimal channel Ψ whose Choi state is given by

$$\begin{aligned} J(\Psi)^\intercal &= (\mathbb{I}_\mathcal{X} \otimes \sigma^{-\frac{1}{2}}) \left(\sqrt{\rho^\intercal} \otimes (\Phi(\rho)^{-1} \# \sigma)\right) J(\Phi) \left(\sqrt{\rho^\intercal} \otimes (\Phi(\rho)^{-1} \# \sigma)\right) (\mathbb{I}_\mathcal{X} \otimes \sigma^{-\frac{1}{2}}) \\ &= (J(\Phi) \star \left(\rho^\intercal \otimes (\Phi(\rho)^{-1} \# \sigma)^2\right) \star_\mathcal{Y} \sigma^{-1} \\ &= J(\Phi) \star_\mathcal{X} \rho^\intercal \star_\mathcal{Y} (\Phi(\rho)^{-1} \# \sigma)^2 \star_\mathcal{Y} \sigma^{-1}. \end{aligned} \tag{5.86}$$

Their proof technique involves formulating the above optimization in terms of Lagrange multipliers to derive the closed form.

We first provide a shorter proof, using marginal projection, for their main result followed by a geometric interpretation. By Theorem 5.3.1, we have that every bipartite positive definite matrix is uniquely associated with a channel-state pair, which can be written as

$$Q = J(\Psi)^\intercal \star_\mathcal{Y} \sigma \quad \iff \quad [Q]_\mathcal{Y} = \sigma \text{ and } \Pi_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}[Q] = J(\Psi)^\intercal, \tag{5.87}$$

We can then rewrite the main optimization problem (Eq. (5.85)) as

$$\max_{Q:[Q]_\mathcal{Y}=\sigma} \mathrm{F}(J(\Phi) \star_\mathcal{X} \rho^\intercal, Q). \tag{5.88}$$

This is the same as the Bures projection of the form $Q = \hat{\Pi}_{\mathcal{Y},\sigma}[J(\Phi) \star_\mathcal{X} \rho^\intercal]$. Denote $P \equiv J(\Phi) \star_\mathcal{X} \rho^\intercal$ for brevity. We will now use the closed form for marginal projection to obtain the optimal Q. We have

$$Q = \hat{\Pi}_{\mathcal{Y},\sigma}[P] = P \star_\mathcal{Y} ([P]_\mathcal{Y}^{-1} \# \sigma)^2 = P \star_\mathcal{Y} (\Phi(\rho)^{-1} \# \sigma)^2, \tag{5.89}$$

where we used the fact $[P]_\mathcal{Y} = \Phi(\rho)$. To obtain the associated channel $\Psi \in \mathrm{CPT}(\mathcal{Y}, \mathcal{X})$, we now simply marginal-project Q to $\mathrm{CPT}(\mathcal{Y}, \mathcal{X})$:

$$\begin{aligned} \Pi_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}[Q] = Q \star_\mathcal{Y} Q_\mathcal{Y}^{-1} &= (P \star_\mathcal{Y} (\Phi(\rho) \# \sigma^{-1})^2) \star_\mathcal{Y} \sigma^{-1} \\ &= J(\Phi) \star_\mathcal{X} \rho^\intercal \star_\mathcal{Y} \left(\Phi(\rho)^{-1} \# \sigma\right)^2 \star_\mathcal{Y} \sigma^{-1}, \end{aligned} \tag{5.90}$$

where we have used the fact that $[Q]_\mathcal{Y} = \sigma$. This is exactly the closed-form obtained for $J(\Psi)^\intercal$ in Eq. (5.86). We also see that if $[\Phi(\rho), \sigma] = 0$, we can write the above equation as

$$J(\Psi)^\intercal = \Pi_{\mathrm{CPT}(\mathcal{Y},\mathcal{X})}[Q] \equiv J(\Phi) \star_\mathcal{X} \rho^\intercal \star_\mathcal{Y} \Phi(\rho)^{-1}, \tag{5.91}$$

which is the transpose of the Choi matrix of the Petz recovery map.

Thus, using the closed-form for marginal projection, we have a shorter proof

for the central result of [BBS24], which originally was derived using the method of Lagrange multipliers. Indeed this also gives their result the following geometric interpretation. The channel of interest is the (transpose of Choi matrix of) the nearest channel to the projection of the joint state $J(\Phi)\star_{\mathcal{X}}\rho^{\intercal}$ to the spectrahedron defined by $\mathcal{Y}$-marginal as the *output* reference state σ.

5.5 Non-contractivity of Bures projection

An important caveat, particularly in constrained optimization tasks, is that Bures projections onto convex sets need not always be *contractive*. We say that a projection is contractive if the projection brings points closer. That is, $\mathrm{d}(x,y) \geq \mathrm{d}(\Pi(x),\Pi(y))$. The easiest way to see projections with respect to Bures distance are not necessarily contractive is by considering trace-projection, as shown in the following proposition theorem.

Proposition 5.5.1 (Non-contractivity of Bures projections)**.** *There exists a convex and closed set $\mathcal{C} \subset \mathbb{P}_d$ such that $\Pi_{\mathcal{C}}$ is non-contractive:*

$$\mathrm{B}(P,Q) < \mathrm{B}(\Pi_{\mathcal{C}}[P],\Pi_{\mathcal{C}}[Q]), \tag{5.92}$$

where the projection Π is with respect to Bures distance.

Proof. The proof is by explicit example. Let $\mathcal{C} = \mathbb{D}_d$ be the convex and closed set of $d \times d$ density matrices. The projection, as we have seen, is given by trace-normalization: $\Pi_{\mathbb{D}_d}[P] = P/\mathrm{Tr}[P]$. Choose $P = \epsilon\rho$ and $Q = \epsilon\sigma$ where $\epsilon > 0$ can be thought of as a small positive scalar and $\rho, \sigma \in \mathbb{D}_d$. Indeed Bures projecting to the set of density matrices gives $\Pi_{\mathbb{D}_d}[\epsilon\rho] = \rho$ and $\Pi_{\mathbb{D}_d}[\epsilon\sigma] = \sigma$, whence it follows

$$\mathrm{B}(P,Q) = \mathrm{B}(\epsilon\rho,\epsilon\sigma) = \epsilon\mathrm{B}(\rho,\sigma) = \epsilon\mathrm{B}(\Pi_{\mathbb{D}_d}[P],\Pi_{\mathbb{D}_d}[Q]). \tag{5.93}$$

On choosing any $0 < \epsilon < 1$, we have the required example, thereby showing that Bures projection onto convex sets need not always be contractive. □

This contrasts with the Euclidean distance, where projections onto convex sets are always contractive, aiding in the proofs of convergence results for projected optimization algorithms. Consequently, additional effort might be required when proving convergence results involving Bures projection. Also see [Alt+21, Appendix B.3] for a brief discussion on how the geometry of the Bures manifold might play a role in the non-contractivity of projections (although there they are

concerned with projections onto *geodesic convex*[5] sets).

We conclude with a visualization of the non-contractivity of Bures projection. View the set of positive semidefinite matrices as a vertically oriented cone, in which case trace-normalization corresponds to scaling a point along a line connecting that point to the apex (the pointy end). The set of density matrices can be then visualized as a horizontal slice at some height. If the pre-projection points are 'between' this slice and the apex, then the scaling (Bures projection onto density matrices) will take them further apart, thereby showing that it is non-contractive.

5.6 Conclusion

In this chapter, we studied the problem of projecting positive matrices onto certain convex and compact subsets of positive semidefinite matrices with respect to Bures distance. These sets are defined as the preimage of a given *output state* under a given channel. Using the algebraic condition for the saturation of data processing inequality for fidelity, we construct a function we call the *Gamma map* which, for certain channels, serves as a closed-form for the projection and derive sufficient conditions for the Gamma map to yield the projection.

We then give explicit instances for which the Gamma map serves as the projection closed-form. This includes partial trace (including trace), pinching channels (including completely dephasing map), and projective measurements. We then study the explicit form of the projection for these channels.

The projection with respect to the partial trace channel is of particular importance. Essentially this gives the Bures projection to the set of states with a given marginal. We then show the closed-form for this *marginal projection* has multiple applications such as

1. Closed-form for projection to set of Choi matrices of channels with applications in quantum process tomography.

2. Closed-form for projecting a given ensemble to the set of ensembles that sum to a fixed state.

3. A method for generating random matrices and ensembles with specific properties.

The same closed-form also gives the following geometric interpretations to existing results in quantum information.

[5] A subset of a Riemannian manifold is geodesic convex if for any two points the geodesic connecting them lies entirely within the set. In the Euclidean scenario, geodesic convexity coincides with (regular) convexity.

1. Geometric and operational interpretations for pretty good measurement.

2. A geometric interpretation for the Petz recovery map.

3. A geometric interpretation and simpler proof for the *Quantum minimal change principle*.

Note that all closed-form projections also saturate the DPI for fidelity (and thus also sandwiched Rényi divergence of order $\alpha = 1/2$) thereby our results also give explicit examples for the saturation of DPI for these figures of merit.

Our results open up new questions of interest, some of which we list below.

1. Are there other channels for which Γ map yields the closed-form for Bures projection?

2. Are there simpler ways to characterize the conditions when the Γ map will coincide with the projection?

3. The marginal projection has proven to be versatile with various applications and manifestations. Are there other applications for marginal projection?

4. All of our applications pertain to the partial trace channel. Are there concrete applications for the other channels such as pinching maps and projective measurement? For example, does the problem of finding the closest state with a given diagonal to a given state have applications in quantum information?

An important caveat we then discuss is that Bures projections onto convex sets need not be contractive, as we have demonstrated with an example. That is, it does not always hold that $\mathrm{B}(P, Q) \geq \mathrm{B}(\Pi[P], \Pi[Q])$. Contractivity upon projection to convex sets is a crucial property enjoyed by Euclidean projection (see [htta] and [Bub15, Lemma 3.1]) which eases the analysis of projected optimization algorithms like projected gradient descent. Due to non-contractivity, the analysis of optimization algorithms based on Bures projection can be more difficult. However, convergence analysis for such algorithms is an active field of research [Alt+21; Che+20].

Chapter 6

Relating average fidelity and generalized fidelity

In this short chapter, we elucidate some relations between the results discussed in the different chapters of this thesis. We first show that the SDP used to construct the block-matrix characterization of generalized fidelity is a special case of the SDP from average fidelity maximization (Eq. (3.3)).

6.1 Block-matrix characterization and average fidelity SDP

In Chapter 4, we discussed the block-matrix characterization of generalized fidelity. We now show this characterization is closely related to the semidefinite program for optimal average fidelity.

Recall that the block-matrix characterization of generalized fidelity $\mathrm{F}_R(P,Q)$ is based on the following block matrix from Theorem 4.4.1.

$$\begin{aligned}\mathrm{X}_\star &:= \begin{pmatrix} P & P^{\frac{1}{2}}U_P U_Q^* Q^{\frac{1}{2}} & P^{\frac{1}{2}}U_P R^{\frac{1}{2}} \\ Q^{\frac{1}{2}}U_Q U_P^* P^{\frac{1}{2}} & Q & Q^{\frac{1}{2}}U_Q R^{\frac{1}{2}} \\ R^{\frac{1}{2}}U_P^* P^{\frac{1}{2}} & R^{\frac{1}{2}}U_Q^* Q^{\frac{1}{2}} & R \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{G}_R(P,P) & \mathrm{G}_R(Q,P) & \mathrm{G}_R(R,P) \\ \mathrm{G}_R(P,Q) & \mathrm{G}_R(Q,Q) & \mathrm{G}_R(R,Q) \\ \mathrm{G}_R(P,R) & \mathrm{G}_R(Q,R) & \mathrm{G}_R(R,R) \end{pmatrix}\end{aligned} \tag{6.1}$$

where we have denoted $\mathrm{G}_R(P,Q) \equiv Q^{\frac{1}{2}}U_Q U_P^* P^{\frac{1}{2}} = \sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}$ which is the matrix whose trace gives the generalized fidelity: $\mathrm{Tr}[\mathrm{G}_R(P,Q)] = \mathrm{F}_R(P,Q)$.

Now let us recall the SDP for average fidelity from Definition 3.3.1, whose

form, for a fixed state σ, is

$$\mathrm{X}_\star := \begin{pmatrix} \rho_1 & \cdots & \rho_1^{\frac{1}{2}} U_1 U_n^* \rho_n^{\frac{1}{2}} & \rho_1^{\frac{1}{2}} U_1 \sigma^{\frac{1}{2}} \\ \vdots & \ddots & \vdots & \vdots \\ \rho_n U_n U_1^* \rho_1 & \cdots & \rho_n & \rho_n^{\frac{1}{2}} U_n \sigma^{\frac{1}{2}} \\ \sigma^{\frac{1}{2}} U_1^* \rho_1^{\frac{1}{2}} & \cdots & \sigma^{\frac{1}{2}} U_n^* \rho_n^{\frac{1}{2}} & R \end{pmatrix}, \tag{6.2}$$

where $U_i = \mathrm{Pol}\left(\rho_i^{\frac{1}{2}} \sigma^{\frac{1}{2}}\right)$ for each $i \in [n]$. For a distribution supported on two states P and Q, and an arbitrary input state R this takes the form

$$\mathrm{X}_\star := \begin{pmatrix} P & P^{\frac{1}{2}} U_P U_Q^* Q^{\frac{1}{2}} & P^{\frac{1}{2}} U_P R^{\frac{1}{2}} \\ Q^{\frac{1}{2}} U_Q U_P^* P^{\frac{1}{2}} & Q & Q^{\frac{1}{2}} U_Q R^{\frac{1}{2}} \\ R^{\frac{1}{2}} U_P^* P^{\frac{1}{2}} & R^{\frac{1}{2}} U_Q^* Q^{\frac{1}{2}} & R \end{pmatrix}, \tag{6.3}$$

with $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}} R^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}} R^{\frac{1}{2}}\right)$. Thus we see that the two block matrices are the same. Essentially, for a distribution supported on two states, the input to the barycenter function is equivalent to the base in the generalized fidelity problem. In the barycenter problem, we are interested in terms of the form $P^{\frac{1}{2}} U_P R^{\frac{1}{2}}$, as these are the terms whose trace yields fidelity. Whereas in the generalized fidelity problem, we are interested in terms of the form $P^{\frac{1}{2}} U_P U_Q^* Q^{\frac{1}{2}}$, as the trace of these terms yields generalized fidelity.

For the barycenter problem where the distribution is supported on n points we have, for any state σ,

$$\begin{aligned} \mathrm{X}(\sigma) &= \begin{pmatrix} \mathrm{G}_\sigma(\rho_1, \rho_1) & \cdots & \mathrm{G}_\sigma(\rho_1, \rho_n) & \mathrm{G}_\sigma(\rho_1, \sigma) \\ \vdots & \ddots & \vdots & \vdots \\ \mathrm{G}_\sigma(\rho_n, \rho_1) & \cdots & \mathrm{G}_\sigma(\rho_n, \rho_n) & \mathrm{G}_\sigma(\rho_n, \sigma) \\ \mathrm{G}_\sigma(\sigma, \rho_1) & \cdots & \mathrm{G}_\sigma(\sigma, \rho_n) & \mathrm{G}_\sigma(\sigma, \sigma) \end{pmatrix} \\ &= \sum_{i,j=1}^{n} |i\rangle\langle j| \otimes \mathrm{G}_\sigma(\rho_i, \rho_j) + \sum_{i=1}^{n} |i\rangle\langle n+1| \otimes \mathrm{G}_\sigma(\rho_i, \sigma) + |n+1\rangle\langle i| \otimes \mathrm{G}_\sigma(\sigma, \rho_i). \end{aligned} \tag{6.4}$$

We thus see that the SDP formulation for optimal fidelity has fundamental concepts from generalized fidelity and the Riemannian geometry of the Bures manifold hidden away in plain sight.

6.2 Generalized fidelity, multivariate fidelities, and fidelity barycenters

We now relate fidelity barycenters (Chapter 3) and generalized fidelity (Chapter 4) to recent results by Nuradha, Mishra, Leditzky, and Wilde [Nur+24] which introduces and studies notions of *multivariate fidelities*.

