UNIVERSITY OF CALGARY

Convex Analysis in Quantum Information

by

Mark William Girard

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

August, 2017

Abstract

Convexity arises naturally in the study of quantum information. As a result, many useful tools from convex analysis can be used to give important results regarding aspects of quantum information. This thesis builds up methods using core concepts from convex analysis, including convex optimization problems, convex roof constructions, and conic programming, to study mathematical problems related to quantum entanglement. In this thesis, I develop a method for solving convex optimization problems that arise in quantum information theory by analyzing the corresponding converse problem. That is, given an element in a convex set, I determine a family of convex functions that are minimized at this point. This method is used find explicit formulas for the relative entropy of entanglement, as well as other important quantities used to quantify entanglement, and allows one to show important relationships between them. I also construct a practical algorithm that can be used to compute these quantities. This thesis also presents a method to compute convex roofs of arbitrary entanglement measures evaluated on highly symmetric bipartite states. I also establish a framework for completely characterizing quantum resource theories that are convex. For resource theories with a simple mathematical structure, this gives rise to a complete set of resource monotones that can be computed in practice using semidefinite programs. This has applications to the study of entanglement transformations.

Acknowledgements

First and foremost, I must express my sincere gratitude to my advisor, Dr. Gilad Gour, without whose direction this thesis would have not been possible. I hope that at least some of his unending patience and calmness, in addition to his wisdom and knowledge, have rubbed off on me in my time working with him.

Thanks must go to my coauthors, Yuriy Zinchenko and Shmuel Friedland, whose work was the foundation for much of Chapter 4 of this thesis.

I am incredibly thankful for the numerous scientific discussions (as well as non-academic distractions) from the other students and postdocs in my group during my time in Calgary, including Mehdi Ahmadi, Hoan Dang, Ish Dhand, Varun Narasimhachar, Priyaa Srinivasan and Alireza Poostindouz.

I must acknowledge my various funding sources that supported me throughout my studies, including scholarships from Alberta Innovates–Technology Futures and the Killam Trust.

Above all, I am eternally grateful for my wife, Hannah, whose love and support helped me get through the tough times and helped me enjoy the good times.

To Hannah, the love of my life.

Table of Contents

List of Figures and Illustrations

List of Symbols, Abbreviations, and Nomenclature

Symbol or abbreviation	Definition
$\mathrm{L}(\mathcal{H})$	Space of linear operators on a hilbert space $\mathcal{H}$
$\mathrm{H}(\mathcal{H})$	Space of hermitian operators on $\mathcal{H}$
$\mathrm{H}(\mathcal{H})_+$, $\mathrm{H}(\mathcal{H})_{++}$	Cones of positive and positive definite operators on $\mathcal{H}$
$\mathrm{D}(\mathcal{H})$	Set of density operators on $\mathcal{H}$
$\mathrm{Sep}(\mathcal{H}_\mathsf{A} : \mathcal{H}_\mathsf{B})$	Cone of separable operators on $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$
$\mathrm{SepD}(\mathcal{H}_\mathsf{A} : \mathcal{H}_\mathsf{B})$	Set of separable density operators on $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$
$\mathrm{PPTD}(\mathcal{H}_\mathsf{A} : \mathcal{H}_\mathsf{B})$	Set of PPT density operators on $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$
$\mathbb{1}_\mathsf{A}$	Identity operator on $\mathcal{H}_\mathsf{A}$
id_A	Identity map on $\mathrm{L}(\mathcal{H}_\mathsf{A})$
$f'(X;Y)$	Directional derivative
$f^{[1]}(\Lambda)$	Matrix of divided differences
$\mathfrak{D}_{f,\sigma}$	Fréchet derivative of f at σ
$\mathrm{cl}(\mathcal{X})$	Closure of $\mathcal{X}$
$\mathrm{int}(\mathcal{X})$	Interior of $\mathcal{X}$
$\mathrm{conv}(\mathcal{X})$	Convex hull of $\mathcal{X}$
$\mathrm{cone}(\mathcal{X})$	Conic hull of $\mathcal{X}$
$\mathrm{relint}(\mathcal{X})$	Relative interior of $\mathcal{X}$
$\mathrm{bd}(\mathcal{X})$	Relative boundary of $\mathcal{X}$
H_d	Space of $d \times d$ hermitian matrices
$\mathrm{H}_{d,+}$	Cone of positive semidefinite $d \times d$ hermitian matrices
$\mathrm{H}_{d,++}$	Positive definite $d \times d$ hermitian matrices
$\mathrm{H}_{d,+,1}$	Set of $d \times d$ density matrices
$B_p(x,1)$	Unit ball centered at x in the p-norm
$\mathrm{orb}_\mathcal{G}(\rho)$	The orbit of the action of a group $\mathcal{G}$ on ρ
$\mathcal{T}_\mathcal{G}$	The $\mathcal{G}$-twirling operator for group $\mathcal{G}$
$\mathcal{X}^*$	Dual cone to $\mathcal{X}$
$\mathcal{E}^*$, Φ^*	Dual linear map

Chapter 1

Introduction

Quantum information theory is the study of how information can be stored, communicated, and manipulated using the laws of quantum mechanics. While quantum information differs from classical information in many ways, the two key components that lead to its most interesting and useful properties are superpositions and entanglement. A superposition of quantum states is, mathematically, just a linear combination of vectors and is thus very well-understood. On the other hand, entanglement is a property of positive operators on the tensors product of vector spaces and is much more difficult to analyze.

Although entanglement is one of the defining features of quantum physics, a complete understanding of this phenomenon remains elusive. Entanglement was originally defined in terms of correlations: a quantum system is said to be entangled if the results of measurements performed on the separate subsystems exhibit correlations that are stronger than what could possibly occur classically [SB35]. These correlations can exist even when the systems are far apart from each other. Making use of these long-distance correlations constitutes the heart of entanglement theory applications and understanding how entanglement can be manipulated is one of the main problems in quantum information theory. The true value of entanglement was first recognized in its power to help perform efficient simulations of quantum systems [Fey82]. This ushered in an era of new developments in quantum computation with the prospect of constructing quantum computers that will one day be able to solve problems much more quickly than is possible with the best-known classical algorithms [RB01]. It was quickly recognized that many-body entanglement is a necessary resource for quantum computers to have computational speedups over their classical counterparts [Vid03]. Furthermore, entanglement has been discovered to be useful in quantum information processing schemes involving the teleportation of quantum states [BBC$^+$93], as well as secure quantum communication and cryptography [BB84].

Entanglement is a necessary ingredient for many quantum information processing protocols, but not all entanglement is equivalent [ES14, FGR11, GMN$^+$15a]. Only certain entangled states are known to be useful for particular applications. Furthermore, typical experimental realizations of quantum information processing are constrained in how the information of the entangled particles can be manipulated [CLM$^+$14]. It is therefore necessary for us to understand the mathematical structure of entanglement imposed by this restriction and to determine when it is possible to transform one entangled resource into another within this hierarchy [VPRK97, Hor01, NV01].

Convexity naturally arises in many places in quantum information theory. The sets of all possible states, processes and measurements for quantum systems are all convex sets. Furthermore, many important quantities in quantum information are defined in terms of convex optimization problems. In particular, entanglement is an important resource and quantifying entanglement is a problem that is often cast in terms of convex optimization problems. Entanglement, however, is not the only property of quantum systems that can be studied as a resource. Many more quantum "resource theories" can be defined in terms of convex sets of free states and free operations. As a result of this convexity, numerous tools from convex analysis can be used to study the important questions that arise in these resource theories. The primary tools from convex analysis that are utilized in this thesis are those used to study convex optimization problems, convex roofs of functions, and conic programming.

Using techniques from convex analysis, for example, we can characterize the conditions for when a particular density matrix is the optimal one for computing the minimum of given convex function. This allows us to find closed-form solutions to many convex optimization problems that would be otherwise unobtainable analytically. In particular, closed formula for the relative entropy of entanglement as well as the Rains bound (an upper bound to distillable entanglement) can be found using this method, and these quantities can be compared. Making use of certain types of symmetry in bipartite quantum states, we can also use methods from convex analysis to compute so-called *convex roofs* of entanglement measures on many entangled states. This analysis can also be used to determine necessary and sufficient conditions that determine when transformation between certain symmetric entangled states is possible. Lastly, tools from conic programming can be used to study transformability of resources in arbitrary resource theories in quantum information. Such results can be used to study approximations to entanglement theory. These results listed above are the primary ones that are presented in this thesis.

The main organization for the remainder of this thesis is as follows. Chapters 2 and 3 present the necessary background for quantum information theory and convex analysis. The main new contributions of this thesis are contained in Chapters 4, 5, and 6. These chapters are mostly independent of one another and can be read separately. A brief chapter-by-chapter breakdown of the thesis is given below.

Chapter 2: Quantum information theory and quantum entanglement This chapter introduces the mathematical basics necessary for dealing with quantum information: state vectors, density operators, superoperators, and so on. A thorough overview of quantum entanglement is also given. Finally, the definitions of a generalized *resource theory* in quantum information are given and notions regarding convertibility of resources within a resources within a resource theory are discussed.

Chapter 3: Convex analysis This chapter discusses the main concepts from convex analysis that will be used throughout the thesis. The main definitions needed to analyze convex sets and convex functions are outlined. The Fréchet derivative is introduced which is used to write down necessary and sufficient conditions for a point in a convex set to minimize a convex function. Important definitions and facts for operator convex functions (i.e., functions of matrices) will be introduced, and the separating hyperplane theorem will be mentioned. Lastly, the study of cones and conic programming will be reviewed.

Chapter 4: Convex optimization problems in quantum information theory This chapter consists largely of the work originally presented in [GGF14] and [GZFG15], but many results are expanded upon and additional examples are provided. Results about finding necessary and sufficient conditions for optimization of certain convex optimization problems that arise in quantum information theory are presented. The case of trace-type functionals is studied first in general manner, which gives rise to the specific results regarding optimization criteria for studying the quantum relative entropies. Closed formulae for the relative entropy of entanglement of certain states is given by solving the converse problem. That is, given a state σ on the boundary of the separable states and a supporting hyperplane of the separable states at that state, it is possible to find all entangled states ρ whose closest separable state (with respect to the relative entropy) is σ. Similar results are used to study the Rains bound and the Rényi relative entropies of entanglement. Additionaly, an algorithm is presented that is used to numerically estimate these quantities. This algorithm uses the *cutting plane* technique and implements numerical convex programming tools.

Chapter 5: Entanglement of bipartite symmetric states This chapter presents the work of [GG17]. The *convex roof* construction is discussed and techniques for computing convex roofs of under symmetry are introduced. Certain classes of symmetric bipartite entangled states are analyzed. Most importantly, methods for computing the convex roof of arbitrary entanglement measures for states with these types of symmetries are provided. We also give necessary and sufficient conditions for conversion of a pure state to a Werner state by LOCC.

Chapter 6: Conic resource theories This chapter presents new work that has not yet been published. The main concepts of conic resource theories are first introduced. Necessary and sufficient conditions for conversion among resources in different convex quantum resource theories are given. These conditions are can be written purely in terms of the dual cone of witnesses to the convex set of free operations. A complete family of computable resource monotones are also be defined for these resource theories. Examples of convex resource theories are presented, and applications involving approximations to the resource theory of entanglement are discussed. In particular, a new class of PPT-monotones that can be computed as semidefinite programs are introduced and analyzed.

Chapter 2

Quantum information theory and entanglement

This chapter is devoted to introducing the basic mathematical tools and framework that will be used throughout this work. We first recall some basic notions from linear algebra and use this to develop the necessary tools for analyzing quantum information theory for finite-dimensional quantum systems. Then we will consider the notion of entanglement in bipartite quantum systems and discuss some of its properties.

2.1 Linear algebra

2.1.1 Vectors and matrices

The n-dimensional complex Euclidean space will be denoted $\mathcal{H} = \mathbb{C}^n$. The 'bra-ket' notation from quantum mechanics will be used as follows. Elements of $\mathbb{C}^n$ are column vectors and will denoted by *kets* $|v\rangle \in \mathbb{C}^n$. The *standard basis* of $\mathbb{C}^n$ is typically denoted by the collection of vectors $\{|1\rangle, |2\rangle, \ldots, |n\rangle\}$ defined by

$$|1\rangle = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \quad |2\rangle = \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \quad \ldots, \quad |n\rangle = \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}.$$

Arbitrary elements can be expanded in this basis as $|v\rangle = \alpha_1|1\rangle + \cdots + \alpha_n|n\rangle$ for some complex numbers α_i. The *bras* $\langle v| := |v\rangle^\dagger$ represent the dual (row) vectors, where $(\cdot)^\dagger$ is the conjugate transpose. The inner product of two vectors $|v\rangle, |w\rangle \in \mathbb{C}^n$ is denoted by $\langle v|w\rangle$ which is conjugate-linear in the first argument and linear in the second.

The space of linear maps from $\mathcal{H}_{\mathsf{A}}$ to $\mathcal{H}_{\mathsf{B}}$ will be denoted by $\mathrm{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$. For $\mathcal{H}_{\mathsf{A}} = \mathbb{C}^m$ and $\mathcal{H}_{\mathsf{B}} = \mathbb{C}^n$, this space of linear operators is is identified with the space $\mathrm{L}(\mathbb{C}^m, \mathbb{C}^n) \simeq \mathrm{M}_{n,m}$ of $n \times m$ matrices with complex entries. We use the shorthand $\mathrm{M}_n := \mathrm{M}_{n,n}$ to denote the space of $n \times n$ matrices. Both the operator notation $\mathrm{L}(\mathcal{H}_{\mathsf{A}})$ and the matrix notation M_n will be used depending on convenience. The identity operator will be denoted $\mathbb{1}_n \in \mathrm{M}_n$ or as $\mathbb{1}_{\mathsf{A}} \in \mathrm{L}(\mathcal{H}_{\mathsf{A}})$ depending on whether the size of the matrix or the name of the space that it operates on is being emphasized.