Let $\mathcal{R} := \{\rho_1, \dots, \rho_n\} \subset \mathbb{D}_d$ be a collection of quantum states. In [Nur+24], the authors introduced and studied various notions of multivariate fidelities—fidelities between a collection of $n \geq 2$ states. One particular version of multivariate fidelity they define and study is the normalized sum of distinct pairwise (Uhlmann) fidelities:

$$\mathrm{F}(\rho_1, \dots, \rho_n) := \frac{1}{n(n-1)} \sum_{i,j=1; i \neq j}^{n} \mathrm{F}^{\mathrm{U}}(\rho_i, \rho_j)\,. \tag{6.5}$$

This quantity has previously appeared in [NR18] as well.

Given a base $\sigma \in \mathbb{D}_d$, one can then define the analogous generalized multivariate fidelity at as

$$\mathrm{F}_\sigma(\rho_1, \dots, \rho_n) := \frac{1}{n(n-1)} \sum_{i,j=1; i \neq j}^{n} \mathrm{F}_\sigma(\rho_i, \rho_j)\,. \tag{6.6}$$

Define the *total fidelity* function as $g(\sigma) := \sum_{i=1}^n \mathrm{F}^{\mathrm{U}}(\rho_i, \sigma)$. Observe that the maximizer of this function over $\mathbb{D}_d$ is identical to the maximizer of the average fidelity function $f(\sigma) := \frac{1}{n}\sum_{i=1}^n \mathrm{F}^{\mathrm{U}}(\rho_i, \sigma)$ as we have uniform probability weights.

Suppose all the states in the ensemble $\mathcal{R}$ are full-rank and $\sigma_\sharp$ is the average fidelity maximizer of the (uniform) distribution over these points. That is, $\sigma_\sharp = \operatorname{argmax}_{\sigma \in \mathbb{D}_d} f(\sigma)$. By Theorem 3.3.5 we have that $\sigma_\sharp$ uniquely satisfies the fixed-point equation:

$$\sigma_\sharp = \frac{1}{f(\sigma_\sharp)} \sum_{i=1}^n \frac{1}{n} \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}} = \frac{1}{g(\sigma_\sharp)} \sum_{i=1}^n \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}}, \tag{6.7}$$

where we have used the fact that $g(\sigma) = n f(\sigma)$ for any $\sigma \in \mathbb{D}_d$. One can relate the total fidelity of this maximizer and the generalized multivariate fidelity as shown in the following theorem.

Theorem 6.2.1. *Let $\{\rho_1, \ldots, \rho_n\} \subset \mathbb{D}_d$ be a collection of full-rank states and $\sigma_\sharp := \operatorname{argmax}_{\sigma \in \mathbb{D}_d} g(\sigma)$ be the barycenter of the collection. Then*

$$g(\sigma_\sharp)^2 = \sum_{i,j=1}^{n} \mathrm{F}_{\sigma_\sharp}(\rho_i, \rho_j)\,. \tag{6.8}$$

Moreover, the generalized multivariate fidelity is related to the average fidelity of the maximizer as follows:

$$\mathrm{F}_{\sigma_\sharp}(\rho_1, \ldots, \rho_n) = \frac{g(\sigma_\sharp)^2 - n}{n^2 - n}. \tag{6.9}$$

Proof. Observe that

$$\begin{aligned} \sigma_\sharp = \sigma_\sharp \sigma_\sharp^{-1} \sigma_\sharp &= \left(\frac{1}{g(\sigma_\sharp)} \sum_{i=1}^{n} \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}} \right) \sigma_\sharp^{-1} \left(\frac{1}{g(\sigma_\sharp)} \sum_{j=1}^{n} \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_j \sigma_\sharp^{\frac{1}{2}}} \right) \\ &= \frac{1}{g(\sigma_\sharp)^2} \sum_{i,j=1}^{n} \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}} \sigma_\sharp^{-1} \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_j \sigma_\sharp^{\frac{1}{2}}}, \end{aligned} \tag{6.10}$$

where the second equality is by Equation (6.7). Take trace across and rearrange to obtain

$$g(\sigma_\sharp)^2 = \sum_{i,j=1}^{n} \operatorname{Tr}\left[\sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_i \sigma_\sharp^{\frac{1}{2}}} \sigma_\sharp^{-1} \sqrt{\sigma_\sharp^{\frac{1}{2}} \rho_j \sigma_\sharp^{\frac{1}{2}}} \right] = \sum_{i,j=1}^{n} \mathrm{F}_{\sigma_\sharp}(\rho_i, \rho_j)\,, \tag{6.11}$$

which proves the first claim. Note that $n = \sum_{i=1}^{n} \mathrm{F}_{\sigma_\sharp}(\rho_i, \rho_i)$ and subtract this quantity from both sides (respectively) to obtain

$$g(\sigma_\sharp)^2 - n = \sum_{i,j=1; i \neq j}^{n} \mathrm{F}_{\sigma_\sharp}(\rho_i, \rho_j)\,. \tag{6.12}$$

Divide throughout by $n(n-1)$ to obtain

$$\frac{g(\sigma_\sharp)^2 - n}{n^2 - n} = \frac{1}{n(n-1)} \sum_{i,j=1; i \neq j}^{n} \mathrm{F}_{\sigma_\sharp}(\rho_i, \rho_j) =: \mathrm{F}_{\sigma_\sharp}(\rho_1, \ldots, \rho_n), \tag{6.13}$$

which concludes the proof. □

Chapter 7

Conclusion

In this thesis, we studied three problems related to quantum fidelities. Succinctly these problems can be described as follows.

1. In Chapter 3, we studied *maximizing average Uhlmann fidelity.*
2. In Chapter 4 we *generalize quantum fidelities.*
3. In Chapter 5 we *project with respect to Uhlmann fidelity.*

To elaborate, in **In Chapter 3** we presented results related to the state that maximizes average (Uhlmann) fidelity over a finite distribution of quantum states. This is equivalent to finding the quantum state that minimizes squared Bures distance over the ensemble. We formulated the problem as a semidefinite program and also found two fixed-point algorithms for finding the state that maximizes average fidelity (fidelity barycenter). The fixed-point algorithms offer superior numerical performance compared to the SDP. We showed that one of the fixed-point algorithms can be seen as a projected Riemannian gradient descent on the Bures manifold.

We also presented easier-to-compute upper and lower bounds for the optimal average fidelity. Apart from being tight, these bounds coincide with optimal average fidelity if the distribution is supported on a set of commuting states. We then studied the applications of these results, especially in relation to Bayesian quantum tomography. The theoretical results are complemented with numerical experiments studying the performance of these fixed-point algorithms and the tightness of bounds.

Next, in **Chapter 4**, we discuss the second set of results which tackle the more ambitious goal of finding a *generalized (quantum) fidelity* which can recover various existing notions of quantum fidelities. We achieve this by studying the Riemannian geometry of the Bures–Wasserstein manifold. Essentially, one can *linearize* the manifold at an arbitrary *base* point by mapping other points to the

tangent space. Then one can compute the natural distance (the distance induced by the metric tensor on the tangent space) between these tangent vectors whose functional form turns out to be similar to that of Bures distance. This distance is called the generalized Bures distance, and the *fidelity part* is identified as the generalized fidelity.

The rest of the chapter is devoted to studying various mathematical properties and characterizations of generalized fidelity and generalized Bures distance. Apart from basic properties, we study some geometric results related to generalized fidelity, including various invariance and covariance properties as the base moves along certain geodesic-related paths for fixed input matrices. Other properties we study include block-matrix characterization of generalized fidelity and generalized Bures distance and characterization of generalized fidelity in terms of purifications (an *Uhlmann-like theorem* for generalized fidelity). We also define and study a family of fidelities parametrized by a single real-number parameter called *polar fidelities* that generalizes Uhlmann-, Holevo-, and Matsumoto-fidelity. Generalized fidelities are further *generalized* by *Interior fidelities*. Specifically, an interior fidelity between two points is defined by taking the convex combination of generalized fidelity between the two points at different base points. This imparts generalized fidelities with the interpretation of being the *extreme points* of the family of interior fidelities. We propose metric learning as a possible application to generalized fidelity. We end the chapter with various open problems such as convexity, data processing inequality, and certain monotonicity properties.

The final set of results is discussed in **Chapter 5** where we study the problem of *Bures projection*—projecting an arbitrary positive matrix to certain convex and compact sets with respect to the Bures distance. A simple example is projecting a positive semidefinite matrix to the convex set of positive semidefinite matrices with a fixed trace. In this simple case, the projection in this case is given by appropriate scaling. We also present closed-form for the projections with respect to partial trace, pinching channels, and projective measurements.

The projection with respect to the partial trace map, which we call the *marginal projection*, is of particular importance. Using it, we study a wide range of applications—including quantum process tomography, random state and ensemble generation, and projection for generic optimization methods—as well as its manifestations, such as novel geometric and operational interpretations of the pretty good measurement, a new geometric interpretation of the Petz recovery map, and a shorter proof of the central results of the recent article on *quantum minimal change principle* [BBS24]. Our results also provide explicit examples for the saturation of DPI for fidelity and sandwiched Rényi divergence of order $\alpha = 1/2$.

As an epilogue, **Chapter 6** discusses the relations between results from the previous chapters. We show that the SDP for average fidelity and the block-matrix characterization of generalized fidelity are closely related. We also show certain relations between generalized fidelity, multivariate fidelity, and the Uhlmann fidelity barycenter.

In conclusion, we study various problems related to quantum fidelities. Our results hold both practical and theoretical value. Practically, it turns out to be useful in the tomography of quantum states and channels and the generation of various random states and ensembles of interest in quantum information. Theoretically, they uncover fundamental geometric properties of quantum states, channels, and ensembles. They also lead to a panoply of related open problems with geometric and practical implications.

For future work, we plan to study some of the open problems we discussed. Specifically, we aim to use the closed-form for Choi projection to construct channel tomography protocols. The first one aims to follow the lines of *Projected Least Squares channel tomography* [Sur+22]. In the second algorithm, we aim to do Bayesian tomography of quantum channels. This would be a combination (and extension) of results from Chapter 3 and Chapter 5 by replacing the trace projection in projected RGD with a marginal projection to the Choi matrices of quantum channels. We plan to study the convergence analysis more rigorously as well. We also plan to extend the results of Chapter 5 to more channels and study other manifestations of the partial trace projection in quantum information.

Appendix A

Additional preliminaries

In this section, we list some additional mathematical preliminaries and supporting results.

A.1 Deriving the Bures metric tensor

We now derive the form of the Bures metric tensor $\langle U, V\rangle_P^{\mathrm{Bu}}$ starting from the square of the *line element* $(\mathrm{d}s)^2$ at $P \in \mathbb{P}_d^+$ [BŻ17, Eq. 9.43]:

$$(\mathrm{d}s)^2 \equiv (\mathrm{d}s|_P)^2 := \frac{1}{2}\langle \mathrm{d}P, \mathcal{L}_P(\mathrm{d}P)\rangle_P \equiv \frac{1}{2}\langle V, \mathcal{L}_P(V)\rangle_P. \tag{A.1}$$

Here $\mathrm{d}P$ is to be interpreted as a tangent vector, and thus we denote it by $V \in \mathrm{T}_P\mathbb{P}_d^+$. Recall that $\mathcal{L}_P(V)$ is the unique Hermitian solution to the (implicit) matrix Lyapunov equation $V = \mathcal{L}_P(V)P + P\mathcal{L}_P(V)$ [Syl84; BR97]. Since V is a tangent vector, its norm can be computed as $\|V\|_P^2 = \mathrm{B}(P, P+V) = \frac{1}{2}\langle P, \mathcal{L}_P(V)\rangle$. By parallelogram equality (which holds for any norm induced by an inner product), we may write

$$\|U+V\|_P^2 = \|U\|_P^2 + \|V\|_P^2 + \langle U, V\rangle_P + \langle V, U\rangle_P = \|U\|_P^2 + \|V\|_P^2 + 2\langle U, V\rangle_P, \tag{A.2}$$

where the last equality follows from the fact that $\mathrm{T}_P\mathbb{P}_d^+ \cong \mathbb{H}_d$ is a real vector space, and thus the inner product must be real (and thus symmetric). The LHS is given by

$$\begin{aligned}\|U+V\|_P^2 &= \frac{1}{2}\langle U+V, \mathcal{L}_P(U+V)\rangle \\ &= \frac{1}{2}\langle U, \mathcal{L}_P(U)\rangle + \frac{1}{2}\langle V, \mathcal{L}_P(V)\rangle + \frac{1}{2}\langle U, \mathcal{L}_P(V)\rangle + \frac{1}{2}\langle V, \mathcal{L}_P(U)\rangle \\ &= \|U\|_P^2 + \|V\|_P^2 + \frac{1}{2}\langle U, \mathcal{L}_P(V)\rangle + \frac{1}{2}\langle V, \mathcal{L}_P(U)\rangle\end{aligned} \tag{A.3}$$

By comparison, we have

$$\frac{1}{2}\operatorname{Tr}[U\mathcal{L}_P(V) + V\mathcal{L}_P(U)] = \langle U, V\rangle_P + \langle V, U\rangle_P = 2\langle U, V\rangle_P. \tag{A.4}$$

Recall that $U = \mathcal{L}_P(U)P + P\mathcal{L}_P(U)$, and we thus have

$$\begin{aligned}\operatorname{Tr}[U\mathcal{L}_P(V)] &= \operatorname{Tr}[(\mathcal{L}_P(U)P + P\mathcal{L}_P(U))\mathcal{L}_P(V)] \\ &= \operatorname{Tr}[\mathcal{L}_P(U)P\mathcal{L}_P(V)] + \operatorname{Tr}[\mathcal{L}_P(V)P\mathcal{L}_P(U)] \\ &= 2\operatorname{Re}\operatorname{Tr}[\mathcal{L}_P(U)P\mathcal{L}_P(V)]\end{aligned} \tag{A.5}$$

By symmetry we have $\operatorname{Tr}[V\mathcal{L}_P(U)] = 2\operatorname{Re}\operatorname{Tr}[\mathcal{L}_P(U)P\mathcal{L}_P(V)]$, which allows us to conclude

$$\langle U, V\rangle_P = \operatorname{Re}\operatorname{Tr}[\mathcal{L}_P(U)P\mathcal{L}_P(V)] = \frac{1}{2}\operatorname{Tr}[U\mathcal{L}_P(V)] = \frac{1}{2}\operatorname{Tr}[V\mathcal{L}_P(U)]. \tag{A.6}$$

as required.

A.2 A Lemma on the unitary factor of polar decomposition

The first lemma characterizes the unitary factor of $P^{\frac{1}{2}}Q^{\frac{1}{2}}$ for $P, Q \in \mathbb{P}_d^+$.

Lemma A.2.1. *Let $P, Q \in \mathbb{P}_d^+$ and let $U := \operatorname{Pol}\left(P^{\frac{1}{2}}Q^{\frac{1}{2}}\right)$. Then,*

$$U^*P^{\frac{1}{2}}Q^{\frac{1}{2}} = \sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}} = Q^{\frac{1}{2}}P^{\frac{1}{2}}U, \tag{A.7}$$

$$U\sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}}U^* = \sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}}, \tag{A.8}$$

$$U^* = \operatorname{Pol}\left(Q^{\frac{1}{2}}P^{\frac{1}{2}}\right), \tag{A.9}$$

$$U = \operatorname{Pol}\left(P^{-\frac{1}{2}}Q^{-\frac{1}{2}}\right), \tag{A.10}$$

$$P \cdot (P^{-1}\#Q) = \sqrt{PQ} = P^{\frac{1}{2}}UQ^{\frac{1}{2}} \tag{A.11}$$

$$(P^{-1}\#Q) \cdot P = \sqrt{QP} = Q^{\frac{1}{2}}U^*P^{\frac{1}{2}}. \tag{A.12}$$

Proof. The first statement is a direct consequence of the definition of polar decomposition. To prove the second statement, we begin with the first statement. Multiply the first and last sides together and square the middle side to get

$$U^*P^{\frac{1}{2}}QP^{\frac{1}{2}}U = Q^{\frac{1}{2}}PQ^{\frac{1}{2}}. \tag{A.13}$$

Conjugate both sides with U and take the square root to obtain the required

relation.

For the third statement, we begin with the polar decomposition of $Q^{\frac{1}{2}}P^{\frac{1}{2}}$ to get

$$Q^{\frac{1}{2}}P^{\frac{1}{2}} = V\sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}} = \sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}}U^*, \tag{A.14}$$

where we denote $V = \text{Pol}\left(Q^{\frac{1}{2}}P^{\frac{1}{2}}\right)$ and the last equality comes from Eq. (A.7). Multiply by V^* on the left to get

$$\sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}} = V^*\sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}}U^*. \tag{A.15}$$

Comparing the above equation with Equation (A.8), we get $V = U^*$ as claimed.

To prove the fourth statement, begin with the polar decomposition of $P^{\frac{1}{2}}Q^{\frac{1}{2}}$ and take the inverse across to get

$$Q^{-\frac{1}{2}}P^{-\frac{1}{2}} = \sqrt{Q^{-\frac{1}{2}}P^{-1}Q^{-\frac{1}{2}}}U^*. \tag{A.16}$$

Now take the adjoint to get the desired result.

For the fifth statement, we will use the fact that for any matrix A, not necessarily Hermitian, with positive eigenvalues, there exists a unique matrix B such that $B^2 = BB = A$. Thus, we may denote $B \equiv \sqrt{A}$. See [Bha09, Excercise 4.5.2] for further details. To prove the first part, observe that

$$\begin{aligned}(P \cdot P^{-1}\#Q)\left(P \cdot P^{-1}\#Q\right) &= \left(P \cdot P^{-\frac{1}{2}}\sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}}P^{-\frac{1}{2}}\right)\left(P \cdot P^{-\frac{1}{2}}\sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}}P^{-\frac{1}{2}}\right) \\ &= PQ.\end{aligned} \tag{A.17}$$

To prove the second equality, it is sufficient to prove the squared version:

$$PQ \stackrel{?}{=} P^{\frac{1}{2}}UQ^{\frac{1}{2}}P^{\frac{1}{2}}UQ^{\frac{1}{2}}, \tag{A.18}$$

which can be proven by noting that $Q^{\frac{1}{2}}P^{\frac{1}{2}}U = U^*P^{\frac{1}{2}}Q^{\frac{1}{2}}$ by Statement 1. Substituting, the RHS reduces to PQ, thus completing the proof. The final equality is proven in the same manner as the previous equality, and hence, we skip the proof. □

Appendix B

Further results related to averaging fidelities

B.1 Alternate semidefinite program for optimal fidelity

A more numerically tractable SDP (which is still not as tractable as the fixed-point algorithms) to solve the maximization problem Eq. (3.1) can be constructed as follows. Let $\{\rho_i\}_{i=1}^n \subset \mathbb{D}_d$ be a collection of states. The alternate SDP for optimal fidelity is formulated as n different fidelity SDPs (Watrous [Wat18, Theorem 3.17]), with the constraint that the matrix variable σ is the same in each of the n SDPs. That is, given the weighted ensemble $(\mathcal{R}, p)$ with $\mathcal{R} = \{\rho_1, \ldots, \rho_n\}$,

$$\begin{aligned} &\text{Primal problem} \\ \text{maximize}: &\sum_{i=1}^{n} p_i \operatorname{Re}\operatorname{Tr}[X_i] \\ \text{subject to}: &\begin{pmatrix} \rho_i & X_i \\ X_i^* & \sigma \end{pmatrix} \geq 0, \quad \text{for each } i \in [n], \\ &\operatorname{Tr}[\sigma] = 1. \end{aligned} \tag{B.1}$$

This SDP also achieves the primal optimum $\max_{\sigma \in \mathbb{D}_d} f(\sigma)$ while being more numerically tractable. We compare the performance of this SDP along with the original SDP and the FP algorithm in Fig. 3.2. The relation between the original SDP and the alternate SDP can be seen using concepts from chordal graph-structured SDPs [Fuk+01; Zhe+20; VA+15].

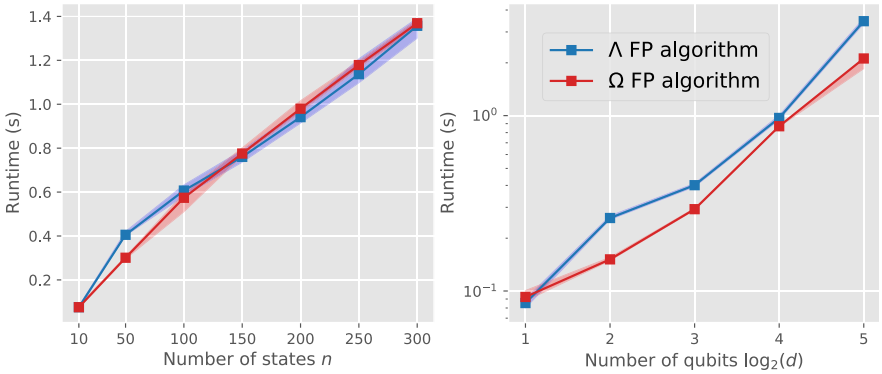

Figure B.1: Runtime comparison of Λ FP algorithm Eq. (3.33) and Ω FP algorithm Eq. (3.37) as a function of (a) Number of states for $d = 8$ and (b) Number of qubits ($\log_2(d)$) for $n = 50$. Both FP algorithms perform similarly in terms of runtime. Each data point is the median of 50 runs and interquartile regions are shaded.

B.2 Runtime comparison of fixed point algorithms

When all the states in the ensemble are full rank, both the FP algorithms are numerically seen to converge to the optimal state for any full-rank starting point. We now turn to numerics to compare the performance (runtime) of the two fixed-point methods. As seen in Fig. B.1, the runtime performance of Ω fixed point algorithm Eq. (3.37) and Λ fixed point algorithm Eq. (3.33) are comparable. Here we choose stopping tolerance $\epsilon = 10^{-5}$. Since the convergence of the Ω fixed-point algorithm is theoretically guaranteed, it should be preferred over Λ fixed-point algorithm even though the former has a more complicated form than the latter.

Appendix C

Proofs related Generalized fidelity

Here we present the proofs related to generalized fidelity.

C.1 Proofs of basic properties

In this section, we prove the basic properties of generalized fidelity stated in Section 4.1. For the rest of the section we arbitrarily choose and fix $P, Q, R \in \mathbb{P}_d^+$ and define $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}}R^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}}R^{\frac{1}{2}}\right)$.

1. **Conjugate Symmetry**. $\mathrm{F}_R(P,Q) \stackrel{?}{=} \mathrm{F}_R(Q,P)^*$. The claim easily follows from the definition:

$$\begin{aligned} \mathrm{F}_R(P,Q) :&= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right] \\ &= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\right]^* =: \mathrm{F}_R(Q,P)^*. \end{aligned} \tag{C.1}$$

2. **Equivalent forms**:

$$\begin{aligned} \mathrm{F}_R(P,Q) &:= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right] \\ &\stackrel{?}{=} \mathrm{Tr}\left[Q^{\frac{1}{2}}U_QU_P^*P^{\frac{1}{2}}\right] \\ &\stackrel{?}{=} \mathrm{Tr}\left[(R^{-1}\#Q)R(R^{-1}\#P)\right], \end{aligned} \tag{C.2}$$

   To show the first equality, we simply take the polar decompositions $\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}} = U_P^*P^{\frac{1}{2}}R^{\frac{1}{2}}$ and $\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}} = R^{\frac{1}{2}}Q^{\frac{1}{2}}U_Q$, substitute in the definition of generalized fidelity, and then use the cyclic nature of the trace map. To prove the

third equivalent form

$$\mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right] \stackrel{?}{=} \mathrm{Tr}\left[(R^{-1}\#Q)R(R^{-1}\#P)\right], \tag{C.3}$$

substitute the formula for geometric mean. The relation immediately follows.

3. **Simplified form for pure states.** Assume $P = |\psi\rangle\langle\psi|$ and $Q = |\phi\rangle\langle\phi|$. Consider the term

$$\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}} = \sqrt{|u\rangle\langle u|} = \frac{|u\rangle\langle u|}{\sqrt{\langle u,u\rangle}}, \tag{C.4}$$

where we have defined $|u\rangle := R^{\frac{1}{2}}|\psi\rangle$. A similar calculation reveals

$$\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}} = \frac{|v\rangle\langle v|}{\sqrt{\langle v,v\rangle}}, \tag{C.5}$$

for $|v\rangle := R^{\frac{1}{2}}|\phi\rangle$. Thus we have

$$\begin{aligned}
\mathrm{F}_R(P,Q) &= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right] \\
&= \frac{1}{\|u\|\|v\|}\,\mathrm{Tr}\left[|u\rangle\langle u|\,R^{-1}\,|v\rangle\langle v|\right] \\
&= \frac{1}{\|u\|\|v\|}\,\mathrm{Tr}\left[R^{\frac{1}{2}}\,|\psi\rangle\langle\psi|\,R^{\frac{1}{2}}R^{-1}R^{\frac{1}{2}}\,|\phi\rangle\langle\phi|\,R^{\frac{1}{2}}\right] \\
&= \frac{\langle\psi,\phi\rangle\langle\phi,R\psi\rangle}{\mathrm{F}^{\mathrm{U}}(P,R)\,\mathrm{F}^{\mathrm{U}}(Q,R)},
\end{aligned} \tag{C.6}$$

where in the last equality, we used $\|u\| = \sqrt{\langle\psi,R\psi\rangle} = \mathrm{F}^{\mathrm{U}}(P,R)$ and the analogous identity for $\|v\|$.

4. **Commutation with base implies reality.** Without loss of generality, assume $[R,P] = 0$. Then we have $\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}} = P^{\frac{1}{2}}R^{\frac{1}{2}}$, which implies

$$\mathrm{F}_R(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}}R^{-\frac{1}{2}}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right], \tag{C.7}$$

which is the inner product of two positive matrices, and thus the generalized fidelity is real (and positive).

5. **Multiplicativity.** For $P_1,Q_1,R_1 \in \mathbb{P}_{d_1}$ and $P_2,Q_2,R_2 \in \mathbb{P}_{d_2}$, the claim is

$$\mathrm{F}_{R_1\otimes R_2}(P_1\otimes P_2, Q_1\otimes Q_2) \stackrel{?}{=} \mathrm{F}_{R_1}(P_1,Q_1)\cdot\mathrm{F}_{R_2}(P_2,Q_2)\,. \tag{C.8}$$

Denote $P \equiv P_1 \otimes P_2, Q \equiv Q_1 \otimes Q_2$, and $R \equiv R_1 \otimes R_2$. Then,

$$\mathrm{F}_{R_1 \otimes R_2}(P_1 \otimes P_2, Q_1 \otimes Q_2) = \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}} P R^{\frac{1}{2}}} R^{-1} \sqrt{R^{\frac{1}{2}} Q R^{\frac{1}{2}}}\right]. \tag{C.9}$$

Since the matrix square root, product, and inverse factors out with respect to the tensor product, we have

$$\begin{aligned}&\sqrt{R^{\frac{1}{2}} P R^{\frac{1}{2}}} R^{-1} \sqrt{R^{\frac{1}{2}} Q R^{\frac{1}{2}}} \\ &\qquad = \left[\sqrt{R_1^{\frac{1}{2}} P_1 R_1^{\frac{1}{2}}} R_1^{-1} \sqrt{R_1^{\frac{1}{2}} Q_1 R_1^{\frac{1}{2}}}\right] \otimes \left[\sqrt{R_2^{\frac{1}{2}} P_2 R_2^{\frac{1}{2}}} R_2^{-1} \sqrt{R_2^{\frac{1}{2}} Q_2 R_2^{\frac{1}{2}}}\right].\end{aligned} \tag{C.10}$$

Take trace across and using the identity $\mathrm{Tr}[A \otimes B] = \mathrm{Tr}[A]\,\mathrm{Tr}[B]$ to obtain $\mathrm{F}_R(P, Q) = \mathrm{F}_{R_1}(P_1, Q_1) \cdot \mathrm{F}_{R_2}(P_2, Q_2)$.

6. **Additivity.** Here the claim is

$$\mathrm{F}_{R_1 \oplus R_2}(P_1 \oplus P_2, Q_1 \oplus Q_2) \stackrel{?}{=} \mathrm{F}_{R_1}(P_1, Q_1) + \mathrm{F}_{R_2}(P_2, Q_2). \tag{C.11}$$

Denote $P \equiv P_1 \oplus P_2, Q \equiv Q_1 \oplus Q_2$, and $R \equiv R_1 \oplus R_2$. From the properties of direct sum we have,

$$\begin{aligned}&\sqrt{R^{\frac{1}{2}} P R^{\frac{1}{2}}} R^{-1} \sqrt{R^{\frac{1}{2}} Q R^{\frac{1}{2}}} \\ &\qquad = \left[\sqrt{R_1^{\frac{1}{2}} P_1 R_1^{\frac{1}{2}}} R_1^{-1} \sqrt{R_1^{\frac{1}{2}} Q_1 R_1^{\frac{1}{2}}}\right] \oplus \left[\sqrt{R_2^{\frac{1}{2}} P_2 R_2^{\frac{1}{2}}} R_2^{-1} \sqrt{R_2^{\frac{1}{2}} Q_2 R_2^{\frac{1}{2}}}\right].\end{aligned} \tag{C.12}$$

Take trace across and use the identity that $\mathrm{Tr}[A \oplus B] = \mathrm{Tr}[A] + \mathrm{Tr}[B]$, we have $\mathrm{F}_R(P, Q) = \mathrm{F}_{R_1}(P_1, Q_1) + \mathrm{F}_{R_2}(P_2, Q_2)$.

7. **Unitary invariance.** Given $P, Q, R \in \mathbb{P}_d^+$ and $U \in \mathbb{U}_d$, we want to show that

$$\mathrm{F}_{URU^*}(UPU^*, UQU^*) = \mathrm{F}_R(P, Q). \tag{C.13}$$

Observe that $\sqrt{VAV^*} = V\sqrt{A}V^*$ for any $V \in \mathbb{U}_d$ and $A \in \mathbb{P}_d^+$. Moreover, $(URU^*)^{-1} = UR^{-1}U^*$. Substituting this in the definition, we get the required relation.

8. **Unitary contravariance.** Given $P, Q, R \in \mathbb{P}_d^+$ and $U \in \mathbb{U}_d$, we want to show that

$$\mathrm{F}_{URU^*}(P, Q) = \mathrm{F}_R(U^*PU, U^*QU). \tag{C.14}$$

By definition, we have

$$\begin{aligned}\mathrm{F}_{URU^*}(P,Q) &:= \mathrm{Tr}\left[\sqrt{(UR^{\frac{1}{2}}U^*)P(UR^{\frac{1}{2}}U^*)}(UR^{-1}U^*)\sqrt{(UR^{\frac{1}{2}}U^*)Q(UR^{\frac{1}{2}}U^*)}\right]\\ &= \mathrm{Tr}\left[U\sqrt{R^{\frac{1}{2}}U^*P^{\frac{1}{2}}UR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}U^*Q^{\frac{1}{2}}UR^{\frac{1}{2}}}U^*\right]\\ &= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}\left(U^*P^{\frac{1}{2}}U\right)R^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}\left(U^*Q^{\frac{1}{2}}U\right)R^{\frac{1}{2}}}\right]\\ &= \mathrm{F}_R(U^*PU, U^*QU)\,.\end{aligned} \tag{C.15}$$

9. **Scaling**. For positive scalars $p, q, r \in \mathbb{R}_+$, we want to show that

$$\mathrm{F}_{rR}(pP, qQ) = \sqrt{pq}\,\mathrm{F}_R(P,Q). \tag{C.16}$$

The result follows directly from substitution.