The space of operators $\mathrm{M}_n = \mathrm{L}(\mathbb{C}^n)$ is itself a Euclidean space with inner product defined by $\langle X, Y\rangle = \mathrm{Tr}(X^\dagger Y)$, called the *Hilbert-Schmidt* inner product, for $X, Y \in \mathrm{L}(\mathcal{H})$. The space M_n is n^2-dimensional and is spanned by $\{|j\rangle\langle k|\}$ for $1 \leq i, j \leq n$. Given an operator $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$, its adjoint $X^\dagger$ is the unique linear operator such that $\langle w|X|v\rangle = (X^\dagger|w\rangle)^\dagger|v\rangle$ for all $|v\rangle \in \mathcal{H}_{\mathsf{A}}$ and $|w\rangle \in \mathcal{H}_{\mathsf{B}}$. In matrix representation, $X^\dagger$ is the conjugate transpose. An operator $X \in \mathrm{L}(\mathbb{C}^n)$ is called *hermitian* if $X^\dagger = X$. An hermitian operator X is called *positive semidefinite* (or just *positive* for short) if $\langle v|X|v\rangle \geq 0$ for all $|v\rangle \in \mathbb{C}^n$ and is called *positive definite* if that inequality is strict. Positive semi-definiteness and positive definiteness of a matrix X is denoted by $X \geq 0$ and $X > 0$ respectively. Given hermitian matrices X and Y, we will write $X \geq Y$ to mean $X - Y \geq 0$. The set of hermitian $n \times n$ matrices is denoted H_n and the set of positive semidefinite matrices is denoted $\mathrm{H}_{n,+}$. When the dimension of the underlying space does not need to be emphasized, we will also use the notation $\mathrm{H}(\mathcal{H}_\mathsf{A})$ and $\mathrm{H}(\mathcal{H}_\mathsf{A})_+$ to denote the space of hermitian operators and set of positive semidefinite operators acting on $\mathcal{H}_\mathsf{A}$.

The space $\mathrm{H}(\mathbb{C}^n) = \mathrm{H}_n$ of hermitian operators is itself an n^2-dimensional *real* Euclidean space with the same inner product. This space is spanned by the n^2 hermitian matrices $\{H_{j,k} \,|\, 1 \leq j, k \leq n\}$ defined by

$$H_{j,k} = \begin{cases} |j\rangle\langle j| & \text{if } j = k \\ \frac{|j\rangle\langle k| + |k\rangle\langle j|}{\sqrt{2}} & \text{if } j < k \\ \frac{|j\rangle\langle k| - i|k\rangle\langle j|}{\sqrt{2}} & \text{if } k < j, \end{cases}$$

which are orthonormal with respect to the inner product defined above. We will now state some useful characterizations of matrices using the Hilbert-Schmidt inner product. Let $X \in \mathrm{L}(\mathcal{H})$. Then $X \geq 0$ if and only if $\langle X, P\rangle$ for all positive operators $P \geq 0$. Similarly, it holds that $X = 0$ if and only if $\langle X, Y\rangle = 0$ for all $Y \in \mathrm{H}(\mathcal{H})$. Lastly, for any vector $|v\rangle \in \mathcal{H}$ it holds that $\Big\langle |v\rangle\langle v|, X \Big\rangle = \langle v|X|v\rangle$.

Every hermitian operator $X \in \mathrm{H}(\mathbb{C}^n)$ can be diagonalized into its spectral decomposition $X = \sum_{j=1}^n \lambda_j |u_j\rangle\langle u_j|$ where λ_j are the real eigenvalues of X and $|v_j\rangle$ are the normalized eigenvectors. A real-valued function $f : \Omega \to \mathbb{R}$ from a subset $\Omega \subseteq \mathbb{R}$ can be extended to functions of hermitian matrices whose eigenvalues are contained in Ω by $f(X) = \sum_j f(\lambda_j)|u_j\rangle\langle u_j|$. The modulus of a matrix X is defined as $|X| = \sqrt{X^\dagger X}$, which is allowed since $X^\dagger X \geq 0$ for any matrix X. Every hermitian matrix X has a unique *Jordan-Hahn decomposition* given by $X = X_+ - X_-$, where $X_+ \geq 0$ and $X_- \geq 0$ are the unique positive semidefinite matrices satisfying $X = X_+ - X_-$ and $X_+ X_- = 0$. These matrices can be given by

$$X_+ = \frac{|X| + X}{2} \qquad \text{and} \qquad X_- = \frac{|X| - X}{2}.$$

2.1.2 The tensor product

The tensor product to two complex Euclidean spaces $\mathcal{H}_{\mathsf{A}} = \mathbb{C}^m$ and $\mathcal{H}_{\mathsf{B}} = \mathbb{C}^n$, written as $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$, is isomorphic to the mn-dimensional Euclidean space $\mathbb{C}^{mn}$. Given bases $\{|v_1\rangle, \ldots, |v_m\rangle\}$ and $\{|w_1\rangle, \ldots, |w_n\rangle\}$ of $\mathbb{C}^m$ and $\mathbb{C}^n$, arbitrary elements of $\mathbb{C}^m \otimes \mathbb{C}^n$ can be written as $|v\rangle = \sum_{i=1}^m \sum_{j=1}^n \alpha_{ij} |v_i\rangle \otimes |w_j\rangle$. The tensor products of the standard basis elements are denoted by $|ij\rangle = |i\rangle \otimes |j\rangle$ for short. The tensor product of matrix spaces has the isomorphism $\mathrm{L}(\mathcal{H}_{\mathsf{A}}) \otimes \mathrm{L}(\mathcal{H}_{\mathsf{B}}) \simeq \mathrm{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$, and an arbitrary matrix X in this space can be expanded in the standard bases as

$$X = \sum_{i,j=1}^{m} \sum_{j,k=1}^{n} x_{ijkl} |i\rangle\langle j| \otimes |k\rangle\langle l|.$$

Given a matrix $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$, its *partial trace* $\mathrm{Tr}_{\mathsf{B}}(X)$ obtained from tracing over system B is given by

$$\begin{aligned} \mathrm{Tr}_{\mathsf{B}}(X) = &= \sum_{i,j,k,l} x_{ijkl} |i\rangle\langle j| \, \mathrm{Tr}(|k\rangle\langle l|) \\ &= \sum_{i,j,k} x_{ijkk} |i\rangle\langle j|. \end{aligned}$$

Note that the partial trace is a linear map $\mathrm{Tr}_{\mathsf{B}} \colon \mathrm{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A}})$. The partial trace Tr_{A} is defined analogously.

When writing elements of and operators on $\mathbb{C}^m$ and $\mathbb{C}^n$ as vectors and matrices, the tensor product is given by the Kronecker product. In particular, the tensor product of the space of $m \times m$ matrices with the space of $n \times n$ matrices is isomorphic to the space of $mn \times mn$ matrices $\mathrm{M}_m \otimes \mathrm{M}_n \simeq \mathrm{M}_{mn}$. Given $X \in \mathrm{M}_m$ and $Y \in \mathrm{M}_n$ their tensor product is written as

$$
\begin{aligned}
X \otimes Y = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mm} \end{pmatrix} \otimes \begin{pmatrix} x_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nn} \end{pmatrix} &= \begin{pmatrix} x_{11}Y & \cdots & x_{1m}Y \\ \vdots & \ddots & \vdots \\ x_{m1}Y & \cdots & x_{mm}Y \end{pmatrix} \\
&= \begin{pmatrix} x_{11}y_{11} & \cdots & x_{1m}y_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1}y_{n1} & \cdots & x_{mm}y_{nn} \end{pmatrix}.
\end{aligned}
$$

We will now state a useful fact about vectors in tensor product spaces. This fact makes use of the *singular value decomposition*. That is, for any matrix $X \in \mathrm{M}_{m,n}$ there exist unitary matrices $U \in \mathrm{M}_m$ and $V \in \mathrm{M}_n$ and a pseudo-diagonal matrix $D \in \mathrm{M}_{m,n}$ such that $X = UDV^\dagger$, where the diagonal entries of D are nonnegative. The diagonal entries of D are unique and are called the *singular values* of X.

Theorem 2.1 (Schmidt decomposition)**.** *Let $|x\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ be a (not necessarily normalized) vector and let $d = \min\{\dim(\mathcal{H}_\mathsf{A}), \dim(\mathcal{H}_\mathsf{B})\}$. Then there exist positive numbers $\lambda_1 \geq \ldots \lambda_d \geq 0$ and orthonormal sets of vectors $\{|u_1\rangle, \ldots, |u_d\rangle\}$ and $\{|v_1\rangle, \ldots, |v_d\rangle\}$ such that*

$$|x\rangle = \sum_{j=1}^{d} \sqrt{\lambda_j} |u_j\rangle \otimes |v_j\rangle.$$

Furthermore, if $|x\rangle = \sum_{j,k} x_{jk} |j\rangle \otimes |k\rangle$ then the λ_j are the eigenvalues of the positive matrix $XX^\dagger$ where X is the matrix whose entries are x_{jk}.

Proof. We can assume without loss of generality that $d = \dim(\mathcal{H}_\mathsf{A}) \leq \dim(\mathcal{H}_\mathsf{B})$. Consider the singular value decomposition $X = UDV^\dagger$ and note that the $d \times d$ matrix $XX^\dagger = UDD^\dagger U^\dagger$ has entries given by

$$x_{jk} = \sum_{l} \sqrt{\lambda_l} u_{jl} \overline{v_{kl}}$$

where U_{jl} and v_{kl} are the entries of U and V respectively. Then

$$\begin{aligned}|x\rangle &= \sum_{j,k,l} \sqrt{\lambda_l} u_{jl} \overline{v_{kl}} |j\rangle \otimes |k\rangle \\ &= \sum_{l} \sqrt{\lambda_l} \Big(\sum_{j} u_{jl} |j\rangle \Big) \otimes \Big(\sum_{k} \overline{v_{kl}} |k\rangle \Big) \\ &= \sum_{l=1}^{d} \sqrt{\lambda_j} |u_l\rangle \otimes |v_l\rangle,\end{aligned}$$

where $|u_l\rangle = \sum_j u_{jl}|j\rangle$ and $|v_l\rangle = \sum_k \overline{v_{kl}}|k\rangle$ are the l^{th} columns of U and $\overline{V}$ respectively. Since U and V are unitary matrices, the sets of vectors $\{|u_1\rangle, \ldots, |u_d\rangle\}$ and $\{|v_1\rangle, \ldots, |v_d\rangle\}$ are orthonormal. □

In the future, we will often assume without loss of generality that a normalized vector $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ is in *Schmidt form* $|\psi\rangle = \sum_{j=1}^{d} \sqrt{\lambda_j} |jj\rangle$ with Schmidt coefficients in decreasing order $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$, and satisfying $\sum_{j=1}^{d} \lambda_j = 1$ since $|\psi\rangle$ is normalized. The vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ of Schmidt coefficients of a normalized vector $|\psi\rangle$ is unique up to ordering. We will later see that Schmidt coefficients are useful for characterizing entanglement of bipartite pure states.

2.1.3 Superoperators

Linear maps of operators $\Phi\colon \mathrm{L}(\mathcal{H}_{\mathsf{A}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{B}})$ are called *superoperators*. The identity map on $\mathrm{L}(\mathcal{H}_{\mathsf{A}})$ will be denoted id_{A}. The *dual* of a superoperator is the unique linear operator $\Phi^*\colon \mathrm{L}(\mathcal{H}_{\mathsf{B}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A}})$ such that

$$\langle Y, \Phi(X) \rangle = \langle \Phi^*(Y), X \rangle$$

for all matrices $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A}})$ and all $Y \in \mathrm{L}(\mathcal{H}_{\mathsf{B}})$.

There are numerous useful definitions regarding superoperators that will be used in this thesis that are stated below. Note that $\langle X, \mathbb{1}_{\mathsf{A}} \rangle = \mathrm{Tr}(X)$ for any matrix $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A}})$.

Definition 2.2. Let $\Phi\colon \mathrm{L}(\mathcal{H}_{\mathsf{A}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{B}})$ be a superoperator. Then Φ is said to be *trace preserving* if $\mathrm{Tr}(\Phi(X)) = \mathrm{Tr}(X)$ for all $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A}})$.

Given two superoperators $\Phi\colon \mathrm{L}(\mathcal{H}_{\mathsf{A}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{B}})$ and $\Psi\colon \mathrm{L}(\mathcal{H}_{\mathsf{B}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{B}'})$ we can define their product superoperator $\Phi\otimes\Psi\colon \mathrm{L}(\mathcal{H}_{\mathsf{A}}\otimes\mathcal{H}_{\mathsf{B}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{B}}\otimes\mathcal{H}_{\mathsf{B}'})$ that acts on tensor products of operators by

$$\Phi \otimes \Psi(X \otimes Y) = \Phi(X) \otimes \Psi(Y)$$

and extends linearly to all operators in $\mathrm{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$. The identity superoperator on $\mathrm{L}(\mathbb{C}^m)$ will be denoted $\mathrm{id}_m : \mathrm{L}(\mathbb{C}^m) \to \mathrm{L}(\mathbb{C}^m)$. We now define the notions of *positivity* and *complete positivity* of superoperators which will be important for applications in quantum information.

Definition 2.3. Let $\Phi\colon \mathrm{L}(\mathcal{H}_{\mathsf{A}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{B}})$ be a superoperator. Then Φ is said to be

1. *positivity preserving* (or just *positive* for short) if $\mathrm{Tr}(\Phi(X)) \geq 0$ for all $X \geq 0$;
2. *k-positive* if the map $\Phi \otimes \mathrm{id}_k$ is positive;
3. *completely positive* if Φ is k-positive for all integers $k \geq 1$.

Example 2.4. An example of a linear map that is positive but not completely positive is the transpose map. Let $\mathcal{H} = \mathbb{C}^d$ and consider the transpose map $T\colon \mathrm{M}_d \to \mathrm{M}_d$ given by $T(X) = X^T$, where $\mathrm{L}(\mathbb{C}^d) = \mathrm{M}_d$ is the space of $d \times d$ matrices. The transpose map is clearly positive, since the transpose of a matrix has the same eigenvalues as the original matrix, so positivity is indeed preserved under the transpose. However, the transpose is not completely positive. Define the (unnormalized) vector $|\phi_d^+\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$

$$|\phi_d^+\rangle = \sum_{j=1}^{d} |j\rangle \otimes |j\rangle.$$

Applying the extended map $\mathrm{id}_d \otimes T$ to $|\phi_d^+\rangle\langle\phi_d^+|$, we find

$$\begin{aligned} T \otimes \mathrm{id}_d(|\phi^+\rangle\langle\phi^+|) &= \sum_{j,k=1}^{d} |k\rangle\langle j| \otimes |j\rangle\langle k| \\ &= \sum_{j,k=1}^{d} |kj\rangle\langle jk| = W_d, \end{aligned}$$

where W_d is the *swap operator* on $\mathbb{C}^d \otimes \mathbb{C}^d$ having the property that $W_d(|u\rangle \otimes |v\rangle) = |v\rangle \otimes |u\rangle$ for all vectors $|u\rangle, |v\rangle \in \mathbb{C}^d$. We will make use of the swap operator later in this thesis, and we will simply write W if the dimension d is implicit from context. The swap operator is both unitary and Hermitian, having eigenvalues 1 and -1 and satisfies $W^2 = \mathbb{1}$. Using the Hahn decomposition of W, the projection matrices onto the positive and negative eigenspaces of W will be denoted W_+ and W_- respectively. The positive eigenspace of W is spanned by the $\binom{d+1}{2}$ vectors in

$$\{|jj\rangle \mid 1 \le j \le d\} \cup \left\{|\psi_{jk}^+\rangle \,\middle|\, 1 \le j < k \le d\right\} \tag{2.1}$$

and the negative eigenspace of W is spanned by the $\binom{d}{2}$ vectors in

$$\left\{|\psi_{jk}^-\rangle \,\middle|\, 1 \le j < k \le d\right\} \tag{2.2}$$

where $|\psi_{jk}^+\rangle$ and $|\psi_{jk}^-\rangle$ are the unit vectors in $\mathbb{C}^d$ defined by

$$|\psi_{jk}^+\rangle = \frac{|jk\rangle + |kj\rangle}{\sqrt{2}} \qquad \text{and} \qquad |\psi_{jk}^-\rangle = \frac{|jk\rangle - |kj\rangle}{\sqrt{2}} \tag{2.3}$$

for $1 \le j < k \le d$. In particular we see that $\mathrm{id}_d \otimes T(|\phi^+\rangle\langle\phi^+|)$ is the swap operator, which is not a positive matrix since it has negative eigenvalues even though $|\phi^+\rangle\langle\phi^+|$ is positive. Hence T is not completely positive. For $d = 2$, the matrix form of the swap operator is explicitly given by

$$T \otimes \mathrm{id}_2 \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

An important result regarding complete positivity of maps on finite-dimensional spaces is that it suffices to check only that the map is m-positive, where m is the dimension of the input space.

Theorem 2.5 ([Cho75])**.** *Let $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ be finite-dimensional Hilbert spaces with $d = \dim(\mathcal{H}_\mathsf{A})$, and let $\Phi\colon \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ be a superoperator. It holds that Φ is completely positive if and only if Φ is d-positive.*

In this thesis, two well known representations of superoperators will be used. These are the *Choi representation* and the *Kraus representation*. Here, we first introduce these representations then state how these can be used to characterize quantum channels.

Choi representation

Let $\mathcal{H}_\mathsf{A} = \mathbb{C}^d$ and let $\Phi\colon \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ be a linear map, and let $\{|1\rangle, \dots |d\rangle\}$ be an orthonormal basis for $\mathcal{H}_\mathsf{A}$. The *Choi matrix* of Φ is the operator $J(\Phi) \in \mathrm{L}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B})$ defined by

$$\begin{aligned} J(\Phi) &= \Phi \otimes \mathrm{id}_\mathsf{A} \left(\sum_{j,k=1}^{d} |j\rangle\langle k| \otimes |j\rangle\langle k| \right) \\ &= \sum_{j,k=1}^{d} \Phi(|j\rangle\langle k|) \otimes |j\rangle\langle k|. \end{aligned} \tag{2.4}$$

When $\mathcal{H}_\mathsf{A} = \mathbb{C}^d$,the Choi matrix can be written as $J(\Phi) = \Phi \otimes \mathrm{id}_d(|\phi^+\rangle\langle\phi^+|)$. Note that J is an isomorphism from the space of linear maps from $\mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ to the tensor product space of operators $\mathrm{L}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B})$. Indeed, the map Φ can be uniquely recovered from its Choi matrix by

$$\Phi(X) = \mathrm{Tr}_\mathsf{A}\Big(J(\Phi)(\mathbb{1}_\mathsf{B} \otimes X^T) \Big) \tag{2.5}$$

for any $X \in \mathrm{L}(\mathcal{H}_\mathsf{A})$, and J is known as the *Choi-Jamiolkowski* isomorphism. Furthermore, for any $Y \in \mathrm{L}(\mathcal{H}_\mathsf{B})$, it holds that

$$\langle \Phi(X), Y \rangle = \langle J(\Phi), Y \otimes X^T \rangle.$$

Kraus representation

Let $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ be finite dimensional Hilbert spaces, and let $\{K_1, \ldots, K_n\} \subseteq \mathrm{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ and $\{L_1, \ldots, L_n\} \subseteq \mathrm{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ be collections of operators. These operators define a superoperator $\Phi : \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ by

$$\Phi(X) = \sum_{j=1}^{n} K_j X L_j^\dagger,$$

and this representation is a *Kraus* representation of Φ. In fact, every superoperator $\Phi : \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ has such a Kraus representation with at most $n = mm'$ operators, where $m = \dim(\mathcal{H}_\mathsf{A})$ and $m' = \dim(\mathcal{H}_\mathsf{B})$, although we will now prove this here.

Given an operator $X \in \mathrm{L}(\mathcal{H}, \mathcal{H}')$, the *adjoint mapping* is the linear map $\mathrm{Ad}_X : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H}')$ defined by

$$\mathrm{Ad}_X(Y) = XYX^\dagger. \tag{2.6}$$

Clearly, Ad_X is completely positive for and X. With this notation, the Kraus representation of a channel can be written as $\Phi(X) = \sum_j \mathrm{Ad}_{K_j}(X)$.

Characterizations of quantum channels

We now state the conditions for a superoperator Φ to be completely positive and trace preserving in terms of the Choi and Kraus representations.

Theorem 2.6 ([Cho75]). *Let $\mathcal{H}_\mathsf{A} = \mathbb{C}^m$ and $\mathcal{H}_\mathsf{B} = \mathbb{C}^m$, and let $\Phi \colon \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathcal{H}_\mathsf{B}$ be a linear map. The following are equivalent.*

1. *Φ is completely positive.*

2. It holds that $J(\Phi) \geq 0$.

3. There exist operators $\{K_1, \ldots, K_{mn}\} \in \mathrm{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ *such that* $\Phi(X) = \sum_{l=1}^{mn} K_l X K_l^\dagger$.

A superoperator $\Phi : \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ is said to be *hermiticity preserving* if $\Phi(X)$ is hermitian whenever $X \in \mathrm{L}(\mathcal{H}_\mathsf{A})$ is hermitian. It is clear that if Φ is positive then it is hermiticity preserving. Importantly, the Choi matrix $J(\Phi)$ is hermitian if and only if the superoperator Φ is hermiticity preserving. We now state the conditions for a hermiticity preserving map to be trace preserving.

Theorem 2.7. *Let* $\mathcal{H}_\mathsf{A} = \mathbb{C}^m$ *and* $\mathcal{H}_\mathsf{B} = \mathbb{C}^n$, *and let* $\Phi : \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ *be a hermiticity preserving linear map. The following are equivalent.*

1. Φ *is trace preserving.*

2. The Choi matrix $J(\Phi)$ *of* Φ *satisfies* $\mathrm{Tr}_\mathsf{B}(J(\Phi)) = \mathbb{1}_\mathsf{A}$.

3. There exist operators $\{K_1, \ldots, K_d\} \subseteq \mathrm{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ *and* $\{L_1, \ldots, L_d\} \subseteq \mathrm{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ *such that* $\Phi(X) = \sum_{j=1}^{d} K_j X L_j^\dagger$ *and* $\sum_j L_j^\dagger K_j = \mathbb{1}_\mathsf{A}$.

Definition 2.8. A superoperator $\Phi \colon \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ is said to be a *quantum channel* if it is completely positive and trace preserving. The set of quantum channels from $\mathcal{H}_\mathsf{A}$ to $\mathcal{H}_\mathsf{B}$ will be denoted $\mathrm{C}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$.

Quantum channels represent physical operations that act on quantum systems. This will be discussed further in the next section. Putting together Theorems 2.6 and 2.7 yields the following characterization of quantum channels.

Theorem 2.9. *Let* $\mathcal{H}_\mathsf{A} = \mathbb{C}^m$ *and* $\mathcal{H}_\mathsf{B} = \mathbb{C}^n$, *and let* $\Phi \colon \mathrm{L}(\mathcal{H}_\mathsf{A}) \to \mathrm{L}(\mathcal{H}_\mathsf{B})$ *be a linear map. The following are equivalent.*

1. Φ *is a quantum channel.*

2. The Choi matrix $J(\Phi)$ *is satisfies* $J(\Phi) \geq 0$ *and* $\mathrm{Tr}_\mathsf{A}(J(\Phi)) = \mathbb{1}_\mathsf{B}$.

3. There exist $\{K_j\} \subseteq \mathrm{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ *such that* $\Phi(X) = \sum_j K_j X K_j^\dagger$ *and* $\sum_j K_j^\dagger K_j = \mathbb{1}_\mathsf{A}$.

2.2 Quantum states, measurements, and channels

In order to mathematically represent quantum systems, a finite dimensional Hilbert space $\mathcal{H}$ is associated to each (finite dimensional) quantum system.

Definition 2.10. Let $\mathcal{H}$ be a finite dimensional Hilbert space. A *quantum state* (or a *density operator*) is a positive operator $\rho \in \mathrm{L}(\mathcal{H})$ satisfying $\mathrm{Tr}(\rho) = 1$. The set of density operators on a system $\mathcal{H}$ will be denoted $\mathrm{D}(\mathcal{H})$.

The set of density operators on a system is convex and closed. The *pure states* of the system are denoted by unit vectors $|\psi\rangle \in \mathcal{H}$, or equivalently by their density operator $|\psi\rangle\langle\psi|$. The density operators of pure states are rank one and are exactly the extreme points of the set of density operators. In particular, the set of states is the convex hull of the pure states

$$\mathrm{D}(\mathcal{H}) = \mathrm{conv}\{|\psi\rangle\langle\psi| \,|\, |\psi\rangle \in \mathcal{H},\, \langle\psi|\psi\rangle = 1\},$$

and any density operator can be decomposed (non-uniquely) into a convex combination of projectors onto pure states

$$\rho = \sum_i p_i |\psi\rangle\langle\psi|$$

such that $p_i \geq 0$ and $\sum_i p_i = 1$.

Definition 2.11. An *ensemble* is a (countable) collection of pairs $\{(p_i, \rho_i)\}$ where the $p_i \geq 0$ are probabilities satisfying $\sum_i p_i = 1$ and the ρ_i are quantum states of a system $\mathcal{H}$. An ensemble $\{(p_i, \rho_i)\}$ is said to be a *decomposition* of another state ρ if $\rho = \sum_i p_i \rho_i$.

As we will see later, the concept of decompositions of a quantum state is important for analyzing entanglement of mixed states.

2.2.1 Distance measures in quantum information

There are several measures that are used in quantum information for the difference between two density operators ρ and σ on a system $\mathcal{H}$. As requirements of such measures, we impose the following properties for a function $f(\rho, \sigma)$ of two operators to be a useful distance measure.

Let $\mathcal{H}$ be a finite dimensional Hilbert spaces and $f : \mathrm{H}(\mathcal{H})_+ \times \mathrm{H}(\mathcal{H})_+ \to \mathbb{R}$ a function of positive operators onto nonnegative real numbers.

- f is *unitary invariant* if $f(U\rho U^\dagger, U\sigma U^\dagger) = f(\rho, \sigma)$ for all unitary operators U.
- f is said to be *jointly convex* if, for all $\rho_1, \rho_2, \sigma_1, \sigma_2 \in \mathrm{H}(\mathcal{H})_+$ and all $t \in (0, 1)$ it holds that
$$f(t\rho_1 + (1-t)\rho_2), t\sigma_1 + (1-t)\sigma_2) \leq tf(\rho_1, \sigma_1) + (1-t)f(\rho_2, \sigma_2). \tag{2.7}$$
- f is said to be *additive* if $f(\rho_1 \otimes \rho_2, \sigma \otimes \sigma_2) = f(\rho_1, \sigma_1) + f(\rho_2, \sigma_2)$ for all $\rho_1, \rho_2, \sigma_1, \sigma_2$.
- f is *monotonic* (or said to satisfy the *data processing inequality*) if
$$f(\Phi(\rho), \Phi(\sigma)) \leq f(\rho, \sigma) \tag{2.8}$$
holds for all ρ, σ and all quantum channels Φ.