10. **Upper bound on absolute value.** For any triple $P, Q, R \in \mathbb{P}_d^+$, we want to show that

$$|\,\mathrm{F}_R(P,Q)| \leq \mathrm{F}^{\mathrm{U}}(P,Q). \tag{C.17}$$

We have

$$|\,\mathrm{F}_R(P,Q)| = \left|\mathrm{Tr}\left[P^{\frac{1}{2}}U_PU_Q^*Q^{\frac{1}{2}}\right]\right| \leq \max_{V\in\mathbb{U}_d}\left|\mathrm{Tr}\left[P^{\frac{1}{2}}VQ^{\frac{1}{2}}\right]\right| = \mathrm{F}^{\mathrm{U}}(P,Q), \tag{C.18}$$

where the first equality comes from the alternative characterization of generalized fidelity, and the last equality comes from the variational characterization of Uhlmann fidelity.

11. **Reduction to named fidelities** We now show that for specific choices of the base R, we can recover the Uhlmann-, Holevo-, and Matsumoto fidelities.

    (a) **Uhlmann fidelity:** $R \in \{P, Q\}$. Without loss of generality, choose $R = P$. We then have

$$\begin{aligned}\mathrm{F}_R(P,Q) &:= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right]\\ &= \mathrm{Tr}\left[P\cdot P^{-1}\sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}}\right] = \mathrm{F}^{\mathrm{U}}(P,Q).\end{aligned} \tag{C.19}$$

    The case where $R = Q \implies \mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{U}}(P,Q)$ is derived similarly.

    (b) **Holevo fidelity**. Choose $R = \mathbb{I}$. We then have

$$\begin{aligned}\mathrm{F}_R(P,Q) &:= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right]\\ &= \mathrm{Tr}\left[P^{\frac{1}{2}}Q^{\frac{1}{2}}\right] = \mathrm{F}^{\mathrm{H}}(P,Q).\end{aligned} \tag{C.20}$$

(c) **Matsumoto fidelity:** $R \in \{P^{-1}, Q^{-1}\}$. Without loss of generality, choose $R = P^{-1}$. We then have

$$\begin{aligned} \mathrm{F}_R(P,Q) &:= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right] \\ &= \mathrm{Tr}\left[\mathbb{I}\cdot P\sqrt{P^{-\frac{1}{2}}QP^{-\frac{1}{2}}}\right] = \mathrm{Tr}[P\#Q] = \mathrm{F}^{\mathrm{M}}(P,Q). \end{aligned} \tag{C.21}$$

The case where $R = Q^{-1} \implies \mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q)$ is derived similarly.

C.2 Useful lemmas

In this section, we list various supporting results used in the main proofs.

The following proposition characterizes the *named fidelities* between P and Q in terms of $P^{\frac{1}{2}}, Q^{\frac{1}{2}}$ and a unitary matrix.

Proposition C.2.1. *Let $P, Q \in \mathbb{P}_d$. Then we have*

$$\mathrm{F}^{\mathrm{U}}(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}}UQ^{\frac{1}{2}}\right], \tag{C.22}$$

$$\mathrm{F}_{\mathrm{M}}(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}}VQ^{\frac{1}{2}}\right], \tag{C.23}$$

$$\mathrm{F}_{\mathrm{H}}(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}}\mathbb{I}Q^{\frac{1}{2}}\right], \tag{C.24}$$

where $U, V \in \mathbb{U}_d$ such that $U = \mathrm{Pol}\left(P^{\frac{1}{2}}Q^{\frac{1}{2}}\right)$ and $V = \mathrm{Pol}\left(P^{-\frac{1}{2}}Q^{\frac{1}{2}}\right)$. Moreover, V is the unique unitary such that $P^{\frac{1}{2}}VQ^{\frac{1}{2}}$ is Hermitian.

Proof. The proof of the first statement directly follows from the definition of Uhlmann fidelity. Indeed by Lemma A.2.1, we have

$$Q^{\frac{1}{2}}P^{\frac{1}{2}}U = \sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}}, \tag{C.25}$$

whence it follows

$$P^{\frac{1}{2}}UQ^{\frac{1}{2}} = Q^{-\frac{1}{2}}\sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}}Q^{\frac{1}{2}}. \tag{C.26}$$

Take trace and use its cyclic nature to obtain

$$\mathrm{F}^{\mathrm{U}}(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}}UQ^{\frac{1}{2}}\right]. \tag{C.27}$$

For the second statement, recall that $\mathrm{F}_{\mathrm{M}}(P,Q) = \mathrm{Tr}[P\#Q]$. The fact that there exists a unique unitary V such that $P\#Q = P^{\frac{1}{2}}VQ^{\frac{1}{2}}$ is proven in multiple sources including [Bha09, Proposition 4.1.8] and [CS20]. Thus, we omit the proof

of this part and instead prove that $V = \mathrm{Pol}(P^{-\frac{1}{2}}Q^{\frac{1}{2}})$. To this end consider the geometric mean of P and Q:

$$P\#Q = P^{\frac{1}{2}}\sqrt{P^{-\frac{1}{2}}QP^{-\frac{1}{2}}}P^{\frac{1}{2}} = P^{\frac{1}{2}}VQ^{\frac{1}{2}} \in \mathbb{P}_d^+ \tag{C.28}$$

for some unitary $V \in \mathbb{U}_d$. Now left and right multiply by $P^{-\frac{1}{2}}$ to obtain

$$P^{-\frac{1}{2}}(P\#Q)P^{-\frac{1}{2}} = \sqrt{P^{-\frac{1}{2}}QP^{-\frac{1}{2}}} = VQ^{\frac{1}{2}}P^{-\frac{1}{2}}. \tag{C.29}$$

By definition, the polar decomposition of $Q^{\frac{1}{2}}P^{-\frac{1}{2}}$ is

$$Q^{\frac{1}{2}}P^{-\frac{1}{2}} = W\left|Q^{\frac{1}{2}}P^{-\frac{1}{2}}\right| = W\sqrt{P^{-\frac{1}{2}}QP^{-\frac{1}{2}}}, \tag{C.30}$$

where $W = \mathrm{Pol}\left(Q^{\frac{1}{2}}P^{-\frac{1}{2}}\right)$. Comparing with the previous equation, we get $V^* = W$. This completes the proof. □

One can use the above Proposition to provide sufficient conditions when the generalized fidelity will equal these fidelities.

Corollary C.2.2. *Let $P, Q, R \in \mathbb{P}_d$. Then*

$$\mathrm{F}_R(Q,P) = \mathrm{Tr}\left[P^{\frac{1}{2}}U_PU_Q^*Q^{\frac{1}{2}}\right] = \begin{cases} \mathrm{F}^{\mathrm{U}}(P,Q) & \text{if} \quad U_PU_Q^* = U \\ \mathrm{F}^{\mathrm{M}}(P,Q) & \text{if} \quad U_PU_Q^* = V \\ \mathrm{F}^{\mathrm{H}}(P,Q) & \text{if} \quad U_PU_Q^* = \mathbb{I} \text{ or } U_P = U_Q \end{cases} \tag{C.31}$$

where

$$\begin{aligned} U &:= \mathrm{Pol}\left(P^{\frac{1}{2}}Q^{\frac{1}{2}}\right), \quad V := \mathrm{Pol}\left(P^{-\frac{1}{2}}Q^{\frac{1}{2}}\right), \\ U_P &:= \mathrm{Pol}\left(P^{\frac{1}{2}}R^{\frac{1}{2}}\right), \quad U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}}R^{\frac{1}{2}}\right). \end{aligned} \tag{C.32}$$

We shall use the above corollary to prove certain geometric results of generalized fidelity. The next lemma lists some defining properties of the Bures–Wasserstein geodesic.

Lemma C.2.3. *Let $A, B \in \mathbb{P}_d^+$ and $C = \gamma_{AB}^{\mathrm{BW}}(t)$ for any $t \in [0,1]$. Then, the following statements hold true.*

$$C = [(1-t)\mathbb{I} + t(A^{-1}\#B)]A[(1-t)\mathbb{I} + t(A^{-1}\#B)], \tag{C.33}$$

$$C = (1-t)\sqrt{C^{\frac{1}{2}}AC^{\frac{1}{2}}} + t\sqrt{C^{\frac{1}{2}}BC^{\frac{1}{2}}}, \tag{C.34}$$

$$\mathbb{I} = (1-t)C^{-1}\#A + tC^{-1}\#B, \tag{C.35}$$

where the fixed-point equation (Eq. (C.34)) is uniquely satisfied by C.

Proof. The first equation has been proven in multiple sources such as [BJL19, Equation 39]. The second equation can be seen as the $n = 2$ version of the fixed-point equation satisfied by the BW barycenter (see [AKF22; Alt+23]). To obtain the third equation, conjugate the LHS and RHS of the second statement with $C^{-\frac{1}{2}}$. □

Next, we present a well-known result regarding Affine-invariant geodesics which states. This, along with the following lemmas, will be useful in proving results related to the geometric mean.

Lemma C.2.4. *Let $A, B \in \mathbb{P}_d$. Then, for any invertible X with matching dimensions,*

$$X(\gamma_{A,B}^{\mathrm{AI}}(t))X^* = \gamma_{XAX^*,XBX^*}^{\mathrm{AI}}(t), \tag{C.36}$$

for any $t \in [0,1]$.

Proof. See [LL24, Theorem 5.1 and Remark 5.2] for proof. □

The following lemma is useful to state a sufficient condition for generalized fidelity to reduce to Matsumoto fidelity.

Lemma C.2.5. *Let $P, Q, R \in \mathbb{P}_d$. Then, the following statements are equivalent.*

1. $R^{-\frac{1}{2}}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}R^{-\frac{1}{2}} = P\#Q.$
2. $\left[R^{\frac{1}{2}}PR^{\frac{1}{2}}, R^{\frac{1}{2}}QR^{\frac{1}{2}}\right] = 0.$
3. $PRQ = QRP.$

Here $[A, B] := AB - BA$ denotes the commutator of A, B.

Proof. We will establish the equivalence by showing (1) $\iff$ (2) and (2) $\iff$ (3). We begin with (1) $\implies$ (2). Assume

$$R^{-\frac{1}{2}}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}R^{-\frac{1}{2}} = P\#Q, \tag{C.37}$$

which implies

$$\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}} = R^{\frac{1}{2}}P\#QR^{\frac{1}{2}} = \sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}, \tag{C.38}$$

where the last equality comes from the fact that $P\#Q$ is positive definite (and thus Hermitian). Hence we have

$$\left[\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}, \sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\right] = 0, \tag{C.39}$$

which implies their squares also commute, thereby completing the proof.

Now for the reverse implication (2) $\implies$ (1), assume $\left[R^{\frac{1}{2}}PR^{\frac{1}{2}}, R^{\frac{1}{2}}QR^{\frac{1}{2}}\right] = 0$. Then their square roots also commute, and thus we have

$$\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}} > 0 \iff R^{-\frac{1}{2}}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}R^{-\frac{1}{2}} > 0. \tag{C.40}$$

By polar decomposition, we can write the RHS as

$$0 < R^{-\frac{1}{2}} \cdot R^{\frac{1}{2}}P^{\frac{1}{2}}U_P \cdot U_Q^*Q^{\frac{1}{2}}R^{\frac{1}{2}} \cdot R^{-\frac{1}{2}} = P^{\frac{1}{2}}U_PU_Q^*Q^{\frac{1}{2}}, \tag{C.41}$$

where $U_P := \mathrm{Pol}\left(P^{\frac{1}{2}}R^{\frac{1}{2}}\right)$ and $U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}}R^{\frac{1}{2}}\right)$. By Proposition C.2.1, if $A, B > 0$ and V is a unitary matrix such that $A^{\frac{1}{2}}VB^{\frac{1}{2}} > 0$, then $A^{\frac{1}{2}}VB^{\frac{1}{2}} = A\#B$. Thus we have

$$R^{-\frac{1}{2}}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}R^{-\frac{1}{2}} = P^{\frac{1}{2}}U_PU_Q^*Q^{\frac{1}{2}} = P\#Q, \tag{C.42}$$

which proves the reverse implication. Now we prove (2) $\iff$ (3). This is easily seen as

$$\left[R^{\frac{1}{2}}PR^{\frac{1}{2}}, R^{\frac{1}{2}}QR^{\frac{1}{2}}\right] = 0 \quad \Leftrightarrow \quad R^{\frac{1}{2}}PRQR^{\frac{1}{2}} = R^{\frac{1}{2}}QRPR^{\frac{1}{2}} \quad \Leftrightarrow \quad PRQ = QRP. \tag{C.43}$$

This concludes the proof. □

We note that triples of matrices of these forms have been studied as Γ-*commuting matrices* in [LL08]. Observe that the first statement of this lemma is a sufficient condition for the generalized fidelity to be equal to the Matsumoto fidelity (take trace across and use cyclicity on the LHS). Thus, we have the fol-

lowing corollary.

Corollary C.2.6. *Let $P, Q, R \in \mathbb{P}_d^+$. If $PRQ = QRP$, then $\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q)$.*

Proof. The proof follows directly from Statements 1 and 3 of Lemma C.2.5. □

C.3 Proofs of geometric properties

In this section, we provide the proofs of the geometric results results discussed in Chapter 4.

Theorem C.3.1 (Path 2; restated from Theorem 4.2.1)**.** *Let $P, Q \in \mathbb{P}_d^+$ be fixed. Let the base $R = \gamma_{PQ}^{\mathrm{BW}}(t)$ for any $t \in [0,1]$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{U}}(P,Q). \tag{C.44}$$

Proof. Let

$$U_P := \mathrm{Pol}\left(P^{\frac{1}{2}}R^{\frac{1}{2}}\right), \quad U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}}R^{\frac{1}{2}}\right), \quad \text{and} \quad U := \mathrm{Pol}\left(P^{\frac{1}{2}}Q^{\frac{1}{2}}\right) \tag{C.45}$$

be polar factors. Using the alternate representation of generalized fidelity, we have

$$\mathrm{F}_R(P,Q) = \mathrm{Tr}\left[Q^{\frac{1}{2}}U_Q U_P^* P^{\frac{1}{2}}\right] \tag{C.46}$$

and noting that $\mathrm{F}^{\mathrm{U}}(P,Q) = \mathrm{Tr}\left[P^{\frac{1}{2}}UQ^{\frac{1}{2}}\right]$, it suffices to show that $U_P U_Q^* = U$. By Eq. (C.35), we have

$$\mathbb{I} = (1-t)R^{-1}\#P + tR^{-1}\#Q \equiv M + N, \tag{C.47}$$

where, for simplicity, we have denoted $M \equiv (1-t)R^{-1}\#P$ and $N \equiv tR^{-1}\#Q$. We thus have $M, N \in \mathbb{P}_d$ and $M + N = \mathbb{I}$, which implies

$$[M,N] = [M, \mathbb{I} - M] = 0, \tag{C.48}$$

which implies $R^{-1}\#P$ commutes with $R^{-1}\#Q$. Thus, the product of these positive matrices is also positive:

$$0 < (R^{-1}\#P)(R^{-1}\#Q) = R^{-\frac{1}{2}}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}R^{-\frac{1}{2}}, \tag{C.49}$$

which implies

$$\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}} = U_Q^*Q^{\frac{1}{2}}P^{\frac{1}{2}}U_P = U_P^*P^{\frac{1}{2}}Q^{\frac{1}{2}}U_Q > 0, \tag{C.50}$$

where the first equality follows from polar decomposition. Conjugating with U_P we get

$$U_PU_Q^*Q^{\frac{1}{2}}P^{\frac{1}{2}} = P^{\frac{1}{2}}Q^{\frac{1}{2}}U_QU_P^* > 0, \tag{C.51}$$

whence it follows that

$$\left(U_PU_Q^*Q^{\frac{1}{2}}P^{\frac{1}{2}}\right)^2 = \left(P^{\frac{1}{2}}Q^{\frac{1}{2}}U_QU_P^*\right)\left(U_PU_Q^*Q^{\frac{1}{2}}P^{\frac{1}{2}}\right) = P^{\frac{1}{2}}QP^{\frac{1}{2}}. \tag{C.52}$$