Additionally, we should require of any good distance measure that $f(\rho, \sigma) = 0$ if and only if $\rho = \sigma$. Functions that satisfy the above properties are useful for constructing resource measures in quantum resource theories, which will be discussed later. We now present a few of the distance measures that will be employed in this thesis.

The most natural distance measure is given by the trace-norm distance $\|\rho - \sigma\|_1$ given in terms of the *trace norm* $\|X\|_1 = \sqrt{\mathrm{Tr}(X^\dagger X)} = \mathrm{Tr}|X|$. Given a matrix $X \in \mathrm{L}(\mathcal{H})$, the *support* of X will be denoted $\mathrm{supp}(X) = \{|v\rangle \mid X|v\rangle \neq 0\}$. Given $\rho, \sigma \in \mathrm{H}(\mathcal{H})$, the *quantum* *relative entropy* of ρ and σ is defined as

$$S(\rho\|\sigma) = \mathrm{Tr}(\rho \log \rho - \rho \log \sigma) \tag{2.9}$$

if $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ and $S(\rho\|\sigma) = \infty$ otherwise. The quantum relative entropy is additive, monotonic under quantum channels, and is jointly convex in its arguments. It also satisfies $S(\rho\|\sigma) \geq 0$ for all ρ and σ with equality if and only if $\rho = \sigma$.

For real numbers $\alpha > 0$ (with $\alpha \neq 1$), it is also possible to define the *relative Rényi entropy of order* α by

$$S_\alpha(\rho\|\sigma) = \frac{1}{1-\alpha} \log \mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha}). \tag{2.10}$$

If ρ is a density operator, the standard relative entropy is recovered from the limit $S(\rho\|\sigma) = \lim_{\alpha\to 1} S_\alpha(\rho\|\sigma)$. The relative Rényi entropies are additive and strictly positive for all $\alpha > 0$, but are only monotonic under quantum channels for $\alpha \in [0, 2]$ and are jointly convex only when $\alpha \in [0, 1]$ [MH11].

2.2.2 Quantum measurements and instruments

A *measurement* on a system $\mathcal{H}$ is a collection of operators $\{K_1, \dots, K_n\} \subseteq \mathrm{L}(\mathcal{H}, \mathcal{H}')$ such that $\sum_{j=1}^n K_j^\dagger K_j = \mathbb{1}_{\mathcal{H}}$. This has the following physical interpretation. The indices $j \in \{1, \dots, n\}$ indicate the possible outcomes of the measurement. If the system is initially prepared in the state $\rho \in \mathrm{D}(\mathcal{H})$, outcome j will be observed with probability $p_j = \mathrm{Tr}(K_j^\dagger K_j \rho)$. If outcome j is observed, the resulting system will be in the state $\rho_j = \frac{1}{p_j} K_j \rho K_j^\dagger$. After performing the measurement, the system can be seen to be in the ensemble $\{(p_j, \rho_j)\}$.

This notion of a measurement is closely related to the Kraus representation of a quantum channel. Indeed, any measurement $\{K_j\}$ can be made into a channel $\mathcal{E}(\rho) = \sum_j K_j \rho K_j^\dagger$ which is equivalent to performing the measurement and subsequently "forgetting" the outcome. Equivalently, a Kraus representation of a channel can be interpreted as a measurement.

More generally, a *quantum instrument* $\mathfrak{I}$ from $\mathcal{H}$ to $\mathcal{H}'$ is a finite collection of superoperators $\mathfrak{J} = \{\mathcal{E}_j\}_{j\in\mathcal{I}}$ for some finite set $\mathcal{I}$, where each superoperator $\mathcal{E}_j : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H}')$ is completely positive (but not necessarily trace-preserving), such that their sum $\mathcal{E} = \sum_{j\in\mathcal{I}} \mathcal{E}_j$ is completely positive. With this definition, a measurement $\{K_j\}$ is just a quantum instrument in which each of the superoperators is an adjoint map $\mathcal{E}_j = \mathrm{Ad}_{K_j}$ defined by $\mathrm{Ad}_{K_j}(X) = K_j X K_j^\dagger$. Similarly, a quantum channel $\mathcal{E}$ is just a quantum instrument with one superoperator $\mathfrak{J} = \{\mathcal{E}\}$. Finally, any quantum instrument $\mathfrak{J} = \{\mathcal{E}_j\}_{j\in\mathcal{I}}$ can be made into a quantum channel $\mathcal{E}_{\mathfrak{J}} \in \mathrm{C}(\mathcal{H}, \mathcal{H}')$ defined by $\mathcal{E}_{\mathfrak{J}}(X) = \sum_{j\in\mathcal{I}} \mathcal{E}_j(X)$.

Quantum instruments have a similar interpretation to quantum measurements. If a system $\mathcal{H}$ in state ρ is measured using the instrument $\mathfrak{J} = \{\mathcal{E}_j\}_{j\in\mathcal{I}}$, outcome j will be observed with probability $p_j = \mathrm{Tr}(\mathcal{E}_j(\rho))$. If outcome j is observed, the resulting system will be in the state $\rho_j = \frac{1}{p_j}\mathcal{E}_j(\rho)$. After performing a measurement corresponding to an instrument $\mathfrak{J} = \{\mathcal{E}_j\}$, the system can be considered to be in the ensemble $\{(p_j, \rho_j)\}$.

A quantum instrument $\mathfrak{J}'$ from $\mathcal{H}$ to $\mathcal{H}''$ is said to be *conditioned* on another instrument $\mathfrak{J} = \{\mathcal{E}_j\}_{j\in\mathcal{I}}$ from $\mathcal{H}$ to $\mathcal{H}'$ if, for each $j \in \mathcal{I}$ there is an instrument $\mathfrak{J}^{(j)} = \{\mathcal{E}_{k|j}\}_{k\in\mathcal{I}^{(j)}}$ from $\mathcal{H}'$ to $\mathcal{H}''$ such that

$$\mathfrak{J}' = \{\mathcal{E}_{k|j} \circ \mathcal{E}_j\}_{j\in\mathcal{I}, k\in\mathcal{I}^{(j)}}.$$

Such an instrument has the following interpretation: After a measurement corresponding to the instrument $\mathfrak{J}$ is performed, a subsequent measurement is performed. The second instrument used is conditioned on the outcome of the measurement from the initial instrument. The resulting state $\rho_{j,k} = \frac{1}{p_{j,k}}\mathcal{E}_{k|j}(\mathcal{E}_j(\rho))$ depends on the outcome of the two instruments. This characterization of quantum instruments will be useful for defining the class of LOCC operations.

2.3 Quantum entanglement

Within quantum information theory, the theory of entanglement [BŻ06, HHHH09] is one of the most important and active areas of research. Entanglement is a necessary ingredient for many quantum information processing tasks, including the teleportation of quantum states [BBC+93], superdense coding [BW92], and numerous uses in quantum cryptography protocols [BB84]. Two principal features of entanglement are that it cannot be created among distant parties when there is none to begin with, and that it is depleted in the implementation of such protocols. Not all entanglement is created equal. Some entangled states may be more useful for certain applications than other entangled states. It is therefore of great interest to develop a detailed understanding of the properties of entanglement in terms of its nature as a *resource* [PV07]. In this section we will introduce the mathematical formulation of entanglement in bipartite quantum systems.

A pure state $|\psi\rangle \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ is said to be *separable* if it can be written as an elementary tensor $|\psi\rangle = |u\rangle \otimes |v\rangle$ for some $|u\rangle \in \mathcal{H}_{\mathsf{A}}$ and $|v\rangle \in \mathcal{H}_{\mathsf{B}}$. Otherwise $|\psi\rangle$ is said to be *entangled*. We have already seen that any bipartite pure state can be written in Schmidt form as

$$|\psi\rangle = \sum_{j=1}^{d} \sqrt{\lambda_j} |u_j\rangle \otimes |v_j\rangle,$$

where $\lambda_1 \geq \cdots \geq \lambda_d$ are the Schmidt coefficients of $|\psi\rangle$, and $\{|u_j\rangle\}$ and $\{|v_j\rangle\}$ are orthonormal. The number of non-zero Schmidt coefficients is known as the *Schmidt rank* of $|\psi\rangle$. Clearly, a pure state is separable if and only if its Schmidt rank is 1.

In the case of mixed states, a density operator $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ is said to be separable if it can be written as a convex combination of separable pure states

$$\rho = \sum_j p_j |u_j\rangle\langle u_j| \otimes |v_j\rangle\langle v_j|$$

for some unit vectors $|u_j\rangle \in \mathcal{H}_{\mathsf{A}}$ and $|v_j\rangle \in \mathcal{H}_{\mathsf{B}}$ and some probabilities $p_j \geq 0$ such that $\sum_j p_j = 1$. Otherwise ρ is said to be entangled. More generally, an operator $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ is separable if it can be written in the form $X = \sum_j Y_j \otimes Z_j$ for some positive operators $Y_j, Z_j \geq 0$ on $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ respectively. The set of separable operators on $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ will be denoted $\mathrm{Sep}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$, and we will write $\mathrm{SepD}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$ to denote the set of separable density operators on $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$,

$$\mathrm{SepD}(\mathcal{H}_\mathsf{A} : \mathcal{H}_\mathsf{B}) = \mathrm{D}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}) \cap \mathrm{Sep}(\mathcal{H}_\mathsf{A} : \mathcal{H}_\mathsf{B}).$$

2.3.1 Local Operations and Classical Communication

In the paradigmatic setting for the study of entanglement, distant parties (named Alice and Bob) jointly share a state of a composite quantum system. The systems that each party has access to are labeled $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ respectively. Procedures that can be performed in such a setting are limited to those that can be implemented through local operations (LO) on the subsystems and exchange of classical communication (CC) between the parties. The class of operations that can be performed in this manner is known as *LOCC*. Here we will describe this class of operations using the notion of quantum instruments defined earlier (see, e.g., [CLM+14] for more details regarding this constriction).

A quantum instrument $\mathfrak{I}$ from $\mathcal{H}_\mathsf{AB}$ to $\mathcal{H}_\mathsf{A'B'}$ is said to be *one-way* A*-local* if it is of the form

$$\mathfrak{I} = \{\mathcal{E}_j \otimes \Phi_j\}_{j \in \mathcal{I}}$$

for some completely positive (but not necessarily trace preserving) superoperators on Alice's system $\mathcal{E}_j : \mathrm{L}(\mathcal{H}_\mathsf{A}) \rightarrow \mathrm{L}(\mathcal{H}_\mathsf{A'})$ and some quantum channels $\Phi_j \in \mathrm{C}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{B'})$ on Bob's, such that the sum $\mathcal{E} = \sum_{j \in \mathcal{I}} \mathcal{E}_j$ is trace preserving. That is $\{\mathcal{E}_j\}_{j \in \mathcal{I}}$ is itself an instrument from $\mathcal{H}_\mathsf{A}$ to $\mathcal{H}_\mathsf{A'}$. We give A-local instruments the following interpretation. Alice performs a measurement on her system corresponding to the instrument $\{\mathcal{E}_j\}_{j \in \mathcal{I}}$ and she subsequently communicates the outcome of the measurement to Bob, who then performs the channel Φ_j on his system. The definition of a *one-way* B*-local* instrument is analogous, where Bob performs a measurement corresponding to some quantum instrument and communicates the result to Alice. A channel that is one-way A- or B-local is an LOCC channel that can be implemented with one round of communication. To define the full set of LOCC instruments, we must define a notion of local conditioning.

An instrument $\mathfrak{J}$ from $\mathcal{H}_{\mathsf{AB}}$ to $\mathcal{H}_{\mathsf{A''B''}}$ is said to be A-*locally conditioned* on an instrument $\mathfrak{J}' = \{\mathcal{E}_j\}_{j\in\mathcal{I}}$ from $\mathcal{H}_{\mathsf{AB}}$ to $\mathcal{H}_{\mathsf{A'B'}}$ if $\mathfrak{J}$ is conditioned on $\mathfrak{J}'$ by A-local channels. More specifically, for each $j \in \mathcal{I}$ there is a one-way A-local instrument $\mathfrak{J}^{(j)} = \{\mathcal{E}_{k|j}\}_{k\in\mathcal{I}^{(j)}}$ from $\mathcal{H}_{\mathsf{A'B'}}$ to $\mathcal{H}_{\mathsf{A''B''}}$ such that $\mathfrak{J} = \{\mathcal{E}_{k|j} \circ \mathcal{E}_j\}_{j\in\mathcal{I},k\in\mathcal{I}^{(j)}}$. We give A-local conditioning the following interpretation. Alice performs a measurement conditioned on the result of the measurement corresponding to $\mathfrak{J}'$ and communicates the final result to Bob. That is, if outcome j is obtained from the first measurement, Alice performs measurement $\mathfrak{J}^{(j)}$ and obtains outcome k. The definition of B-locally conditioning is analogous.

The full definition of LOCC instruments is now stated. A quantum instrument $\mathfrak{J}$ from $\mathcal{H}_{\mathsf{AB}}$ to $\mathcal{H}_{\mathsf{A'B'}}$ is said to be

- in LOCC_1 if it is one-way A- or B-local;
- in LOCC_r for $r > 1$ if it is A- or B-locally conditioned on an instrument in LOCC_{r-1};
- and in LOCC if it is in LOCC_r for some $r \in \mathbb{N}$.

If $\mathfrak{J} \in \mathrm{LOCC}_r$ for some $r \in \mathbb{N}$, it can be performed by LOCC with r rounds of communication between the parties.

A channel $\Lambda \in \mathrm{C}(\mathcal{H}_{\mathsf{AB}}, \mathcal{H}_{\mathsf{A'B'}})$ is said to be LOCC if it is of the form $\Lambda(X) = \sum_{j\in\mathcal{I}} \Lambda_j(X)$ for some LOCC instrument $\{\Lambda_j\}_{j\in\mathcal{I}}$. We write $\mathrm{LOCC}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{A'}} : \mathcal{H}_{\mathsf{B}}, \mathcal{H}_{\mathsf{B'}})$ to denote the set of all LOCC channels from $\mathcal{H}_{\mathsf{AB}}$ to $\mathcal{H}_{\mathsf{A'B'}}$. Given two bipartite states $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{AB}})$ and $\rho' \in \mathrm{D}(\mathcal{H}_{\mathsf{A'B'}})$, we say that ρ can be converted into ρ' by LOCC operations if there exists an LOCC channel Λ such that $\Lambda(\rho) = \rho'$. If this holds, then we write $\rho \xrightarrow{\mathrm{LOCC}} \rho'$.

Given two states ρ and σ of a bipartite quantum system, the fundamental question that we want to answer is the following: Can we obtain σ from ρ using only LOCC? The possible transformations of resources establishes a partial order on the set of all possible states.

2.3.2 Entanglement measures

Axioms for entanglement measures

In light of the view of entanglement as a resource that can be used up, it is important to understand the resourcefulness of entanglement by quantifying it. The standard setting for understanding entanglement is the LOCC paradigm, so we will require that any measure of entanglement be in some sense monotonic with respect to local operations and classical communication. There are two distinct ways of axiomatizing the definition of an entanglement measure that we will list here. For the following definitions, let E be a function on bipartite states.

(E1) **Monotonicity under (deterministic) LOCC**. For all bipartite states ρ and all LOCC channels Λ it holds that $E(\rho) \geq E(\Lambda(\rho))$.

(E2) **Monotonicity under probabilistic LOCC**. For all bipartite states ρ and all LOCC instruments $\{\Lambda_j\}$ it holds that $E(\rho) \geq \sum_j p_j E(\frac{1}{p_j}\Lambda_j(\rho))$, where $p_j = \mathrm{Tr}(\Lambda_j(\rho_j))$.

Both conditions imply that E cannot increase under LOCC operations. It is clear that (E2) implies (E1), since any LOCC channel Λ can be represented by an LOCC instrument with exactly one output $\{\Lambda\}$. Historically, axiom (E2) was originally used as the definition for entanglement measures [HHHH09]. But the first axiom (E1) is more fundamental, since it gives information about entanglement of the *state* ρ, while (E2) gives information about the average entanglement of an ensemble $\{(p_i, \rho_i)\}$, which is less operational than the notion of a state. Furthermore, axiom (E1) is all that is needed when examining convertibility of states in the single-shot setting, which is primarily the cases in this thesis.

Since any separable state can be obtained using LOCC from any other bipartite state, both axioms (E1) and (E2) imply that any entanglement measure E must be constant on the set of separable states. For convenience, we will require that this constant must be zero, which we state as an additional axiom for entanglement measures.

(E0) **Vanishing on separable states**. $E(\rho) = 0$ for all separable bipartite states ρ.

An entanglement measure is said to be *faithful* if $E(\rho) = 0$ implies that ρ is separable. Faithful measures of entanglement give useful criteria for detecting when a state is entangled, but faithfulness is not a requirement of entanglement measures.

An entanglement measure is said to be *convex* if $E\left(\sum_j p_j \rho_j\right) \leq \sum_j p_j E(\rho_j)$. Convex measures of entanglement have the interpretation that entanglement should not increase under mixing (or lack of knowledge). While we will not require all entanglement measures to be convex, it will nevertheless be useful in some settings to restrict our attention to convex measures.

Entanglement measures for pure states

Entanglement for pure states can be completely characterized by their Schmidt coefficients. The mathematical tool of *majorization* is useful here. Let $\boldsymbol{\lambda}, \boldsymbol{\lambda'} \in \mathbb{R}^d$ be vectors of real numbers whose entries are in decreasing order. We say that $\boldsymbol{\lambda'}$ *majorizes* $\boldsymbol{\lambda}$ if it holds that

$$\sum_{j=1}^{k} \lambda_j \leq \sum_{j=1}^{k} \lambda'_j \qquad \text{for all } k = 1, \ldots, d,$$

and we write $\boldsymbol{\lambda} \prec \boldsymbol{\lambda'}$. This gives a neat way to characterize the convertibility of pure states under LOCC.

Theorem 2.12 ([Nie99])**.** *Let $|\psi\rangle \in \mathbb{C}^m \otimes \mathbb{C}^n$ and $|\psi'\rangle \in \mathbb{C}^{m'} \otimes \mathbb{C}^{n'}$ be bipartite pure states. There exists an LOCC channel Λ such that $\Lambda(|\psi\rangle\langle\psi|) = |\psi'\rangle\langle\psi'|$ if and only if it holds that $\boldsymbol{\lambda} \prec \boldsymbol{\lambda'}$, where $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda'}$ are the vectors of Schmidt coefficients of $|\psi\rangle$ and $|\psi'\rangle$.*

A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be *Schur concave* if $\boldsymbol{\lambda} \prec \boldsymbol{\lambda'}$ implies $f(\boldsymbol{\lambda}) \geq f(\boldsymbol{\lambda'})$ for all $\boldsymbol{\lambda}, \boldsymbol{\lambda'} \in \mathbb{R}^d$. If a function f is concave and symmetric then it is Schur concave, but the converse does not hold. Since Schur concave functions are exactly the functions that are monotonic under majorization, they can be used to quantify entanglement of pure states.

Indeed, given a Schur concave function f, we can define a measure of entanglement on pure states by $E_f(\psi) = f(\boldsymbol{\lambda})$.

Some examples of entanglement measures for pure states include the well-known *entropy of entanglement*, which is simply the Shannon entropy of the Schmidt coefficients

$$E(\psi) = H(\boldsymbol{\lambda}) = -\sum_i \lambda_i \log \lambda_i$$

and the *Rényi entropies* of entanglement

$$E_\alpha(\psi) = H(\boldsymbol{\lambda}) = \frac{1}{1-\alpha} \log\Big(\sum_i \lambda_i^\alpha\Big)$$

for $\alpha \in [0, +\infty]$. Another well-known family of entanglement measures that will be used in this thesis are the *Vidal measures* [Vid00b], which are defined by

$$\begin{aligned} E_k(\psi) &= 1 - \sum_{i=1}^{k} \lambda_i & (2.11) \\ &= \sum_{i=k+1}^{d} \lambda_i & (2.12) \end{aligned}$$

where the Schmidt coefficients are assumed to be in decreasing order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. These are simply the sum of the $(d-k)$- smallest Schmidt coefficients of $|\psi\rangle$.

Examples of entanglement measures

A few common types of measures of entanglement will be introduced here. For a survey of entanglement measures, see [Hor01, HHHH09].

Distillable entanglement and entanglement cost We will first define two fundamental measures of entanglement that have clear operational interpretation. The *entangle-*

ment cost of a bipartite state ρ is defined as

$$E_C(\rho) = \inf\Big\{r \geq 0 \,\Big|\, \exists \Lambda_n \in \text{LOCC s.t. } \lim_{n\to\infty} \Big\| \Lambda_n\Big(|\phi^+\rangle\!\langle\phi^+|^{\otimes \lfloor rn \rfloor}\Big) - \rho^{\otimes n} \Big\|_1 = 0\Big\} \tag{2.13}$$

It is the smallest rate r at which n copies of ρ can be extracted from rn copies of the two-qubit Bell state $|\phi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ in the limit of large n. The *distillable entanglement* of a bipartite state ρ is defined as

$$E_D(\rho) = \sup\Big\{r \geq 0 \,\Big|\, \exists \Lambda_n \in \text{LOCC s.t. } \lim_{n\to\infty} \Big\| \Lambda_n(\rho^{\otimes n}) - |\phi^+\rangle\!\langle\phi^+|^{\otimes \lfloor rn \rfloor} \Big\|_1 = 0\Big\} \tag{2.14}$$

It is the largest rate r at which rn copies of the two-qubit Bell state can be extracted from n copies of ρ in the limit of large n. A state is said to be *distillable* if $E_D(\rho) > 0$. It is clear that $E_D \leq E_C$, and it is also known that $E_D(\psi) = E_C(\psi)$ for all bipartite pure states. While the entanglement cost and distillable entanglement have a clear operational interpretation, they are difficult to compute.

Distance-based measures Given a suitable distance measure for quantum states (see Section 2.2.1, one can define a measure of entanglement based on this distance. Indeed, if f is a suitable distance measure, we can define

$$E_f^{\text{Sep}}(\rho) := \inf_{\sigma\in\text{SepD}} f(\rho,\sigma).$$

Indeed, if f is monotonic under completely positive and trace preserving maps, then E_f is an entanglement measure satisfying (E1). For example, the *relative entropy of entanglement* is defined by

$$E_R^{\text{Sep}}(\rho) := \inf_{\sigma\in\text{SepD}} S(\rho\|\sigma).$$

This is known to be an upper bound to the distillable entanglement [Rai01]. Similarly, we can define the Rényi α-relative entropies by using functions defined in (2.10). In Chapter 4, necessary and sufficient conditions will be given for when a state $\sigma \in \mathrm{SepD}$ satisfies $E_f^{\mathrm{Sep}}(\rho) = f(\rho, \sigma)$ (i.e., when σ minimizes the distance of ρ to the set of separable states for a distance function f). This will also be generalized to convex sets other than the set of separable states.

Convex roof measures Given an entanglement measure E on pure states, we can also consider the following method of constructing an entanglement measure on *all* states by extending it via the so-called convex roof [Uhl10, BL13]:

$$\widehat{E}(\rho) = \inf_{\{(p_j,\psi_j)\}} \sum_j p_j E(\psi_j) \tag{2.15}$$

where the infimum is taken over all pure state decompositions of ρ, i.e. $\rho \sum_j p_j |\psi_j\rangle\langle\psi_j|$ for some pure states $|\psi_j\rangle$ and probabilities $p_j \geq 0$ such that $\sum_j p_j = 1$. It is easy to see that this function is convex. In fact, it is the largest convex function on $\mathrm{D}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B})$ that satisfies $\widehat{E}(\psi) \leq E(\psi)$ on all pure states. The following theorem of Vidal [Vid00a] states the conditions on f for the convex roof of E_f to be an entanglement measure.

Theorem 2.13. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a concave, symmetric function satisfying $f(1, 0, \ldots, 0) = 0$ and define a function E on pure bipartite states as $E(\psi) = f(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is the vector of Schmidt coefficients of $|\psi\rangle$. Then the convex roof of E*

$$\widehat{E}(\rho) = \inf_{\{(p_i,\psi_i)\}} \sum_i p_i E(\psi_i) \tag{2.16}$$

is a full entanglement measure satisfying (E2).

In particular, the function f is Schur concave since it is concave and symmetric, but not all Schur concave functions are of this form. Furthermore, *all* convex roof entanglement measures that are evaluated on pure states must be of the form in Theorem 2.13. While entanglement measures for pure states can be computed easily, actually computing the value of a convex roof measure on arbitrary mixed states is difficult in general. In Chapter 5, it will be shown how to compute arbitrary convex roof monotones on certain classes of symmetric states.

2.4 Approximations to LOCC and separability

LOCC channels are quite messy to represent mathematically, so it is common to work instead with other sets of quantum channels that approximate the set of LOCC channels. In particular, we will work with the sets of *separable* and *PPT* channels.

2.4.1 Separable channels

A superoperator $\Phi : \mathrm{L}(\mathcal{H}_{\mathsf{AB}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A'B'}})$ is said to be *separable* if it has a Kraus representation $\{K_j\}$ in which all of the Kraus operators are product operators $K_j = X_j \otimes Y_j$ for some $X_j \in \mathrm{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{A'}})$ and $Y_j \in \mathrm{L}(\mathcal{H}_{\mathsf{B}}, \mathcal{H}_{\mathsf{B'}})$. The superoperator can then be expressed as

$$\Phi(\rho) = \sum_j X_j \otimes Y_j \rho (X_j \otimes Y_j)^\dagger$$

for all $\rho \in \mathrm{L}(\mathcal{H}_{\mathsf{AB}})$. Equivalently, a superoperator Φ is separable if its Choi matrix $J(\Phi) \in \mathrm{L}(\mathcal{H}_{\mathsf{A'B'}} \otimes \mathcal{H}_{\mathsf{AB}})$ is separable operator with respect to the bipartite splitting $\mathcal{H}_{\mathsf{AA'}} : \mathcal{H}_{\mathsf{BB'}}$, i.e., $J(\Phi) \in \mathrm{Sep}(\mathcal{H}_{\mathsf{AA'}} : \mathcal{H}_{\mathsf{BB'}})$.

Every LOCC channel is a separable channel, but the converse is not true. That is, there are separable channels that cannot be implemented via the LOCC paradigm described above [BDF+99, CLM+14]. The distinction between separable and LOCC channels is still not particularly well-understood, but has been explored in [GG07, GG08, Ghe10]. The study of separable maps are useful since their simple mathematical characterization makes working with them fairly straightforward, and anything that holds for separable channels necessarily also holds for LOCC channels. However, determining the separability of an operator is known to be NP-hard [Gur03], so it is still not computationally feasible to consider the class of all separable superoperators.