Since the positive semidefinite matrices have unique positive semidefinite square roots, it follows that

$$U_PU_Q^*Q^{\frac{1}{2}}P^{\frac{1}{2}} = \sqrt{P^{\frac{1}{2}}QP^{\frac{1}{2}}}, \tag{C.53}$$

which implies $U_PU_Q^* = U = \mathrm{Pol}\left(P^{\frac{1}{2}}Q^{\frac{1}{2}}\right)$, which then further implies

$$\mathrm{F}_R(Q,P) = \mathrm{Tr}\left[U_PU_Q^*Q^{\frac{1}{2}}P^{\frac{1}{2}}\right] = \mathrm{Tr}\left[P^{\frac{1}{2}}UQ^{\frac{1}{2}}\right] = \mathrm{F}^{\mathrm{U}}(P,Q), \tag{C.54}$$

where $R = R_t$ for any $t \in [0,1]$. This completes the proof. □

Theorem C.3.2 (Path 1, restated from 4.2.2)**.** *Let $P, Q \in \mathbb{P}_d^+$ be fixed. Let the base R be any point along the path $[\gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t)]^{-1}$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{U}}(P,Q). \tag{C.55}$$

Proof. This theorem will be proved similarly to the previous one. Let us begin with the form of R:

$$R = [\gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t)]^{-1} \iff R^{-1} = \gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t). \tag{C.56}$$

Similar to the previous proof, let

$$U_P := \mathrm{Pol}\left(P^{\frac{1}{2}}R^{\frac{1}{2}}\right), \quad U_Q := \mathrm{Pol}\left(Q^{\frac{1}{2}}R^{\frac{1}{2}}\right), \quad \text{and} \quad U := \mathrm{Pol}\left(P^{\frac{1}{2}}Q^{\frac{1}{2}}\right). \tag{C.57}$$

We will show that $U_PU_Q^* = U$. By Eq. (C.35) (with $(A,B,C) = (P^{-1}, Q^{-1}, R^{-1})$) we have

$$\mathbb{I} = (1-t)R\#P^{-1} + tR\#Q^{-1} = M + N, \tag{C.58}$$

where we wrote $M \equiv (1-t)R\#P^{-1}$ and $N = (1-t)R\#Q^{-1}$ for simplicity. We

thus have $M, N \in \mathbb{P}_d$ with $M + N = \mathbb{I}$. It follows that

$$[M, N] = [M, \mathbb{I} - M] = 0, \tag{C.59}$$

which implies $P^{-1}\#R$ commutes with $Q^{-1}\#R$. We thus have their product to be Hermitian and positive definite:

$$R^{\frac{1}{2}}\sqrt{R^{-\frac{1}{2}}P^{-1}R^{-\frac{1}{2}}}R^{\frac{1}{2}}R^{\frac{1}{2}}\sqrt{R^{-\frac{1}{2}}Q^{-1}R^{-\frac{1}{2}}}R^{\frac{1}{2}} > 0, \tag{C.60}$$

which implies

$$\sqrt{R^{-\frac{1}{2}}P^{-1}R^{-\frac{1}{2}}}R\sqrt{R^{-\frac{1}{2}}Q^{-1}R^{-\frac{1}{2}}} > 0. \tag{C.61}$$

Now take inverse across to get

$$\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}} = U_P^* P^{\frac{1}{2}}Q^{\frac{1}{2}}U_Q = U_Q^* Q^{\frac{1}{2}}P^{\frac{1}{2}}U_P > 0, \tag{C.62}$$

where the first equality follows from polar decomposition. The rest of the proof mirrors the last part of the previous proof, which we write for the sake of completion. Conjugating $U_Q^* Q^{\frac{1}{2}}P^{\frac{1}{2}}U_P > 0$ with U_Q we get

$$Q^{\frac{1}{2}}P^{\frac{1}{2}}U_P U_Q^* = U_Q U_P^* P^{\frac{1}{2}}Q^{\frac{1}{2}} > 0, \tag{C.63}$$

whence it follows that

$$\left(Q^{\frac{1}{2}}P^{\frac{1}{2}}U_P U_Q^*\right)^2 = \left(Q^{\frac{1}{2}}P^{\frac{1}{2}}U_P U_Q^*\right)\left(U_Q U_P^* P^{\frac{1}{2}}Q^{\frac{1}{2}}\right) = Q^{\frac{1}{2}}PQ^{\frac{1}{2}}. \tag{C.64}$$

Since the positive semidefinite matrices have unique positive semidefinite square roots, we have

$$Q^{\frac{1}{2}}P^{\frac{1}{2}}U_P U_Q^* = \sqrt{Q^{\frac{1}{2}}PQ^{\frac{1}{2}}}, \tag{C.65}$$

which implies $U_P U_Q^* = U = \mathrm{Pol}\left(P^{\frac{1}{2}}Q^{\frac{1}{2}}\right)$. Thus we have

$$\begin{aligned} \mathrm{F}_R(P, Q) &= \mathrm{Tr}\left[\sqrt{R^{\frac{1}{2}}P^1R^{\frac{1}{2}}}R^{-1}\sqrt{R^{\frac{1}{2}}Q^1R^{\frac{1}{2}}}\right] \\ &= \mathrm{Tr}\left[U_P U_Q^* Q^{\frac{1}{2}}P^{\frac{1}{2}}\right] = \mathrm{Tr}\left[P^{\frac{1}{2}}UQ^{\frac{1}{2}}\right] = \mathrm{F}^{\mathrm{U}}(P, Q), \end{aligned} \tag{C.66}$$

where $R = [\gamma^{\mathrm{BW}}_{P^{-1}Q^{-1}}(t)]^{-1}$ for any $t \in [0, 1]$. This completes the proof. □

Theorem C.3.3 (Restated from Theorem 4.2.3)**.** *Let $P, Q \in \mathbb{P}_d$ be fixed. For any fixed $t \in [0,1]$, define*

$$R_1 := \gamma^{\mathrm{BW}}_{PQ^{-1}}(t) \quad \textit{and} \quad R_2 := \gamma^{\mathrm{BW}}_{QP^{-1}}(t). \tag{C.67}$$

Then $\mathrm{F}_{R_1}(P,Q) = \mathrm{F}_{R_2}(P,Q)$.

Proof. Fix arbitrary $P, Q \in \mathbb{P}_d$, $t \in [0,1]$ and define R_1 and R_2 as above. We aim to show that the generalized fidelities between P and Q at these bases are equal. Recall that the generalized fidelity can also be written as

$$\mathrm{F}_{R_1}(P,Q) = \mathrm{Tr}\left[(R_1^{-1}\#Q)R_1(R_1^{-1}\#P)\right], \quad \mathrm{F}_{R_2}(P,Q) = \mathrm{Tr}\left[(R_2^{-1}\#Q)R_2(R_2^{-1}\#P)\right]. \tag{C.68}$$

Thus, to prove the theorem, it suffices to show that the terms inside the trace are equal, which will be done in two steps. We first prove that

$$(R_1^{-1}\#Q)R_1(R_1^{-1}\#P) = \sqrt{QMPM}M^{-1}, \tag{C.69}$$
$$(R_2^{-1}\#Q)R_2(R_2^{-1}\#P) = M^{-1}\sqrt{MQMP}, \tag{C.70}$$

for a particular $M \in \mathbb{P}_d^+$, whose form is described later. In the second step, we show that the two RHS terms are equal.

Let us perform the first step. To this end, we first write down the explicit forms of R_1 and R_2. For an arbitrary fixed $t \in [0,1]$ we have

$$R_1 = \gamma^{\mathrm{BW}}_{PQ^{-1}}(t) = \left[(1-t)\mathbb{I} + tP^{-1}\#Q^{-1}\right] P \left[(1-t)\mathbb{I} + tP^{-1}\#Q^{-1}\right], \tag{C.71}$$

$$R_2 = \gamma^{\mathrm{BW}}_{QP^{-1}}(t) = \left[(1-t)\mathbb{I} + tQ^{-1}\#P^{-1}\right] Q \left[(1-t)\mathbb{I} + tQ^{-1}\#P^{-1}\right]. \tag{C.72}$$

Denote $M \equiv [(1-t)\mathbb{I} + tP^{-1}\#Q^{-1}]$. Thus, and noting that the geometric mean is symmetric, we may write

$$R_1 = MPM \quad \text{and} \quad R_2 = MQM. \tag{C.73}$$

It follows that

$$R_1\#P^{-1} = M = R_2\#Q^{-1}. \tag{C.74}$$

Moreover by Lemma A.2.1, we have that $(A^{-1}\#B)A = \sqrt{BA}$ and $A(A^{-1}\#B) = \sqrt{AB}$ for any $A, B \in \mathbb{P}_d^+$. Thus, we can write

$$(R_1^{-1}\#Q)R_1 = \sqrt{QR_1} \quad \text{and} \quad R_2(R_2^{-1}\#P) = \sqrt{R_2P}. \tag{C.75}$$

By inverting Eq. (C.74), and using the above relation, we have

$$(R_1^{-1}\#Q)R_1(R_1^{-1}\#P) = \sqrt{QR_1}M^{-1} = \sqrt{QMPM}M^{-1}, \tag{C.76}$$
$$(R_2^{-1}\#Q)R_2(R_2^{-1}\#P) = M^{-1}\sqrt{R_2P} = M^{-1}\sqrt{MQMP}, \tag{C.77}$$

as claimed, where we also used the relations $R_1 = MPM$ and $R_2 = MQM$. Now we prove that the RHS terms are equal, which amounts to proving,

$$\sqrt{QMPM}M^{-1} \stackrel{?}{=} M^{-1}\sqrt{MQMP}. \tag{C.78}$$

This is equivalent to proving

$$\sqrt{QMPM} \stackrel{?}{=} M^{-1}\sqrt{MQMP}M. \tag{C.79}$$

To prove this, it suffices to prove the squared version as matrices with positive eigenvalues have a unique square root (see [Bha13, Excercise 4.5.2]). Square the LHS to obtain $QMPM$. Now square the RHS to obtain

$$\left(M^{-1}\sqrt{MQMP}M\right)\left(M^{-1}\sqrt{MQMP}M\right) = M^{-1}(MQMP)M = QMPM, \tag{C.80}$$

which is equal to the square of the LHS. By uniqueness of square root, we have

$$(R_1^{-1}\#Q)R_1(R_1^{-1}\#P) = (R_2^{-1}\#Q)R_2(R_2^{-1}\#P). \tag{C.81}$$

Trace both sides to obtain $\mathrm{F}_{R_1}(P,Q) = \mathrm{F}_{R_2}(P,Q)$ as claimed. □

Theorem C.3.4 (Restated from Theorem 4.2.4)**.** *Let $P, Q \in \mathbb{P}_d^+$ be fixed. For any fixed $t \in [0,1]$, let*

$$R_1 := [\gamma^{\mathrm{BW}}_{P^{-1}Q}(t)]^{-1} \quad \textit{and} \quad R_2 := [\gamma^{\mathrm{BW}}_{Q^{-1}P}(t)]^{-1}. \tag{C.82}$$

Then $\mathrm{F}_{R_1}(P,Q) = \mathrm{F}_{R_2}(P,Q)$.

Proof. The proof works similarly to the previous proof. We aim to show that for arbitrary $P, Q \in \mathbb{P}_d$ and any $t \in [0,1]$,

$$R_1^{-1} = \gamma^{\mathrm{BW}}_{P^{-1}Q}(t) \quad \text{and} \quad R_2^{-1} = \gamma^{\mathrm{BW}}_{Q^{-1}P}(t) \implies \mathrm{F}_{R_1}(P,Q) = \mathrm{F}_{R_2}(P,Q). \tag{C.83}$$

We first recall the alternative form of generalized fidelity:

$$\begin{aligned}\mathrm{F}_{R_1}(P,Q) &= \mathrm{Tr}\left[(R_1^{-1}\#Q)R_1(R_1^{-1}\#P)\right] \\ \mathrm{F}_{R_2}(P,Q) &= \mathrm{Tr}\left[(R_2^{-1}\#Q)R_2(R_2^{-1}\#P)\right].\end{aligned} \tag{C.84}$$

We will show that the terms inside the trace are equal, which will be done in two steps. We first prove that

$$(R_1^{-1}\#Q)R_1(R_1^{-1}\#P) = \sqrt{QN^{-1}PN^{-1}}N, \tag{C.85}$$
$$(R_2^{-1}\#Q)R_2(R_2^{-1}\#P) = N\sqrt{N^{-1}QN^{-1}P}, \tag{C.86}$$

for a particular choice of $N \in \mathbb{P}_d^+$, whose form is described later. In the second step, we show that the two RHS terms are equal.

To this end, we first write down the explicit forms of R_1^{-1} and R_2^{-1}. For an arbitrary fixed $t \in [0,1]$ we have

$$R_1^{-1} = \gamma^{\mathrm{BW}}_{P^{-1}Q}(t) := [(1-t)\mathbb{I} + tP\#Q]\, P^{-1}\, [(1-t)\mathbb{I} + tP\#Q]\,, \tag{C.87}$$

$$R_2 = \gamma^{\mathrm{BW}}_{Q^{-1}P}(t) := [(1-t)\mathbb{I} + tQ\#P]\, Q^{-1}\, [(1-t)\mathbb{I} + tQ\#P]\,. \tag{C.88}$$

Denote $N \equiv [(1-t)\mathbb{I} + tP\#Q]$. Thus, the above relations can be written as

$$R_1^{-1} = NP^{-1}N \quad \text{and} \quad R_2^{-1} = NQ^{-1}N. \tag{C.89}$$

which implies

$$P\#R_1^{-1} = N = Q\#R_2^{-1}. \tag{C.90}$$

Using Lemma A.2.1 as before, we have that $(A^{-1}\#B)A = \sqrt{BA}$ and $A(A^{-1}\#B) = \sqrt{AB}$ for any $A, B \in \mathbb{P}_d^+$. Thus, we can write

$$\begin{aligned}(R_1^{-1}\#Q)R_1 &= \sqrt{QR_1} = \sqrt{QN^{-1}PN^{-1}} \\ R_2(R_2^{-1}\#P) &= \sqrt{R_2P} = \sqrt{N^{-1}QN^{-1}P}.\end{aligned} \tag{C.91}$$

Using Eq. (C.90) and the above relation, we have

$$(R_1^{-1}\#Q)R_1(R_1^{-1}\#P) = \sqrt{QN^{-1}PN^{-1}}N, \tag{C.92}$$
$$(R_2^{-1}\#Q)R_2(R_2^{-1}\#P) = N\sqrt{N^{-1}QN^{-1}P} \tag{C.93}$$

as claimed. Now we prove that the RHS terms are equal:

$$\sqrt{QN^{-1}PN^{-1}}N \stackrel{?}{=} N\sqrt{N^{-1}QN^{-1}P}, \tag{C.94}$$

which is equivalent to proving

$$\sqrt{QN^{-1}PN^{-1}} \stackrel{?}{=} N\sqrt{N^{-1}QN^{-1}P}N^{-1}, \tag{C.95}$$

As in the previous proof, it suffices to prove the squared version of the above

equation. First, square the LHS to obtain $QN^{-1}PN^{-1}$. Now square the RHS to obtain

$$\begin{aligned}\left(N\sqrt{N^{-1}QN^{-1}P}N^{-1}\right)\left(N\sqrt{N^{-1}QN^{-1}P}N^{-1}\right) &= N(N^{-1}QN^{-1}P)N \\ &= QN^{-1}PN^{-1},\end{aligned} \tag{C.96}$$

which is equal to the square of LHS. We have thus shown, by the uniqueness of the positive square root,

$$(R_1^{-1}\#Q)R_1(R_1^{-1}\#P) = (R_2^{-1}\#Q)R_2(R_2^{-1}\#P). \tag{C.97}$$

Taking the trace across, we get $\mathrm{F}_{R_1}(P,Q) = \mathrm{F}_{R_2}(P,Q)$ as claimed. □

Theorem C.3.5 (Restated from Theorem 4.2.5)**.** *Let $P, Q \in \mathbb{P}_d$ and choose $R = P^{-1}\#Q^{-1}$. Then $\mathrm{F}_R(P,Q) = \mathrm{F}_{\mathrm{M}}(P,Q)$.*

Proof. By definition, we have

$$\mathrm{F}_R(P,Q) = \mathrm{F}_R(P,Q) = \mathrm{Tr}\left(\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}}(P\#Q)\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right), \tag{C.98}$$

where we used the property of the geometric mean that $(A\#B)^{-1} = A^{-1}\#B^{-1}$ for any $A, B \in \mathbb{P}_d^+$. Recall that $\mathrm{F}_{\mathrm{M}}(P,Q) = \mathrm{Tr}[P\#Q]$ by definition. Thus it suffices to show that $\sqrt{R^{\frac{1}{2}}PR^{\frac{1}{2}}} \stackrel{?}{=} \left(\sqrt{R^{\frac{1}{2}}QR^{\frac{1}{2}}}\right)^{-1}$. Equivalently, it suffices to show that their squares equal:

$$R^{\frac{1}{2}}PR^{\frac{1}{2}} \stackrel{?}{=} R^{-\frac{1}{2}}Q^{-1}R^{-\frac{1}{2}} \iff RPR \stackrel{?}{=} Q^{-1}. \tag{C.99}$$

By Proposition 2.1.3, we have that the geometric mean $R = P^{-1}\#Q^{-1}$ uniquely satisfies the second condition, and thus we have $R^{\frac{1}{2}}PR^{\frac{1}{2}} = R^{-\frac{1}{2}}Q^{-1}R^{-\frac{1}{2}}$, which implies their matrix square roots equal, which in turn implies

$$\mathrm{F}_R(P,Q) = \mathrm{Tr}[P\#Q] = \mathrm{F}^{\mathrm{M}}(P,Q). \tag{C.100}$$

This completes the proof. □

Now we prove the result for the whole curve $\gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}$.