2.4.2 PPT states and channels

Given a bipartite Hilbert space $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$, we define the *partial transpose* as the superoperator $T_{\mathsf{B}} = \mathrm{id}_{\mathsf{A}} \otimes T$ on $\mathrm{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$, where T is the transposition map on $\mathcal{H}_{\mathsf{B}}$. Given an operator $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$, we write $X^{T_{\mathsf{B}}}$ for its partial transpose. As seen in Example 2.4, the transpose is not a completely positive map, so there are positive operators $X \geq 0$ for which $X^{T_{\mathsf{B}}} \not\geq 0$. However, any *separable* operator $X \in \mathrm{Sep}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$ will have positive partial transpose, since $(Y \otimes Z)^{T_{\mathsf{B}}} = Y \otimes Z^{T} \geq 0$ for any $Y, Z \geq 0$. A positive operator $X \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})_{+}$ is said to be *positive under partial transpose* (or *PPT*) if it holds that $X^{T_{\mathsf{B}}} \geq 0$. In particular, if a state $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{AB}})$ is not PPT ($\rho^{T_{\mathsf{B}}} \not\geq 0$), then it is necessarily entangled.

A superoperator $\Phi : \mathrm{L}(\mathcal{H}_{\mathsf{AB}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A'B'}})$ is said to be PPT if its Choi matrix is PPT with respect to the bipartite splitting $\mathsf{AA'} : \mathsf{BB'}$. That is, if it holds that $J(\Phi)^{T_{\mathsf{BB'}}} \geq 0$. It is clear that any separable superoperator will be PPT, and thus any LOCC channel is PPT. However, unlike for separable superoperators, checking whether an operator is PPT is computationally feasible, and optimizing over PPT operators can be done with a semidefinite program. This makes the class of PPT channels very practical.

Furthermore, it is known that any state $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{AB}})$ that is PPT is not distillable [Hor97, BDM+99], i.e., $E_D(\rho) = 0$ if $\rho^{T_{\mathsf{B}}} \geq 0$. It is also possible to define the *PPT-distillable entanglement* analogously as

$$E_D^{\mathrm{PPT}}(\rho) = \sup\Big\{ r \geq 0 \,\Big|\, \exists \Phi_n \in \mathrm{PPT} \text{ s.t. } \lim_{n\to\infty} \Big\| \Phi_n(\rho^{\otimes n}) - |\phi^+\rangle\langle\phi^+|^{\otimes \lfloor rn \rfloor} \Big\|_1 = 0 \Big\}. \tag{2.17}$$

Since all LOCC channels are PPT channels, it holds that $E_D(\rho) \leq E_D^{\mathrm{PPT}}(\rho)$, so the PPT-distillable entanglement is an upper bound to the standard distillable entanglement. Furthermore, all PPT states are non-distillable, a fact which can easily be shown as follows. If ρ is PPT then $\rho^{\otimes n}$ is PPT for any number of copies of the state n. Since every PPT channel is PPT-preserving on states, $\Phi(\rho^{\otimes n})$ will be a PPT state for all PPT channels Φ and all n. The two-qubit Bell state $|\phi^+\rangle\langle\phi^+|$, however, is bounded away from the set of PPT states, so no PPT channels can ever bring a PPT state close to a Bell state. Hence $E_D^{\text{PPT}}(\rho) = 0$ for all PPT states ρ, and thus the LOCC-distillable entanglement vanishes for these states as well. It remains an open problem as to whether there exist any non-distillable states that are not PPT [PPHH10, Cla06].

Similarly, analogous to the relative entropy of entanglement for separable states, one can also define the relative entropy with respect to the PPT states,

$$E_R^{\text{PPT}}(\rho) = \inf_{\sigma \in \text{PPTD}} S(\rho \| \sigma). \tag{2.18}$$

This is an upper bound to the PPT-distillable entanglement [Rai99b], so it is an upper bound to the LOCC-distillable entanglement as well. Another improved upper bound to the distillable entanglement, introduced in [Rai99a, Rai01], is known as the *Rains bound* and is defined by

$$R(\rho) = \inf_{\sigma \in \text{D}} S(\rho \| \sigma) + \log \| \sigma^{T_{\text{B}}} \|_1$$

where the infimum is taken over all density matrices, not just the PPT ones. Computing $E_R^{\text{PPT}}(\rho)$ and $R(\rho)$ for arbitrary states can be done in practice for states of small dimensions, since they can be cast as convex optimization problems as shown in Chapter 4. Analytical results regarding when a state σ minimizes the relative entropy of PPT entanglement and the Rains bound will be studied in Chapter 4.

2.5 Resource theories of states in quantum information

The study of resource theories originated from the observation that certain properties of physical systems become valuable when the operations that can be performed are restricted so that such properties are difficult to create. The prototypical example of such a property is quantum entanglement, which becomes a key resource for many quantum information processing tasks when one is restricted to only local operations and classical communication (LOCC). In light of the view of quantum entanglement as a resource, much work has been done to generalize the study of resource theories in quantum information. This framework has been applied to various other concepts in quantum information, such as purity [HHO03], coherence [BCP14, CG16, WY16], magic states (or non-stabilizer states) for quantum computation [VHGE14], asymmetry [MS14], and thermodynamics [BHO$^+$13, GMN$^+$15b]. Beyond the quantum framework, resource theories can be studied as mathematical entities in their own right [Fri13, CFS16]. For recent reviews of quantum resource theories, see [HO13, BG15]. In this section, we discuss the general notion of resource theories before giving the definition for the structure of resource theories that will be studied in this theses.

Every resource theory has three main ingredients that define its structure: the *resources* are the class of objects that are to be manipulated in some way, the *free resources* are the resource objects that can be obtained for 'free', and the *free operations* are the allowable ways of manipulating the resources. If one starts with a free object of a resource theory, it is not possible to turn it into a non-free resource using only the allowable free operations. Conversely, it should always be possible to produce a particular free object given any initial starting resource (free or otherwise). The resources which cannot be created by means of the set of free operations naturally acquire some value, and manipulating this resource consumes the value of this resource in the process.

At the core of the study of any resource theory is the question of *resource conversion.* That is, given two resources, we would like to know if it is possible to convert one into the other using the allowable free operations. One method of determining transformability of resources is by quantifying the amount of resource in a state with *resource monotones*, which are functions of resources that are non-increasing under application of free operations. Convertibility of resources in resource theories can also be characterized by resource *conversion witnesses*, which will be discussed in 2.5.1.

In quantum information, the mathematical structure of resource theories typically arise from certain physical restrictions that allow experimenters only to implement a restricted set of physical manipulations of quantum systems. In such resource theories, the collection of resources consists of all of the possible quantum states on some quantum systems, and the free operations comprise some subset of completely positive trace-preserving maps (quantum channels) between the quantum systems. For example, entanglement of bipartite systems can be considered as a resource theory in which the resources are entangled quantum states and parties are only able to perform local operations and communicate classical information. However resource theories can also be considered that do not represent physical constraints. For example, we can consider the "resource theory of separable operations" in which only separable channels (rather than LOCC channels) can be implemented, which do not necessarily represent physical operations. Given a class of free resources, we can also analyze a resource theory where the set of allowed operations is defined as the largest class of operations that do not generate non-free resources from free ones. Such resource theories no longer necessarily have a physical interpretation, but studying them can yield useful results about resource theories of physical interest.

In this thesis, we will only be interested in resource theories within the framework of quantum information, where the collection of resources consists of all quantum states on some Hilbert spaces. Furthermore, all resource theories considered here will be finite-dimensional. The *resources* of a quantum resource theory are all of the density operators on some family of Hilbert spaces, and the *free states* will be some subset density operators. The *free operations* will be some subset of completely positive trace preserving maps (i.e., quantum channels).

There are many notions of resource theories in quantum information that have been studied that will not be of interest in this paper. Quantum resource theories can also be studied in the asymptotic limit of many copies of a single resource [BH15]. In a resource theory where the set of free states is convex, and the allowed operations are the set of all operations that do not asymptotically generate a resource, the asymptotic conversion rate is given in terms of the regularized relative entropy of a resource. In this case, this the unique resource measure in the asymptotic limit of many copies of the state. In this thesis, however, we will only be concerned with the *single-shot* setting, in which we are interested in convertibility of a single copy of a state. Other recent work examines the conditions for the existence of *resource destroying maps* in quantum resource theories [LHL17].

For the purposes of this thesis, we will mathematically define resource theories in the following manner. Recall that $\mathrm{D}(\mathcal{H})$ denotes the set of density operators on a (finite dimensional) Hilbert space $\mathcal{H}$ and $\mathrm{C}(\mathcal{H}, \mathcal{H}')$ denotes the set of quantum channels from the operators on system $\mathcal{H}$ to the operators on $\mathcal{H}'$. A *resource theory* $(\mathscr{F}, \mathscr{O})$ consists of a family $\mathscr{F} = \{\mathscr{F}(\mathcal{H})\}$ of *free states* and a family $\{\mathscr{O}(\mathcal{H}, \mathcal{H}')\}$ of *free operations*:

- $\mathscr{F}(\mathcal{H}) \subseteq \mathrm{D}(\mathcal{H})$ for finite-dimensional Hilbert spaces $\mathcal{H}$
- $\mathscr{O}(\mathcal{H}, \mathcal{H}') \subseteq \mathrm{C}(\mathcal{H}, \mathcal{H}')$ for pairs of Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$

that satisfy the following conditions:

1. If $\rho \in \mathscr{F}(\mathcal{H})$ and $\mathcal{E} \in \mathscr{O}(\mathcal{H}, \mathcal{H}')$, then $\mathcal{E}(\rho) \in \mathscr{F}(\mathcal{H}')$. ("Free states remain free under free operations.")

2. It holds that $\mathrm{id}_{\mathcal{H}} \in \mathscr{O}(\mathcal{H}, \mathcal{H})$. ("The identity operation is free.")

3. If $\mathcal{E} \in \mathscr{O}_{\mathcal{H},\mathcal{H}'}$ and $\mathcal{E} \in \mathscr{O}(\mathcal{H}', \mathcal{H}'')$, then $\mathcal{E}' \circ \mathcal{E} \in \mathscr{O}(\mathcal{H}, \mathcal{H}'')$. ("Composition of free operations is free.")

4. If $\rho \in \mathscr{F}(\mathcal{H})$ then the channel defined by $\mathcal{E}_\rho(X) = \mathrm{Tr}(X)\rho$ is in $\mathscr{O}(\mathcal{H}', \mathcal{H})$. ("Any free state is obtainable from any other state.")

There are numerous other aspects that general resource theories should have in quantum information (for example, if $\rho \in \mathscr{F}(\mathcal{H})$ and $\rho' \in \mathscr{F}(\mathcal{H}')$, then it should hold that $\rho \otimes \rho' \in \mathscr{F}(\mathcal{H} \otimes \mathcal{H}')$). However, the conditions outlined above suffice for the treatment of resource theories in this thesis, since we are primarily concerned with classifying existence of a free operation that convert a single resource state to another .

Given a resource theory of quantum states $(\mathscr{F}, \mathscr{O})$ and density operators $\rho \in \mathrm{D}(\mathcal{H})$ and $\rho' \in \mathrm{D}(\mathcal{H}')$, we write $\rho \xrightarrow{\mathscr{O}} \rho'$ if there exists a free operation $\mathcal{E} \in \mathscr{O}(\mathcal{H}, \mathcal{H}')$ such that $\mathcal{E}(\rho) = \rho'$. That is, it denotes when "ρ can be converted into ρ' under free operations. This induces a pre-order on density operator. Indeed, since the identity channel is always free in any resource theory, it holds that $\rho \xrightarrow{\mathscr{O}} \rho$ for any $\rho \in \mathrm{D}(\mathcal{H})$. Furthermore, if $\rho \xrightarrow{\mathscr{O}} \rho'$ and $\rho' \xrightarrow{\mathscr{O}} \rho''$ then $\rho \xrightarrow{\mathscr{O}} \rho''$ since we can compose free operations. In any resource theory, the primary problem in the single-shot setting is to find necessary and sufficient conditions for the convertibility of two resource states. That is, given ρ and ρ', we would like to find convenient conditions to characterize when $\rho \xrightarrow{\mathscr{O}} \rho'$.

The theory of entanglement under the restriction to LOCC channels can be viewed as a quantum resource theory under this definition, where the Hilbert spaces are all bipartite $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$, the sets of free states are the separable states $\mathscr{F}(\mathcal{H}_{\mathsf{AB}}) = \mathrm{SepD}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$, and the free operations are the LOCC channels. In fact, entanglement theory was the motivating example for studying other resource theories in quantum information. In the bipartite setting, one can also consider the resource theory of separable operations (in which the free states are again the sets of separable states), and the theory of PPT operations (in which the free states are the PPT states). The concept of *convex* resource theories will be studied in Chapter 6.

Finally, it must be noted that the definition of resource theories given here naturally has the structure of a category. In fact, a complete formalism for resource theories using category theory has been constructed, and such an approach may have useful applications in analyzing resource theories [CFS16]. The categorical approach to resources, however, is beyond the scope of this thesis and will not be used here.

2.5.1 Resource monotones and conversion witnesses

This section reviews the concepts of resource monotones and resource conversion witnesses, a generalization of resource monotones. The rich structure of these conversion witnesses is also explored.

Given a resource theory $(\mathscr{F}, \mathscr{O})$ of quantum states, a *resource monotone* is a function M from states to real numbers that is non-increasing under free operations. That is, M is a monotone if, for any free operation $\mathcal{E} \in \mathscr{O}(\mathcal{H}, \mathcal{H}')$ and any state $\rho \in \mathscr{F}(\mathcal{H})$, it holds that $M(\mathcal{E}(\rho)) \leq M(\rho)$. In any resource theory of quantum states, one resource monotone that can be used is the relative entropy with respect to the free states. Analogous to the relative entropy of entanglement, this is defined as

$$M_R^{\mathscr{F}}(\rho) := \min_{\sigma \in \mathscr{F}(\mathcal{H})} S(\rho \| \sigma)$$

for any $\rho \in \mathrm{D}(\mathcal{H})$. A family of monotones $\{M_i\}_{i \in \mathfrak{I}}$ is said to be *complete* if it holds that $\rho \xrightarrow{\mathscr{O}} \rho'$ if and only if $M_i(\rho) \geq M_i(\rho')$ for all $i \in \mathfrak{I}$. In Chapter 6, it will be shown how to construct a complete family of resource monotones for quantum resource theories where the sets of free operations are convex.