Theorem C.3.6 (Path 9, restated from Theorem 4.2.2)**.** *Let $P, Q \in \mathbb{P}_d^+$ and let $R = \gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t)$ for any $t \in [0,1]$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q). \tag{C.101}$$

Proof. Let

$$R = \gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t) = P^{-\frac{1}{2}} \left(P^{\frac{1}{2}} Q^{-1} P^{\frac{1}{2}}\right)^t P^{-\frac{1}{2}} \tag{C.102}$$

for some $t \in [0,1]$. By Corollary C.2.6, it is sufficient to show that

$$PRQ \stackrel{?}{=} QRP \tag{C.103}$$

Let us begin with the LHS. We first left and right multiply by $P^{-\frac{1}{2}}$ to obtain

$$\begin{aligned} P^{-\frac{1}{2}}(PRQ)P^{-\frac{1}{2}} &= P^{\frac{1}{2}} R P^{\frac{1}{2}} \left(\frac{Q}{P}\right) \\ &\stackrel{1}{=} (P^{\frac{1}{2}} \gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}(t) P^{\frac{1}{2}}) \left(\frac{Q}{P}\right) \\ &\stackrel{2}{=} \gamma^{\mathrm{AI}}_{\mathbb{I}, P^{\frac{1}{2}} Q^{-1} P^{\frac{1}{2}}}(t) \left(\frac{Q}{P}\right) \\ &\stackrel{3}{=} \left(\frac{Q}{P}\right)^{-t} \left(\frac{Q}{P}\right) = \left(\frac{Q}{P}\right)^{1-t} = \left(P^{-\frac{1}{2}} Q P^{-\frac{1}{2}}\right)^{1-t} \end{aligned} \tag{C.104}$$

where we use the shorthand $\frac{A}{B} := B^{-\frac{1}{2}} A B^{-\frac{1}{2}}$. Here in (1) we have used the fact that R is an element of the geodesic $\gamma^{\mathrm{AI}}_{P^{-1}Q^{-1}}$, in (2) we have used the Affine invariance property (Lemma C.2.4, and in (3) we used the definition of the AI geodesic. A similar calculation on the RHS reveals $P^{-\frac{1}{2}}(QRP)P^{-\frac{1}{2}} = \left(\frac{Q}{P}\right)^{1-t} = (P^{-\frac{1}{2}} Q P^{-\frac{1}{2}})^{1-t}$, which implies $PRQ = QRP$. In then follows from Corollary C.2.6 that $\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q)$. This concludes the proof.

□

Theorem C.3.7 (Path 10, restated from Theorem 4.2.8)**.** *Let $P, Q \in \mathbb{P}_d^+$ and let $R = \gamma^{\mathrm{Euc}}_{P^{-1}Q^{-1}}(t)$ for any $t \in [0,1]$. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q). \tag{C.105}$$

Proof. By Lemma C.2.5, it suffices to show that $PRQ = QRP$. Recall that R takes the form $R = (1-t)P^{-1} + tQ^{-1}$ for any $t \in [0,1]$. Use this form of R to write

$$\begin{aligned} PRQ &= P\left((1-t)P^{-1} + tQ^{-1}\right) Q = (1-t)Q + tP \\ &= Q\left((1-t)P^{-1} + tQ^{-1}\right) P = QRP. \end{aligned} \tag{C.106}$$

This completes the proof. □

Theorem C.3.8 (Path 11, restated from Theorem 4.2.9)**.** *Let* $P, Q \in \mathbb{P}_d^+$ *and let* $R = [\gamma_{PQ}^{\mathrm{Euc}}(t)]^{-1}$ *for any* $t \in [0,1]$*. Then*

$$\mathrm{F}_R(P,Q) = \mathrm{F}^{\mathrm{M}}(P,Q). \tag{C.107}$$

Proof. Here too we will prove the sufficient condition $PRQ = QRP$, which is equivalent to showing $Q^{-1}R^{-1}P^{-1} = P^{-1}R^{-1}Q^{-1}$. Noting that along the path $[\gamma_{PQ}^{\mathrm{Euc}}(t)]^{-1}$, we have $R = ((1-t)P + tQ)^{-1}$ for $t \in [0,1]$. Thus we have

$$\begin{aligned} P^{-1}R^{-1}Q^{-1} &= P^{-1}\left((1-t)P + tQ\right)Q^{-1} = (1-t)Q^{-1} + tP^{-1} \\ &= Q^{-1}\left((1-t)P + tQ\right)P^{-1} = Q^{-1}R^{-1}P^{-1}, \end{aligned} \tag{C.108}$$

which implies $PRQ = QRP$. By Lemma C.2.5, this completes the proof. □

Appendix D

Proofs related Bures projection

D.1 Proof of Bures projection for pinching channel

Proof. Let $\mathcal{E} = \{E_i\}_{i\in[n]}$ be an arbitrary orthogonal resolution of identity and $\Lambda_\mathcal{E}$ be the corresponding pinching channel. Firstly observe that $P > 0$ implies $\Lambda_\mathcal{E}(P) > 0$. One way to prove this is to observe that the pinching channel is a unital map (as every pinching channel is a mixed-unitary map [Wat18; Tom15]) and unital maps are strictly positive [Bha09, Section 2.2]. Strictly positive maps map positive definite matrices to positive definite matrices. We thus have

$$(\Lambda_\mathcal{E}(P))^{-1} = \left(\sum_{i=1}^n E_i P E_i\right)^{-1} = \sum_{i=1}^n (E_i P E_i)^{-1} = \sum_{i=1}^n (P_i)^{-1}, \tag{D.1}$$

where we denote $P_i \equiv E_i P E_i$ and the inverse is taken in the support. To see this is indeed the inverse, observe that

$$\left(\sum_{i=1}^n E_i P E_i\right) \cdot \left(\sum_{i=j}^n (E_j P E_j)^{-1}\right) = \sum_{i=1}^n E_i P E_i \cdot (E_i P E_i)^{-1} = \sum_{i=1}^n E_i = \mathbb{I}. \tag{D.2}$$

Noting that C is feasible, and thus $C = \Lambda_\mathcal{E}(C)$, compute

$$\Lambda_\mathcal{E}(P)^{-1} \# C = \Lambda_\mathcal{E}(P)^{-1} \# \Lambda_\mathcal{E}(C) = \left(\sum_{i=1}^n P_i^{-1}\right) \# \sum_{j=1}^n C_j = \sum_{i=1}^n P_i^{-1} \# C_i. \tag{D.3}$$

where $C_i \equiv E_i C E_i$ and the last equality comes from the orthogonality of the projectors. We thus have

$$\Gamma_{\Lambda,C}[P] = \left(\sum_{i=1}^n P_i^{-1} \# C_i\right) P \left(\sum_{j=1}^n P_j^{-1} \# C_j\right) = \sum_{i,j=1}^n P_i^{-1} \# C_i P_{ij} P_j^{-1} \# C_j, \tag{D.4}$$

where we denote $P_{ij} = E_i P E_j$. To show that this is a feasible point, let us apply the pinching channel:

$$\begin{aligned}\Lambda_{\mathcal{E}}(\Gamma_{\Lambda,C}[P]) &= \sum_{i=k}^{n} E_k \left(\sum_{i,j=1}^{n} P_i^{-1} \# C_i P_{ij} P_j^{-1} \# C_j \right) E_k \\ &= \sum_{k=1}^{n} P_k^{-1} \# C_k P_k P_k^{-1} \# C_k = \sum_{i=1}^{n} C_k = C,\end{aligned} \tag{D.5}$$

as required. This concludes the proof. □

Bibliography

[htta] Air (https://math.stackexchange.com/users/181046/air). *Prove That Projection Operator Is Non Expansive*. Mathematics Stack Exchange. eprint: https://math.stackexchange.com/q/1426422. URL: https://math.stackexchange.com/q/1426422.

[httb] user754478 (https://math.stackexchange.com/users/754478/user754478). *Proof: A tangent space of the manifold of SPD matrices is the set of symmetric matrices*. Mathematics Stack Exchange. (version: 2020-02-26). eprint: https://math.stackexchange.com/q/3560825. URL: https://math.stackexchange.com/q/3560825.

[AMS08] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[Afh22] A. Afham. *Optimal Average Fidelity - A Repository on GitHub*. https://github.com/afhamash/optimal-average-fidelity. Accessed: 2025-01-28. 2022.

[Afh24] A. Afham. *Generalized Fidelity - A Repository on GitHub*. https://github.com/afhamash/GeneralizedFidelity. Accessed: 2024-10-07. 2024.

[AF24] A. Afham and Chris Ferrie. *Riemannian-geometric generalizations of quantum fidelities and Bures-Wasserstein distance*. 2024. arXiv: 2410.04937 [quant-ph]. URL: https://arxiv.org/abs/2410.04937.

[AKF22] A. Afham, Richard Kueng, and Chris Ferrie. *Quantum mean states are nicer than you think: fast algorithms to compute states maximizing average fidelity*. 2022. arXiv: 2206.08183 [quant-ph]. URL: https://arxiv.org/abs/2206.08183.

[Agr+18] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. "A rewriting system for convex optimization problems". In: *Journal of Control and Decision* 5.1 (2018). DOI: 10.1080/23307706.2017.1397554.

[AC11] Martial Agueh and Guillaume Carlier. "Barycenters in the Wasserstein space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924.

[ASŻ10] V Al Osipov, Hans-Juergen Sommers, and K Życzkowski. "Random Bures mixed states and the distribution of their purity". In: *Journal of Physics A: Mathematical and Theoretical* 43.5 (2010), p. 055302.

[AS66] Syed Mumtaz Ali and Samuel D. Silvey. "A General Class of Coefficients of Divergence of One Distribution from Another". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 28.1 (1966), pp. 131–142. URL: https://www.jstor.org/stable/2984279.

[Alt+21] Jason Altschuler, Sinho Chewi, Patrik R Gerber, and Austin Stromme. "Averaging on the Bures-Wasserstein manifold: dimension-free convergence of gradient descent". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 22132–22145.

[Alt+23] Jason M. Altschuler, Sinho Chewi, Patrik Gerber, and Austin J. Stromme. *Averaging on the Bures-Wasserstein manifold: dimension-free convergence of gradient descent*. 2023. arXiv: 2106.08502 [math.OC]. URL: https://arxiv.org/abs/2106.08502.

[Álv+16] Pedro C Álvarez-Esteban, E Del Barrio, JA Cuesta-Albertos, and C Matrán. "A fixed-point approach to barycenters in Wasserstein space". In: *Journal of Mathematical Analysis and Applications* 441.2 (2016), pp. 744–762.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. 118. Cambridge university press, 2010.

[AD18] Srinivasan Arunachalam and Ronald De Wolf. "Optimal quantum sample complexity of learning algorithms". In: *Journal of Machine Learning Research* 19.71 (2018), pp. 1–36.

[AD13] Koenraad MR Audenaert and Nilanjana Datta. "alpha-z-relative Renyi entropies". In: *arXiv preprint arXiv:1310.7178* (2013).

[AS08] Koenraad MR Audenaert and Stefan Scheel. "On random unitary channels". In: *New Journal of Physics* 10.2 (2008), p. 023011.

[BCD05] Dave Bacon, Andrew M Childs, and Wim van Dam. "Optimal measurements for the dihedral hidden subgroup problem". In: *arXiv preprint quant-ph/0501044* (2005).

[BK15] Joonwoo Bae and Leong-Chuan Kwek. "Quantum state discrimination and its applications". In: *Journal of Physics A: Mathematical and Theoretical* 48.8 (2015), p. 083001.

[BBS24] Ge Bai, Francesco Buscemi, and Valerio Scarani. "Quantum Bayes' Rule and Petz Transpose Map from the Minimal Change Principle". In: *arXiv preprint* (2024). arXiv:2410.00319. URL: https://arxiv.org/abs/2410.00319.

[BC09] Stephen M. Barnett and Sarah Croke. "Quantum state discrimination". In: *Adv. Opt. Photon.* 1.2 (Apr. 2009), pp. 238–278. DOI: 10.1364/AOP.1.000238. URL: https://opg.optica.org/aop/abstract.cfm?URI=aop-1-2-238.

[BK02] Howard Barnum and Emanuel Knill. "Reversing quantum dynamics with near-optimal quantum and classical fidelity". In: *Journal of Mathematical Physics* 43.5 (2002), pp. 2097–2106.

[BAG20] Afrad Basheer, A Afham, and Sandeep K Goyal. "Quantum k-nearest neighbors algorithm". In: *arXiv preprint arXiv:2003.09187* (2020).

[BDL16] Salman Beigi, Nilanjana Datta, and Felix Leditzky. "Decoding quantum information via the Petz recovery map". In: *Journal of Mathematical Physics* 57.8 (2016).

[BG14] Salman Beigi and Amin Gohari. "Quantum achievability proof via collision relative entropy". In: *IEEE Transactions on Information Theory* 60.12 (2014), pp. 7980–7986.

[Bel75] Viacheslav P Belavkin. "Optimal multiple quantum statistical hypothesis testing". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 1.1-4 (1975), pp. 315–345.

[BS82] Viacheslav P Belavkin and P Staszewski. "Algebraic generalization of relative entropy and entropy". In: *Annales de l'institut Henri Poincaré. Section A, Physique Théorique*. Vol. 37. 1982, pp. 51–58.

[Ben98] Ingemar Bengtsson. "Lecture Notes on Geometry of Quantum Mechanics". Unpublished. 1998. URL: https://3dhouse.se/ingemar/.

[BŻ17] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge university press, 2017.

[Bha09] Rajendra Bhatia. "Positive definite matrices". In: *Positive Definite Matrices*. Princeton university press, 2009. DOI: 10.1515/9781400827787.

[Bha13] Rajendra Bhatia. *Matrix analysis*. Vol. 169. Springer Science & Business Media, 2013.

[BGJ19] Rajendra Bhatia, Stephane Gaubert, and Tanvi Jain. "Matrix versions of the Hellinger distance". In: *Letters in Mathematical Physics* 109 (2019), pp. 1777–1804.

[BJL18] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. "Strong convexity of sandwiched entropies and related optimization problems". In: *Reviews in Mathematical Physics* 30.09 (2018), p. 1850014.

[BJL19] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. "On the Bures–Wasserstein distance between positive definite matrices". In: *Expositiones Mathematicae* 37.2 (2019), pp. 165–191.

[BR97] Rajendra Bhatia and Peter Rosenthal. "How and why to solve the operator equation AX- XB= Y". In: *Bulletin of the London Mathematical Society* 29.1 (1997), pp. 1–21.

[Bha43] Anil Bhattacharyya. "On a measure of divergence between two statistical populations defined by their probability distribution". In: *Bulletin of the Calcutta Mathematical Society* 35 (1943), pp. 99–110.