The most general technique for characterizing the convertibility of states is through *conversion witnesses*, a concept that was first introduced as *relative monotones* in [San10] and further investigated in [GMN$^+$15a, sec. II.A.] and [GG15]. Consider a resource theory $(\mathscr{F}, \mathscr{O})$ of quantum states, and let W be a real-valued function on pairs of quantum states. If $W(\rho, \sigma) \geq 0$ implies that $\rho \xrightarrow{\mathscr{O}} \sigma$, then W is said to be a *go witness*. If $W(\rho, \sigma) < 0$ implies that $\rho \not\mapsto \sigma$, then W is said to be a *no-go witness*. Finally, W is said to be a *complete witness* if it is both a go and a no-go witness.

Given a monotone M, we can define a no-go witness by $W_M(\rho, \sigma) = M(\rho) - M(\sigma)$. Indeed $W_M(\rho, \sigma) < 0$ implies $M(\rho) < M(\sigma)$, and thus $\rho \overset{\mathscr{O}}{\not\to} \sigma$ by the monotonicity of M. Hence resource monotones can be considered as a special case of resource conversion witnesses.

The set of all no-go witnesses in a resource theory is endowed with the structure of a partially ordered set. Indeed, given two no-go witnesses W_1 and W_2, we say that $W_1 \succeq W_2$ if

$$W_2(\rho, \sigma) < 0 \implies W_1(\rho, \sigma) < 0 \quad \text{for all } \rho, \sigma. \tag{2.19}$$

That is, $W_1 \succeq W_2$ means that the witness W_1 tells us more information about the convertibility of states than W_2 does. If W_2 detects the inconvertibility $\rho \overset{\mathcal{G}}{\not\to} \sigma$ for some states ρ and σ, this same information can already be obtained by W_1. But W_1 might be able to detect the inconvertibility of other pairs of states that W_2 cannot.

The partial order structure of no-go witnesses is illuminated in the following example: Given a family of no-go witnesses $(W_i)_{i \in \mathcal{I}}$, we can construct a new no-go witness $W_{\mathcal{I}}$ by minimizing over all witnesses in the family

$$W_{\mathcal{I}}(\rho, \sigma) := \min_{i \in \mathcal{I}} W_i(\rho, \sigma).$$

This is indeed a witness, since $W_{\mathcal{I}}(\rho, \sigma) < 0$ implies that $W_i(\rho, \sigma) < 0$ for at least one $i \in \mathcal{I}$ and thus $\rho \not\mapsto \sigma$. Hence $W_{\mathcal{I}} \succeq W_i$ and the resulting witness $W_{\mathcal{I}}$ is an improvement over each of the conversion witnesses W_i. Similarly, given a family $(M_i)_{i \in \mathcal{I}}$ of monotones, one can define a witness

$$W_{\mathcal{I}}(\rho, \sigma) := \min_{i \in \mathcal{I}} \{M_i(\rho) - M_i(\sigma)\} = \min_{i \in \mathcal{I}} W_{M_i}(\rho, \sigma).$$

If the family $(M_i)_{i \in \mathcal{I}}$ is complete, then the resulting $W_{\mathcal{I}}$ is a complete conversion witness. Furthermore, if $\overline{W}$ is a complete witness, then $\overline{W} \succeq W$ for any no-go witness W.

An example hierarchy of no-go conversion witnesses is depicted in Figure 2.1. Note that two no-go conversion witnesses W_1 and W_2 may be incomparable in general. That is, it may be that both $W_1 \not\succeq W_2$ and $W_2 \not\succeq W_1$.

An analogous partial order exists for *go* witnesses. If W_1 and W_2 are two go witnesses

$$\begin{tikzcd}
 & W_{\{1,2,3\}} \arrow[dl] \arrow[d] \arrow[dr] & \\
W_{\{1,2\}} \arrow[d] \arrow[dr] & W_{\{1,3\}} \arrow[dl] \arrow[dr] & W_{\{2,3\}} \arrow[dl] \arrow[d] \\
W_1 & W_2 & W_3
\end{tikzcd}$$

Figure 2.1: An example hierarchy of no-go conversion witnesses in which an arrow between witnesses $X \longrightarrow Y$ denotes $X \succeq Y$. Consider three no-go witnesses W_1, W_2 and W_3, which may be incomparable with respect to the partial order. The witnesses $W_{\{1,2\}}$, $W_{\{1,3\}}$ and $W_{\{2,3\}}$ are obtained by minimizing over the sub-witnesses W_1, W_2 and W_3 respectively. At the top of the partial order is the witness $W_{\{1,2,3\}}$ obtained by minimizing over all three.

and $W_1(\rho, \sigma) \geq 0$ implies that $W_2(\rho, \sigma) \geq 0$ for all states ρ and σ, then we say that $W_1 \succeq W_2$. Given a family of go witnesses $(W_i)_{i \in \mathcal{I}}$, a new go witness

$$W_{\mathcal{I}}(\rho, \sigma) := \max_{i \in \mathcal{I}} W_i(\rho, \sigma)$$

can be constructed such that $W_{\mathcal{I}} \succeq W_i$ for each conversion witness. Additionally, we have $\overline{W} \succeq W$ for any complete witness $\overline{W}$ and any go witness W.

Chapter 3

Convex analysis

This chapter presents the notions from convex analysis that will be used in the remainder of the thesis. Only finite-dimensional analysis will be of interest, and all of the sets considered here will be contained in finite-dimensional real inner product spaces.

3.1 Convex sets and functions

We first present a few important definitions. A (*real*) *euclidean space* $\mathcal{V}$ is a finite-dimensional real inner product space, where the inner product is denoted $\langle\cdot,\cdot\rangle\colon \mathcal{V}\times\mathcal{V}\to\mathbb{R}$ and is bilinear. The norm derived from the inner product is $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. For $\epsilon > 0$ and a point $x \in \mathcal{V}$, the ϵ-ball at x is denoted $B_\epsilon(x) = \{y \in \mathcal{V} \mid \|x-y\| < \epsilon\}$. The set of non-negative real numbers is denoted $\mathbb{R}_+ = [0, +\infty)$.

Definition 3.1. Let $\mathcal{V}$ be a real euclidean space and let $\mathcal{X} \subseteq \mathcal{V}$ be a set.

1. The *affine hull* of $\mathcal{X}$, the set of all affine combinations of elements in $\mathcal{X}$, is denoted

$$\operatorname{aff}(\mathcal{X}) := \Big\{ \sum_{i=1}^{n} t_i x_i \,\Big|\, n \in \mathbb{N},\, x_1,\ldots,x_n \in \mathcal{X}, \text{ and } t_1,\ldots,t_n \in \mathbb{R} \text{ such that } \sum_{i=1}^{n} t_i = 1 \Big\}.$$

2. The *convex hull* of $\mathcal{X}$, the set of all convex combinations of elements in $\mathcal{X}$, is denoted

$$\operatorname{conv}(\mathcal{X}) := \Big\{ \sum_{i=1}^{n} t_i x_i \,\Big|\, n \in \mathbb{N},\, x_1, \ldots, x_n \in \mathcal{X}, \text{ and } t_1, \ldots, t_n \in \mathbb{R}_+ \text{ such that } \sum_{i=1}^{n} t_i = 1 \Big\}.$$

3. The *closure* of $\mathcal{X}$ is

$$\operatorname{cl}(\mathcal{X}) := \{ y \in \mathcal{V} \,|\, \forall \epsilon > 0,\, \exists x \in \mathcal{X} \text{ such that } \|y - x\| < \epsilon \}$$

4. The *interior* of $\mathcal{X}$ is

$$\operatorname{int}(\mathcal{X}) := \{ x \in \operatorname{aff}(\mathcal{X}) \,|\, \exists \epsilon > 0 \text{ such that } B_\epsilon(x) \subseteq \mathcal{X} \}.$$

5. The *relative interior* of $\mathcal{X}$ is

$$\operatorname{relint}(\mathcal{X}) := \{ x \in \operatorname{aff}(\mathcal{X}) \,|\, \exists \epsilon > 0 \text{ such that } B_\epsilon(x) \cap \operatorname{aff}(\mathcal{X}) \subseteq \mathcal{X} \}.$$

6. The (*relative*) *boundary* of $\mathcal{X}$ is $\operatorname{bd}(\mathcal{X}) = \operatorname{cl}(\mathcal{X}) \setminus \operatorname{relint}(\mathcal{X})$.

Note the distinction between the standard notions of the topological interior and boundary of a set and the notions of relative interior and boundary. The concept of *relative interior* is more useful for analyzing optimization problems over convex sets; it is the interior of $\mathcal{X}$ regarded as a subset of $\operatorname{aff}(\mathcal{X})$. The relative interior $\operatorname{relint}(\mathcal{X})$ consists of the points in the affine hull of $\mathcal{X}$ for which there exists an $\epsilon > 0$ such that $y \in \mathcal{X}$ for all $y \in \operatorname{aff}(\mathcal{X})$ and $\|x - y\| \leq \epsilon$. If $\operatorname{aff}(\mathcal{X}) = \mathcal{V}$, the interior coincides with the relative interior.

A set is $\mathcal{X} \subset \mathcal{V}$ is *convex* if $tx + (1-t)y \in \mathcal{X}$ for all $x, y \in \mathcal{X}$ and all $t \in (0, 1)$, or equivalently if $\operatorname{conv}(\mathcal{X}) = \mathcal{X}$. Given two subsets $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{V}$, their *Minkowski sum* is the set

$$\mathcal{X} + \mathcal{Y} := \{ x + y \,|\, x \in \mathcal{X} \text{ and } y \in \mathcal{Y} \}.$$

If $\mathcal{X}$ and $\mathcal{Y}$ are convex, then so is their sum. The *orthogonal complement* of $\mathcal{X}$ is the set $\mathcal{X}^\perp := \{y \in \mathcal{V} \,|\, \langle y, x\rangle = 0\}$.

3.1.1 Convex functions

Let $\mathcal{V}$ be a real euclidean space and let $\mathcal{C} \subseteq \mathcal{V}$ be a convex set. A function $f : \mathcal{C} \to \mathbb{R}$ is *convex* if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{3.1}$$

holds for all $x, y \in \mathcal{C}$ and all real numbers $t \in [0, 1]$, and is *concave* if the function $-f$ is convex. We define the set of extended real numbers as $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ such that any function $f : \mathcal{C} \to \mathbb{R}$ can be extended to a function $f : \mathcal{V} \to \overline{\mathbb{R}}$ by defining $f(x) = +\infty$ if $x \notin \mathcal{C}$. The *domain* of an extended real-valued function $f : \mathcal{V} \to \overline{\mathbb{R}}$ is the set

$$\operatorname{dom}(f) = \{x \in \mathcal{V} \,|\, f(x) \neq +\infty\}.$$

For the remainder of this thesis, by a "convex function" on a real euclidean space $\mathcal{V}$ we shall always mean a "convex function with possibly infinite values which is defined on all of $\mathcal{V}$". This approach has the advantage that technical nuisances about domains can be almost entirely suppressed, and it is the approach taken by most textbooks on convex analysis (see, e.g., [Roc70, BL06, BV04]). However, this approach leads to arithmetic calculations involving $+\infty$ for which new rules must be adopted. These rules are the obvious ones:

- For all $x \in \mathbb{R}$, $-\infty < x < +\infty$.
- $x + (+\infty) = +\infty$ for all $x \in \mathbb{R} \cup \{+\infty\}$.
- $x + (-\infty) = -\infty$ for all $x \in \mathbb{R} \cup \{-\infty\}$.
- $x \cdot (+\infty) = +\infty$ for all $x > 0$, and $x \cdot (+\infty) = -\infty$ for all $x < 0$.
- $0 \cdot (+\infty) = 0$ and $0 \cdot (-\infty) = 0$.

- $-(-\infty) = +\infty$ and $-(+\infty) = -\infty$.

Note that combinations of the form $(+\infty) + (-\infty)$ are left undefined and will be avoided. Then we can say that an extended function $f : \mathcal{V} \to \overline{\mathbb{R}}$ is convex if $\operatorname{dom}(f) \subseteq \mathcal{V}$ is a convex set and (3.1) is satisfied for all $x, y \in \operatorname{dom}(f)$.

Finally, the *epigraph* of a function $f : \mathcal{V} \to \mathbb{R}$ is a subset $\operatorname{epi}(f) \subseteq \mathcal{V} \times \mathbb{R}$ defined by

$$\operatorname{epi}(f) = \{(x,t) \mid f(x) \leq t\}.$$

The function f is convex if and only if its epigraph is convex as a subset of $\mathcal{V} \times \mathbb{R}$.

3.1.2 Directional and Fréchet derivatives

Let $\mathcal{V}$ and $\mathcal{W}$ be real euclidean spaces, let $\mathcal{C} \subset V$ be a set, and let $f : \mathcal{C} \to \mathcal{W}$ be a function. Given $x, y \in \mathcal{C}$, the directional derivative of f at x in the direction of y will be denoted

$$f'(x; y) := \lim_{t \to 0^+} \frac{f(x + ty) - f(x)}{t}.$$

If $\mathcal{C}$ is convex and f is a convex function, then the directional derivatives always exist at every $x \in \mathcal{C}$ and for every $y \in \mathcal{V}$ (although the value of the derivative might be $\pm\infty$).