[Blu10a] Robin Blume-Kohout. "Hedged Maximum Likelihood Quantum State Estimation". In: *Phys. Rev. Lett.* 105 (20 Nov. 2010). DOI: 10.1103/PhysRevLett.105.200504.

[Blu10b] Robin Blume-Kohout. "Hedged maximum likelihood quantum state estimation". In: *Physical review letters* 105.20 (2010), p. 200504.

[Blu10c] Robin Blume-Kohout. "Optimal, reliable estimation of quantum states". In: *New Journal of Physics* 12.4 (Apr. 2010). DOI: 10.1088/1367-2630/12/4/043034.

[Bou23] Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. DOI: 10.1017/CBO9780511804441.

[BRT23] Shrigyan Brahmachari, Roberto Rubboli, and Marco Tomamichel. "A fixed-point algorithm for matrix projections with applications in quantum information". In: *arXiv preprint arXiv:2312.14615* (2023).

[Bru+09] Wojciech Bruzda, Valerio Cappellini, Hans-Jürgen Sommers, and Karol Życzkowski. "Random quantum operations". In: *Physics Letters A* 373.3 (2009), pp. 320–324.

[Bub15] Sébastien Bubeck. *Convex Optimization: Algorithms and Complexity*. 2015. arXiv: 1405.4980 [math.OC]. URL: https://arxiv.org/abs/1405.4980.

[Che23] Hao-Chung Cheng. "Simple and tighter derivation of achievability for classical communication over quantum channels". In: *PRX Quantum* 4.4 (2023), p. 040330.

[Che+20] Sinho Chewi, Tyler Maunu, Philippe Rigollet, and Austin J Stromme. "Gradient descent algorithms for Bures-Wasserstein barycenters". In: *Conference on Learning Theory*. PMLR. 2020, pp. 1276–1304.

[Chi23] Giulio Chiribella. "Extreme quantum states and processes, and extreme points of general spectrahedra in finite dimensional algebras". In: *arXiv preprint arXiv:2311.10929* (2023).

[Cho75] Man-Duen Choi. "Completely positive linear maps on complex matrices". In: *Linear algebra and its applications* 10.3 (1975), pp. 285–290.

[CM20] Dariusz Chruściński and Takashi Matsuoka. "Quantum Conditional Probability and Measurement Induced Disturbance of a Quantum Channel". In: *Reports on Mathematical Physics* 86.1 (2020), pp. 115–128. ISSN: 0034-4877. DOI: https://doi.org/10.1016/S0034-4877(20)30060-4. URL: https://www.sciencedirect.com/science/article/pii/S0034487720300604.

[CGW21] Bryan Coutts, Mark Girard, and John Watrous. "Certifying optimality for convex quantum channel optimization problems". In: *Quantum* 5 (May 2021), p. 448. ISSN: 2521-327X. DOI: 10.22331/q-2021-05-01-448. URL: http://dx.doi.org/10.22331/q-2021-05-01-448.

[CS20] Sam Cree and Jamie Sikora. "A fidelity measure for quantum states based on the matrix geometric mean". In: *arXiv preprint arXiv:2006.06918* (2020).

[CS22a] Sam Cree and Jonathan Sorce. "Approximate Petz recovery from the geometry of density operators". In: *Communications in Mathematical Physics* 392.3 (2022), pp. 907–919.

[CS22b] Sam Cree and Jonathan Sorce. "Geometric conditions for saturating the data processing inequality". In: *Journal of Physics A: Mathematical and Theoretical* 55.13 (2022), p. 135202.

[Csi67] Imre Csiszár. "Information-Type Measures of Difference of Probability Distributions and Indirect Observations". In: *Studia Scientiarum Mathematicarum Hungarica* 2 (1967), pp. 299–318.

[CŻ24] Jakub Czartowski and Karol Życzkowski. *Quantum Pushforward Designs*. 2024. arXiv: 2412.09672 [quant-ph]. URL: https://arxiv.org/abs/2412.09672.

[DP07] G. M. D'Ariano and P. Perinotti. "Optimal Data Processing for Quantum Measurements". In: *Phys. Rev. Lett.* 98 (2 Jan. 2007), p. 020403. DOI: 10.1103/PhysRevLett.98.020403.

[DP15] Nicola Dalla Pozza and Gianfranco Pierobon. "Optimality of square-root measurements in quantum state discrimination". In: *Physical Review A* 91.4 (2015), p. 042334.

[Dav+07] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. "Information-theoretic metric learning". In: *Proceedings of the 24th International Conference on Machine Learning*. ICML '07. Corvalis, Oregon, USA: Association for Computing Machinery, 2007, pp. 209–216. ISBN: 9781595937933. DOI: 10.1145/1273496.1273523. URL: https://doi.org/10.1145/1273496.1273523.

[DB16] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization". In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5. DOI: 10.48550/arxiv.1603.00943.

[DF92] Manfredo Perdigao Do Carmo and J Flaherty Francis. *Riemannian geometry*. Vol. 2. Springer, 1992.

[Dua+19] Bojia Duan, Jiabin Yuan, Juan Xu, and Dan Li. "Quantum algorithm and quantum circuit for a-optimal projection: Dimensionality reduction". In: *Physical Review A* 99.3 (2019), p. 032311.

[Dup+14] Frédéric Dupuis, Lea Kraemer, Philippe Faist, Joseph M Renes, and Renato Renner. "Generalized entropies". In: *XVIIth international congress on mathematical physics*. World Scientific. 2014, pp. 134–153.

[EF01] Yonina C Eldar and G David Forney. "On quantum detection and the square-root measurement". In: *IEEE Transactions on Information Theory* 47.3 (2001), pp. 858–872.

[EMV03] Yonina C Eldar, Alexandre Megretski, and George C Verghese. "Designing optimal quantum detectors via semidefinite programming". In: *IEEE Transactions on Information Theory* 49.4 (2003), pp. 1007–1012.

[EMV04] Yonina C Eldar, Alexandre Megretski, and George C Verghese. "Optimal detection of symmetric mixed quantum states". In: *IEEE Transactions on Information Theory* 50.6 (2004), pp. 1198–1207.

[FF21] Kun Fang and Hamza Fawzi. "Geometric Rényi divergence and its applications in quantum channel capacities". In: *Communications in Mathematical Physics* 384.3 (2021), pp. 1615–1677.

[FR15] Omar Fawzi and Renato Renner. "Quantum conditional mutual information and approximate Markov chains". In: *Communications in Mathematical Physics* 340.2 (2015), pp. 575–611.

[FB16] Christopher Ferrie and Robin Blume-Kohout. *Bayes estimator for multinomial parameters and Bhattacharyya distances*. 2016. arXiv: 1612.07946 [math.ST]. URL: https://arxiv.org/abs/1612.07946.

[Fey82] Richard P. Feynman. "Simulating Physics with Computers". In: *International Journal of Theoretical Physics* 21.6/7 (1982), pp. 467–488. DOI: 10.1007/BF02650179. URL: https://link.springer.com/article/10.1007/BF02650179.

[Fla+12] Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. "Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators". In: *New Journal of Physics* 14.9 (Sept. 2012), p. 095022. DOI: 10.1088/1367-2630/14/9/095022.

[Fuk+01] Mituhiro Fukuda, Masakazu Kojima, Kazuo Murota, and Kazuhide Nakata. "Exploiting sparsity in semidefinite programming via matrix completion I: General framework". In: *SIAM Journal on optimization* 11.3 (2001), pp. 647–674.

[Gil+22] András Gilyén, Seth Lloyd, Iman Marvian, Yihui Quek, and Mark M Wilde. "Quantum algorithm for Petz recovery channels and pretty good measurements". In: *Physical Review Letters* 128.22 (2022), p. 220502.

[Gin65] Jean Ginibre. "Statistical ensembles of complex, quaternion, and real matrices". In: *Journal of Mathematical Physics* 6.3 (1965), pp. 440–449.

[GCC16] Christopher Granade, Joshua Combes, and D G Cory. "Practical Bayesian tomography". In: *New Journal of Physics* 18.3 (Mar. 2016), p. 033024. DOI: 10.1088/1367-2630/18/3/033024.

[Gra+17] Christopher Granade, Christopher Ferrie, Ian Hincks, Steven Casagrande, Thomas Alexander, Jonathan Gross, Michal Kononenko, and Yuval Sanders. "QInfer: Statistical inference software for quantum applications". en-GB. In: *Quantum* 1 (Apr. 2017). Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften, p. 5. DOI: 10.22331/q-2017-04-25-5.

[Gro+10] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. "Quantum State Tomography via Compressed Sensing". In: *Phys. Rev. Lett.* 105 (15 Oct. 2010), p. 150401. DOI: 10.1103/PhysRevLett.105.150401.

[Guţ+20] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. "Fast state tomography with optimal error bounds". In: *Journal of Physics A: Mathematical and Theoretical* 53.20 (2020), p. 204001.

[Han+21] Andi Han, Bamdev Mishra, Pratik Kumar Jawanpuria, and Junbin Gao. "On Riemannian optimization over positive definite matrices with the Bures-Wasserstein geometry". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 8940–8953.

[HSH17] Mehrtash Harandi, Mathieu Salzmann, and Richard Hartley. "Joint dimensionality reduction and metric learning: A geometric take". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1404–1413.

[Hau+96] Paul Hausladen, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William K. Wootters. "Classical information capacity of a quantum channel". In: *Phys. Rev. A* 54 (3 Sept. 1996), pp. 1869–1876. DOI: 10.1103/PhysRevA.54.1869. URL: https://link.aps.org/doi/10.1103/PhysRevA.54.1869.

[HW94] Paul Hausladen and William K Wootters. "A 'pretty good'measurement for distinguishing quantum states". In: *Journal of Modern Optics* 41.12 (1994), pp. 2385–2390.

[HKK06] Masahito Hayashi, Akinori Kawachi, and Hirotada Kobayashi. "Quantum measurements for hidden subgroup problems with optimal sample complexity". In: *arXiv preprint quant-ph/0604174* (2006).

[Hay+04] Patrick Hayden, Richard Jozsa, Denes Petz, and Andreas Winter. "Structure of states which satisfy strong subadditivity of quantum entropy with equality". In: *Communications in mathematical physics* 246 (2004), pp. 359–374.

[HSF24] Kerry He, James Saunderson, and Hamza Fawzi. "A Bregman proximal perspective on classical and quantum Blahut-Arimoto algorithms". In: *IEEE Transactions on Information Theory* (2024).

[HJN20] Teiko Heinosaari, Maria Anastasia Jivulescu, and Ion Nechita. "Random positive operator valued measures". In: *Journal of Mathematical Physics* 61.4 (2020).

[Hel09] E. Hellinger. "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen". In: *Journal für die reine und angewandte Mathematik* 136 (1909), pp. 210–271. DOI: 10.1515/crll.1909.136.210.

[HK69] K-E Hellwig and Karl Kraus. "Pure operations and measurements". In: *Communications in Mathematical Physics* 11.3 (1969), pp. 214–220.

[Hel69] Carl W Helstrom. "Quantum detection and estimation theory". In: *Journal of Statistical Physics* 1 (1969), pp. 231–252.

[Hol78] Alexander Semenovich Holevo. "On asymptotically optimal hypotheses testing in quantum statistics". In: *Teoriya Veroyatnostei i ee Primeneniya* 23.2 (1978), pp. 429–432.

[HJ85] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985. DOI: 10.1017/CBO9780511810817.

[Hra97] Z. Hradil. "Quantum-state estimation". In: *Phys. Rev. A* 55 (3 Mar. 1997), R1561–R1564. DOI: 10.1103/PhysRevA.55.R1561.

[Hsu+24] Ming-Chien Hsu, En-Jui Kuo, Wei-Hsuan Yu, Jian-Feng Cai, and Min-Hsiu Hsieh. "Quantum state tomography via nonconvex riemannian gradient descent". In: *Physical Review Letters* 132.24 (2024), p. 240804.

[HH12] F. Huszár and N. M. T. Houlsby. "Adaptive Bayesian quantum tomography". In: *Phys. Rev. A* 85 (5 May 2012), p. 052120. DOI: 10.1103/PhysRevA.85.052120.

[IRS17] Raban Iten, Joseph M. Renes, and David Sutter. "Pretty Good Measures in Quantum Information Theory". In: *IEEE Transactions on Information Theory* 63.2 (2017), pp. 1270–1279. DOI: 10.1109/TIT.2016.2639521. URL: https://arxiv.org/abs/1608.08229.

[Jam72] Andrzej Jamiołkowski. "Linear transformations which preserve trace and positive semidefiniteness of operators". In: *Reports on mathematical physics* 3.4 (1972), pp. 275–278.

[Jen03] Anna Jencová. "Geodesic distances on density matrices". In: *arXiv preprint math-ph/0312044* (2003).

[Joz94] Richard Jozsa. "Fidelity for mixed quantum states". In: *Journal of modern optics* 41.12 (1994), pp. 2315–2323.

[Jun+18] Marius Junge, Renato Renner, David Sutter, Mark M Wilde, and Andreas Winter. "Universal recovery maps and approximate sufficiency of quantum relative entropy". In: *Annales Henri Poincaré*. Vol. 19. 10. Springer. 2018, pp. 2955–2978.

[Kan42] Leonid V Kantorovich. "On the translocation of masses". In: *Dokl. Akad. Nauk. USSR (NS)*. Vol. 37. 1942, pp. 199–201.

[KW21] Vishal Katariya and Mark M Wilde. "Geometric distinguishability measures limit quantum channel estimation and discrimination". In: *Quantum Information Processing* 20.2 (2021), p. 78.

[Kho72] AS Kholevo. "On quasiequivalence of locally normal states". In: *Theoretical and Mathematical Physics* 13.2 (1972), pp. 1071–1082.

[Kli+16] Martin Kliesch, Richard Kueng, Jens Eisert, and David Gross. "Improving Compressed Sensing With the Diamond Norm". In: *IEEE Transactions on Information Theory* 62.12 (Dec. 2016), pp. 7445–7463. DOI: 10.1109/tit.2016.2606500.

[KRS09] Robert Konig, Renato Renner, and Christian Schaffner. "The operational meaning of min-and max-entropy". In: *IEEE Transactions on Information theory* 55.9 (2009), pp. 4337–4347.

[Kra71] Karl Kraus. "General state changes in quantum theory". In: *Annals of Physics* 64.2 (1971), pp. 311–335.

[KSS21] Alexey Kroshnin, Vladimir Spokoiny, and Alexandra Suvorikova. "Statistical inference for Bures–Wasserstein barycenters". In: *The Annals of Applied Probability* 31.3 (2021), pp. 1264–1298.

[KA80] Fumio Kubo and Tsuyoshi Ando. "Means of positive linear operators". In: *Mathematische Annalen* 246 (1980), pp. 205–224.

[Kuk+21] Ryszard Kukulski, Ion Nechita, Łukasz Pawela, Zbigniew Puchała, and Karol Życzkowski. "Generating random quantum channels". In: *Journal of Mathematical Physics* 62.6 (2021).

[Kul+13] Brian Kulis et al. "Metric learning: A survey". In: *Foundations and Trends® in Machine Learning* 5.4 (2013), pp. 287–364.

[KL51] Solomon Kullback and Richard A Leibler. "On information and sufficiency". In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.

[LL08] Jimmie Lawson and Yongdo Lim. "Γ-Commuting triples of positive definite matrices and midpoint equations on unitary matrices". In: *Linear Algebra and its Applications* 428.8-9 (2008), pp. 2030–2039.

[LL24] Jimmie D Lawson and Yongdo Lim. "The expanding universe of the geometric mean". In: *Acta Scientiarum Mathematicarum* (2024), pp. 1–21.

[Led22] Felix Leditzky. "Optimality of the pretty good measurement for port-based teleportation". In: *Letters in Mathematical Physics* 112.5 (2022), p. 98.

[LRD17] Felix Leditzky, Cambyse Rouzé, and Nilanjana Datta. "Data processing for the sandwiched Rényi divergence: a condition for equality". In: *Letters in Mathematical Physics* 107 (2017), pp. 61–80.

[Lee12] John M Lee. *Smooth manifolds*. Springer, 2012.

[Lee18] John M Lee. *Introduction to Riemannian manifolds*. Vol. 2. Springer, 2018.

[LS13] M. S. Leifer and Robert W. Spekkens. "Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference". In: *Physical Review A* 88.5 (Nov. 2013). ISSN: 1094-1622. DOI: 10.1103/physreva.88.052130. URL: http://dx.doi.org/10.1103/PhysRevA.88.052130.

[Lei06] Matthew S Leifer. "Quantum dynamics as an analog of conditional probability". In: *Physical Review A—Atomic, Molecular, and Optical Physics* 74.4 (2006), p. 042310.

[LPW09] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. Providence, RI: American Mathematical Society, 2009. ISBN: 978-0-8218-4739-8. URL: https://www.ams.org/publications/authors/books/postpub/mbk-78.

[LC19] Yen-Huan Li and Volkan Cevher. "Convergence of the exponentiated gradient method with Armijo line search". In: *Journal of Optimization Theory and Applications* 181.2 (2019), pp. 588–607.

[Lia+20] Jin-Min Liang, Shu-Qian Shen, Ming Li, and Lei Li. "Variational quantum algorithms for dimensionality reduction and classification". In: *Phys. Rev. A* 101 (3 Mar. 2020), p. 032323. DOI: 10.1103/PhysRevA.101.032323. URL: https://link.aps.org/doi/10.1103/PhysRevA.101.032323.

[LCL21] Chien-Ming Lin, Hao-Chung Cheng, and Yen-Huan Li. "Maximum-Likelihood Quantum State Tomography by Cover's Method with Non-Asymptotic Analysis". In: *arXiv preprint arXiv:2110.00747* (2021).

[Lin75] Göran Lindblad. "Completely positive maps and entropy inequalities". In: *Communications in Mathematical Physics* 40 (1975), pp. 147–151.

[LMW13] Noah Linden, Milán Mosonyi, and Andreas Winter. "The structure of Rényi entropic inequalities". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469.2158 (2013), p. 20120737.

[Liu+24] Nana Liu, Qisheng Wang, Mark M Wilde, and Zhicheng Zhang. "Quantum algorithms for matrix geometric means". In: *arXiv preprint arXiv:2405.00673* (2024).

[LZ04] Shunlong Luo and Qiang Zhang. "Informational distance on quantum-state space". In: *Physical Review A—Atomic, Molecular, and Optical Physics* 69.3 (2004), p. 032106.

[MMP18] Luigi Malago, Luigi Montrucchio, and Giovanni Pistone. "Wasserstein Riemannian geometry of positive definite matrices". In: *arXiv preprint arXiv:1801.09269* (2018).

[Mat10] Keiji Matsumoto. "Reverse test and quantum analogue of classical fidelity and generalized fidelity". In: *arXiv preprint arXiv:1006.0302* (2010).

[Mat14] Keiji Matsumoto. "Quantum fidelities, their duals, and convex analysis". In: *arXiv preprint arXiv:1408.3462* (2014).

[Mat15] Keiji Matsumoto. "A new quantum version of f-divergence". In: *Nagoya Winter Workshop: Reality and Measurement in Algebraic Quantum Theory*. Springer. 2015, pp. 229–273.

[Mod16] Klas Modin. “Geometry of matrix decompositions seen through optimal transport and information geometry”. In: *arXiv preprint arXiv:1601.01875* (2016).

[MRL08] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. “Quantum-process tomography: Resource analysis of different strategies”. In: *Phys. Rev. A* 77 (3 Mar. 2008), p. 032322. DOI: 10.1103/PhysRevA.77.032322. URL: https://link.aps.org/doi/10.1103/PhysRevA.77.032322.

[Mon81] Gaspard Monge. “Mémoire sur la théorie des déblais et des remblais”. In: *Mem. Math. Phys. Acad. Royale Sci.* (1781), pp. 666–704.

[MR05] Cristopher Moore and Alexander Russell. “For distinguishing conjugate hidden subgroups, the pretty good measurement is as good as it gets”. In: *arXiv preprint quant-ph/0501177* (2005).

[Mor63] Toshio Morimoto. “Markov Processes and the H-Theorem”. In: *Journal of the Physical Society of Japan* 18.3 (1963), pp. 328–331. DOI: 10.1143/JPSJ.18.328.

[Mül+13] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. “On quantum Rényi entropies: A new generalization and some properties”. In: *Journal of Mathematical Physics* 54.12 (2013).

[Nak19] Barış Nakiboğlu. “The Augustin capacity and center”. In: *Problems of Information Transmission* 55 (2019), pp. 299–342.

[NR18] Rajai Nasser and Joseph M Renes. “Polar codes for arbitrary classical-quantum channels and arbitrary cq-macs”. In: *IEEE Transactions on Information Theory* 64.11 (2018), pp. 7424–7442.

[NP24] Ion Nechita and Sang-Jun Park. “Random covariant quantum channels”. In: *arXiv preprint arXiv:2403.03667* (2024).

[NY83] Arkadij Semenovič Nemirovskij and David Borisovich Yudin. “Problem complexity and method efficiency in optimization”. In: (1983).

[NM10] Hui Khoon Ng and Prabha Mandayam. “Simple approach to approximate quantum error correction based on the transpose channel”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 81.6 (2010), p. 062342.

[NC01] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Vol. 2. Cambridge university press Cambridge, 2001.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Dec. 2010. ISBN: 978-1-139-49548-6.

[Nur+24] Theshani Nuradha, Hemant K. Mishra, Felix Leditzky, and Mark M. Wilde. *Multivariate Fidelities*. 2024. arXiv: 2404.16101 [quant-ph]. URL: https://arxiv.org/abs/2404.16101.

[Oza00] Masanao Ozawa. "Entanglement measures and the Hilbert–Schmidt distance". In: *Physics Letters A* 268.3 (2000), pp. 158–160.

[PZ19] Victor M Panaretos and Yoav Zemel. "Statistical aspects of Wasserstein distances". In: *Annual review of statistics and its application* 6.1 (2019), pp. 405–431.

[PB23] Arthur J. Parzygnat and Francesco Buscemi. "Axioms for retrodiction: achieving time-reversal symmetry with a prior". In: *Quantum* 7 (May 2023), p. 1013. ISSN: 2521-327X. DOI: 10.22331/q-2023-05-23-1013. URL: http://dx.doi.org/10.22331/q-2023-05-23-1013.

[Pet09] Leif E Peterson. "K-nearest neighbor". In: *Scholarpedia* 4.2 (2009), p. 1883.

[Pet86a] Dénes Petz. "Quasi-entropies for finite quantum systems". In: *Reports on mathematical physics* 23.1 (1986), pp. 57–65.

[Pet86b] Dénes Petz. "Sufficient subalgebras and the relative entropy of states of a von Neumann algebra". In: *Communications in mathematical physics* 105 (1986), pp. 123–131.

[Pet88] Dénes Petz. "Sufficiency of channels over von Neumann algebras". In: *The Quarterly Journal of Mathematics* 39.1 (1988), pp. 97–108.

[RG95] Motakuri Ramana and Alan J Goldman. "Some geometric results in semidefinite programming". In: *Journal of Global Optimization* 7.1 (1995), pp. 33–50.

[RHJ01] J. Rehácek, Z. Hradil, and M. Ježek. "Iterative algorithm for reconstruction of entangled states". In: *Phys. Rev. A* 63 (4 Mar. 2001), p. 040303. DOI: 10.1103/PhysRevA.63.040303.

[Řeh+07] Jaroslav Řeháček, Zdeněk Hradil, E. Knill, and A. I. Lvovsky. "Diluted maximum-likelihood algorithm for quantum tomography". In: *Phys. Rev. A* 75 (4 Apr. 2007), p. 042108. DOI: 10.1103/PhysRevA.75.042108. URL: https://link.aps.org/doi/10.1103/PhysRevA.75.042108.

[Rén61] Alfred Rényi. "On Measures of Entropy and Information". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Ed. by Jerzy Neyman. 20 June-30 July 1960. Berkeley, California: University of California Press, 1961, pp. 547–561.

[Ret+23] Soorya Rethinasamy, Rochisha Agarwal, Kunal Sharma, and Mark M Wilde. "Estimating distinguishability measures on quantum computers". In: *Physical Review A* 108.1 (2023), p. 012409.

[Roc93] R Tyrrell Rockafellar. "Lagrange multipliers and optimality". In: *SIAM review* 35.2 (1993), pp. 183–238.

[Roc70] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970. ISBN: 978-0691015866.

[SSP14] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. "Quantum computing for pattern classification". In: *PRICAI 2014: Trends in Artificial Intelligence: 13th Pacific Rim International Conference on Artificial Intelligence, Gold Coast, QLD, Australia, December 1-5, 2014. Proceedings 13*. Springer. 2014, pp. 208–220.

[Sho94] Peter W Shor. "Algorithms for quantum computation: discrete logarithms and factoring". In: *Proceedings 35th annual symposium on foundations of computer science*. Ieee. 1994, pp. 124–134.

[SGS12] John A. Smolin, Jay M. Gambetta, and Graeme Smith. "Efficient Method for Computing the Maximum-Likelihood Quantum State from Measurements with Additive Gaussian Noise". In: *Phys. Rev. Lett.* 108 (7 Feb. 2012), p. 070502. DOI: 10.1103/PhysRevLett.108.070502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.070502.

[Sti55] W Forrest Stinespring. "Positive functions on C*-algebras". In: *Proceedings of the American Mathematical Society* 6.2 (1955), pp. 211–216.

[Sur+22] Trystan Surawy-Stepney, Jonas Kahn, Richard Kueng, and Madalin Guta. "Projected least-squares quantum process tomography". In: *Quantum* 6 (2022), p. 844.

[STH16] David Sutter, Marco Tomamichel, and Aram W Harrow. "Strengthened monotonicity of relative entropy via pinched Petz recovery map". In: *IEEE Transactions on Information Theory* 62.5 (2016), pp. 2907–2913.

[Syl84] James Joseph Sylvester. “Sur l’équation en matrices px= xq”. In: *CR Acad. Sci. Paris* 99.2 (1884), pp. 67–71.

[Tom15] Marco Tomamichel. *Quantum information processing with finite resources: mathematical foundations*. Vol. 5. Springer, 2015.

[Tys09a] Jon Tyson. “Error rates of Belavkin weighted quantum measurements and a converse to Holevo’s asymptotic optimality theorem”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 79.3 (2009), p. 032343.

[Tys09b] Jon Tyson. “Two-sided estimates of minimum-error distinguishability of mixed quantum states via generalized Holevo–Curlander bounds”. In: *Journal of mathematical physics* 50.3 (2009).

[Uhl76] Armin Uhlmann. “The “transition probability” in the state space of a star-algebra”. In: *Reports on Mathematical Physics* 9.2 (1976), pp. 273–279.

[Uhl77] Armin Uhlmann. “Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory”. In: *Communications in Mathematical Physics* 54 (1977), pp. 21–32.

[Uhl10] Armin Uhlmann. “Transition Probability (Fidelity) and Its Relatives”. In: *Foundations of Physics* 41.3 (Jan. 2010), pp. 288–298. ISSN: 1572-9516. DOI: 10.1007/s10701-009-9381-y. URL: http://dx.doi.org/10.1007/s10701-009-9381-y.

[Ume62] Hisaharu Umegaki. “Conditional expectation in an operator algebra, IV (entropy and information)”. In: *Kodai Mathematical Seminar Reports*. Vol. 14. 2. Department of Mathematics, Tokyo Institute of Technology. 1962, pp. 59–85.

[VH14] Tim Van Erven and Peter Harremos. “Rényi divergence and Kullback-Leibler divergence”. In: *IEEE Transactions on Information Theory* 60.7 (2014), pp. 3797–3820.

[VA+15] Lieven Vandenberghe, Martin S Andersen, et al. “Chordal graphs and semidefinite optimization”. In: *Foundations and Trends® in Optimization* 1.4 (2015), pp. 241–433.

[Vil+09] Cédric Villani et al. *Optimal transport: old and new*. Vol. 338. Springer, 2009.

[Vin14] Cynthia Vinzant. “What is... a spectrahedron”. In: *Notices Amer. Math. Soc* 61.5 (2014), pp. 492–494.

[Vis18] Nisheeth K Vishnoi. "Geodesic convex optimization: Differentiation on manifolds, geodesics, and convexity". In: *arXiv preprint arXiv:1806.06373* (2018).

[VNM24] Daniel Volya, Andrey Nikitin, and Prabhat Mishra. "Fast quantum process tomography via Riemannian gradient descent". In: *arXiv preprint arXiv:2404.18840* (2024).

[WS15] Fei Wang and Jimeng Sun. "Survey on distance metric learning and dimensionality reduction in data mining". In: *Data mining and knowledge discovery* 29.2 (2015), pp. 534–564.

[Wan+24] Guan-Ren Wang, Chung-En Tsai, Hao-Chung Cheng, and Yen-Huan Li. "Computing augustin information via hybrid geodesically convex optimization". In: *2024 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2024, pp. 2532–2537.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge: Cambridge University Press, 2018. ISBN: 978-1-107-18056-7. DOI: 10.1017/9781316848142. (Visited on 11/10/2024).

[WKS15] Nathan Wiebe, Ashish Kapoor, and Krysta M Svore. "Quantum nearest-neighbor algorithms for machine learning". In: *Quantum information and computation* 15.3-4 (2015), pp. 318–358.

[Wil15] Mark M Wilde. "Recoverability in quantum information theory". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471.2182 (2015), p. 20150338.

[WWY14] Mark M Wilde, Andreas Winter, and Dong Yang. "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy". In: *Communications in Mathematical Physics* 331 (2014), pp. 593–622.

[Wil13] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139525343.

[Wil18] Mark M. Wilde. "Recoverability for Holevo's Just-as-Good Fidelity". In: *2018 IEEE International Symposium on Information Theory (ISIT)*. 2018, pp. 2331–2335. DOI: 10.1109/ISIT.2018.8437346. URL: https://arxiv.org/abs/1801.02800.

[Wis28] John Wishart. "The generalised product moment distribution in samples from a normal multivariate population". In: *Biometrika* (1928), pp. 32–52.

[YJ06] Liu Yang and Rong Jin. "Distance metric learning: A comprehensive survey". In: *Michigan State Universiy* 2.2 (2006), p. 4.

[YCL22] Jun-Kai You, Hao-Chung Cheng, and Yen-Huan Li. "Minimizing quantum Rényi divergences via mirror descent with Polyak step size". In: *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2022, pp. 252–257.

[YFT19] Akram Youssry, Christopher Ferrie, and Marco Tomamichel. "Efficient online quantum state estimation using a matrix-exponentiated gradient method". In: *New Journal of Physics* 21.3 (2019), p. 033006.

[YKL75] Horace Yuen, Robert Kennedy, and Melvin Lax. "Optimum testing of multiple hypotheses in quantum detection theory". In: *IEEE transactions on information theory* 21.2 (1975), pp. 125–134.

[ZHS16] Pourya Zadeh, Reshad Hosseini, and Suvrit Sra. "Geometric mean metric learning". In: *International conference on machine learning*. PMLR. 2016, pp. 2464–2471.

[ZP19] Yoav Zemel and Victor M. Panaretos. "Fréchet means and Procrustes analysis in Wasserstein space". In: *Bernoulli* 25.2 (2019), pp. 932–976. DOI: 10.3150/17-BEJ1009. URL: https://doi.org/10.3150/17-BEJ1009.

[Zhe+20] Yang Zheng, Giovanni Fantuzzi, Antonis Papachristodoulou, Paul Goulart, and Andrew Wynn. "Chordal decomposition in operator-splitting methods for sparse semidefinite programs". In: *Mathematical Programming* 180.1 (2020), pp. 489–532.

[Zho+25] Juntai Zhou, Stefano Chessa, Eric Chitambar, and Felix Leditzky. *On the distinguishability of geometrically uniform quantum states*. 2025. arXiv: 2501.12376 [quant-ph]. URL: https://arxiv.org/abs/2501.12376.

[ZS01] Karol Zyczkowski and Hans-Jürgen Sommers. "Induced measures in the space of mixed quantum states". In: *Journal of Physics A: Mathematical and General* 34.35 (2001), p. 7111.