The function f is said to be *Fréchet differentiable* at $x \in \mathcal{C}$ if there is a linear operator $\mathfrak{D}_{f,x} : \mathcal{V} \to \mathcal{W}$ such that

$$f'(x; y) = \mathfrak{D}_{f,x}(y)$$

for all $y \in \mathcal{V}$. Equivalently, we say f is Fréchet differentiable at x if there is a linear operator $\mathfrak{D}_{f,x}$ such that

$$\lim_{h \to 0} \frac{\|f(x + h) - f(h) - \mathfrak{D}_{f,x}(h)\|}{\|h\|} = 0.$$

If f is Fréchet differentiable at x, the operator $\mathfrak{D}_{f,x}$ is called the *Fréchet derivative* of f at x. In particular, if f is Fréchet differentiable at x then there must be an open set around

x on which f is finite, so x must be on the interior of $\operatorname{dom}(f)$. For $x \in \operatorname{bd}(\operatorname{dom}(f))$ on the boundary of the domain, the directional derivatives $f'(x;y)$ will all exist but f will not be Fréchet differentiable. However in most cases we can still approximate the derivative with a linear operator.

In the case where $f : \mathcal{V} \to \mathbb{R}$ is a functional from $\mathcal{V}$ to the real numbers, then the Fréchet derivative of f coincides with the gradient (if it exists). In this case,

$$\mathfrak{D}_{f,x}(y) = \langle \nabla f(x), y \rangle$$

where $\nabla f(x)$ is the standard notion of the gradient of f at x.

3.1.3 Conditions for optimization

In this section, we state the necessary and sufficient conditions for a convex function to be minimized over a convex, compact subset.

Let $\mathcal{V}$ be a real euclidean space. A *convex optimization problem* is a pair $(f, \mathcal{C})$, where $\mathcal{C} \subset \mathcal{V}$ is a (nonempty) convex, compact subset and $f : \mathcal{V} \to \overline{\mathbb{R}}$ is a convex function with $\operatorname{dom}(f) = \mathcal{C}$. The associated optimization problem is to compute

$$\inf\{f(x') \,|\, x' \in \mathcal{V}\}. \tag{3.2}$$

Since the domain of f is convex and compact, a (not necessarily unique) optimal point $x \in \operatorname{dom}(f)$ exists such that $f(x) \leq f(x')$ for all $x' \in \operatorname{dom}(f)$. Furthermore, since f is a convex function, if a point x is a local minimum then it must also be a global minimum. Theorem 3.2 formalizes this concept. See for example [BL06, Proposition 17.3] and [Roc70].

Theorem 3.2. *Let $\mathcal{V}$ be a real euclidean space, let $\mathcal{C} \subset \mathcal{V}$ a convex compact subset, and let $f : \mathcal{V} \to \overline{\mathbb{R}}$ be a convex function. An element $x \in \mathcal{C}$ minimizes f over $\mathcal{C}$, i.e. $f(x) =$*

$\min_{x' \in \mathcal{C}} f(x')$, *if and only if*

$$f'(x; x' - x) \geq 0 \tag{3.3}$$

holds for all $x \in \mathcal{C}$.

In particular, Theorem 3.2 states that a point $x \in \mathcal{C}$ is a minimum if the directional derivative along any direction in $\mathcal{C}$ is nonnegative (i.e., the point x is a local minimum). If f is Fréchet differentiable, this directional derivative condition can be given in terms of the Fréchet derivative operator, as shown in Corollary 3.3.

Corollary 3.3. *Let $\mathcal{V}$ be a real euclidean space, let $\mathcal{C} \subset \mathcal{V}$ a convex compact subset, and let $f : \mathcal{V} \to \overline{\mathbb{R}}$ be a convex function. If f is Fréchet differentiable at $x \in \mathcal{C}$, then x minimizes f over $\mathcal{C}$ if and only if*

$$\mathfrak{D}_{f,x}(x') \geq \mathfrak{D}_{f,x}(x) \tag{3.4}$$

holds for all $x' \in \mathcal{C}$, where $\mathfrak{D}_{f,x}$ is the Fréchet derivative of f at x.

This follows directly from Theorem 3.2 since $f'(x; x' - x) = \mathfrak{D}_{f,x}(x - x')$ and $\mathfrak{D}_{f,x}$ is linear. Indeed, if f is Fréchet differentiable at x then it is differentiable where the Fréchet derivative coincides with the gradient $\mathfrak{D}_{f,x}(x') = \langle \nabla f(x), x' \rangle$. Hence (3.4) is equivalent to

$$\langle \nabla f(x), x' \rangle \geq \langle \nabla f(x), x \rangle,$$

i.e., $\nabla f(x)$ defines a supporting hyperplane of $\mathcal{C}$ at $x \in \mathcal{C}$.

3.2 Matrix-valued functions

Recall that H_d denotes the real euclidean space of $d \times d$ hermitian matrices. For real numbers $a < b$, we let $\mathrm{H}_d(a, b)$ denote the subset of hermitian matrices whose eigenvalues are all contained in the interval (a, b). For example, with this notation the cone of positive definite matrices can be denoted $\mathrm{H}_{d,++} = \mathrm{H}_d(0, \infty)$.

Any function $f : (a, b) \to \mathbb{R}$ of real numbers can be extended to a function on hermitian matrices $\mathrm{H}_d(a, b) \to \mathrm{H}_d$. We now discuss the notion of convexity that is important for real-valued functions that are extended to matrices.

Definition 3.4. A function $f : (a, b) \to \mathbb{R}$ is said to be *operator monotone* if, for all integers d and all $A, B \in \mathrm{H}_d(a, b)$, it holds that

$$A \geq B \quad \Rightarrow \quad f(A) \geq f(B), \tag{3.5}$$

and f is said to be *operator convex* if for all integers d, for all $A, B \in \mathrm{H}_d(a, b)$, and for all $t \in (0, 1)$ it holds that

$$f(tA + (1-t)B) \leq tf(A) + (1-t)f(B). \tag{3.6}$$

We say that f is *operator concave* if $-f$ is operator convex.

Note that any operator monotonic function $f : (a, b) \to \mathbb{R}$ must be monotonic as a function of real numbers. The same holds for operator convexity and operator concavity. Most functions that will be considered here will be defined on the positive real numbers $f : (0, \infty) \to \mathbb{R}$. There are a great many examples of operator convex functions functions that are useful for studying quantum information. The following well known result lists the ones that are of predominant use.

Theorem 3.5 (Löwener-Heinz Theorem)**.** *Consider the following functions $f : (0, \infty) \to \mathbb{R}$.*

i) For $p \in [-1, 0]$, the function $f(t) = t^p$ is operator monotone and operator convex.

ii) For $p \in [0, 1]$, the function $f(t) = t^p$ is operator monotone and operator concave.

iii) For $p \in [1, 2]$, the function $f(t) = t^p$ is operator convex.

iv) The function $f(t) = \log t$ is operator concave while $f(t) = t \log t$ is operator convex.

For example, that log is operator concave can be derived from the operator concavity t^p for $p \in [0,1]$, since $\log A = \lim_{p\to 0^+} \frac{A^p - I}{p}$ and concavity is preserved in the limit. There are also stronger notions of operator monotonicity and convexity that will also be considered in this thesis. We first introduce the following notation. For $A \in \mathrm{H}_d$, we write $A \gneq 0$ whenever $A \geq 0$ and $A \neq 0$. Similarly, we write $A \gneq B$ if $A - B \gneq 0$.

Definition 3.6. Let $f : (a,b) \to \mathbb{R}$. Then f is said to be

1. *strict operator monotone* if $A > B$ implies $f(A) > f(B)$ for all $A, B \in \mathrm{H}_d(a,b)$.

2. *strong operator monotone* if $A \gneq B$ implies $f(A) \gneq f(B)$ for all $A, B \in \mathrm{H}_d(a,b)$.

3. *strict operator convex* if, for all $A, B \in \mathrm{H}_d(a,b)$ with $\mathrm{rank}(A - B) = d$ and all $t \in (0,1)$, it holds that
$$f(tA + (1-t)B) < tf(A) + (1-t)f(B).$$

4. *strong operator convex* if, for all $A, B \in \mathrm{H}_d(a,b)$ with $A \neq B$ and all $t \in (0,1)$, it holds that
$$f(tA + (1-t)B) \lneq tf(A) + (1-t)f(B)$$

The functions $f(t) = -t^p$ for $p \in (0,1]$ and $f(t) = -\log t$ are not only strictly monotone and strictly convex on the interval $(0,+\infty)$ [HJ94], but they are strongly convex as well [FGR11]. This fact is important for proving the strict convexity of the functions on positive operators of the form $X \mapsto \mathrm{Tr}(Af(X))$ presented in Chapter 4, where $A > 0$ is a positive operator and the function $f : (0,+\infty) \to \mathbb{R}$ is strongly operator convex.

3.3 Separating and supporting hyperplanes

This section presents the version of the well known separating hyperplane theorem from convex analysis that will be used in this thesis.

Theorem 3.7 (Hyperplane separation)**.** *Let $\mathcal{C} \subseteq \mathcal{V}$ be a closed convex subset of a real euclidean space $\mathcal{V}$, and let $b \in \mathcal{V}$ such that $b \notin \mathcal{C}$. There exists a nonzero $v \in \mathcal{V}$ and a real number c such that*

$$\langle v, b\rangle < c \leq \langle v, x\rangle$$

for all $x \in \mathcal{C}$. If $\mathcal{C}$ is also a cone, we may take $c = 0$.

A *hyperplane* in $\mathcal{V}$ is an affine set $\{x \in \mathcal{V} \,|\, \langle v, x\rangle = c\}$ that is defined by some $v \in \mathcal{V}$ and $c \in \mathbb{R}$. The interpretation of Theorem 3.7 is that v defines a hyperplane that *separates* the vector b from the set $\mathcal{C}$. The vector v is also said to *witness* for $\mathcal{C}$, since it indicates the non-membership of b in $\mathcal{C}$. As the following corollary shows, hyperplanes can also be used to give a condition for when a point is on the boundary of a convex set.

Corollary 3.8. *Let $\mathcal{C} \subseteq \mathcal{V}$ be a convex subset of a real euclidean space $\mathcal{V}$, and let $x \in \mathcal{C}$. Then x is on the boundary of $\mathcal{C}$ if and only if there exists a nonzero $v \in \mathcal{V}$ such that*

$$\langle v, x'\rangle \geq \langle v, x\rangle \tag{3.7}$$

for all $x' \in \mathcal{C}$.

Given a point x on the boundary of a convex set $\mathcal{C} \subset \mathcal{V}$, a vector $v \in \mathcal{V}$ is said to define a *supporting hyperplane* of $\mathcal{C}$ at x if it satisfies (3.7) for all $x' \in \mathcal{C}$. The corresponding supporting hyperplane is the affine set $\{x' \in \mathcal{V} \,|\, \langle v, x'\rangle = c\}$, where $c = \langle v, x\rangle$. That is, a vector v defines a supporting hyperplane of $\mathcal{C}$ if there is a $c \in \mathbb{R}$ such that

$$\langle v, x'\rangle \geq c \qquad \text{for all } x' \in \mathcal{C}$$

and there exists an $x \in \mathcal{C}$ such that $\langle v, x\rangle = c$. In the special case when $c = 0$, we will say that v is a *witness* for non-membership in $\mathcal{C}$. If v is a witness for $\mathcal{C}$ and $\langle v, y\rangle < 0$ for some $y \in \mathcal{V}$, then $y \notin \mathcal{C}$.

3.4 Cones and conic programming

This section will introduce notation and states a few necessary definitions and properties regarding cones in finite-dimensional real inner product spaces. The concept of general

3.4.1 Cones

Definition 3.9. Let $\mathcal{V}$ be a real euclidean space.

(i) A subset $\mathcal{K} \subset \mathcal{V}$ is said to be a *cone* if $tx \in \mathcal{K}$ for all $x \in \mathcal{K}$ and all positive real numbers $t > 0$.

(ii) Let $\mathcal{X} \subseteq \mathcal{V}$ be any subset. The (*convex*) *conical hull* of $\mathcal{X}$ is the set

$$\operatorname{cone}(\mathcal{X}) := \left\{ \sum_{i=1}^{n} t_i x_i \,\middle|\, n \in \mathbb{N},\, x_1, \ldots, x_n \in \mathcal{X}, \text{ and } t_1, \ldots, t_n \in \mathbb{R}_+ \right\}$$

(iii) Let $\mathcal{X} \subseteq \mathcal{V}$ be any subset. The *dual cone* to $\mathcal{X}$ is the set

$$\mathcal{X}^* := \{ y \in \mathcal{V} \,|\, \forall x \in \mathcal{X},\, \langle y, x \rangle \geq 0 \},$$

and a cone is said to be *self-dual* if $\mathcal{K}^* = \mathcal{K}$.

We now state a few useful facts about cones. Note that $\mathcal{X}^*$ is a closed convex cone for any subset $\mathcal{X} \subseteq \mathcal{V}$. If $\mathcal{K}$ is a closed and convex cone, then $\mathcal{K}^{**} = \mathcal{K}$. If $\mathcal{K}$ and $\mathcal{K}'$ are closed and convex cones in $\mathcal{V}$, their Minkowski sum $\mathcal{K} + \mathcal{K}'$ is also a closed and convex cone. Given two closed and convex cones $\mathcal{K} \subseteq \mathcal{V}$ and $\mathcal{K}' \subseteq \mathcal{V}'$ in two real euclidean spaces, the direct sum of the cones

$$\mathcal{K} \oplus \mathcal{K}' \subseteq \mathcal{V} \oplus \mathcal{V}'$$

is also a closed and convex cone. The dual cone of the direct sum of cones is the direct sum of the duals, i.e.,

$$(\mathcal{K} \oplus \mathcal{K}')^* = \mathcal{K}^* \oplus \mathcal{K}'^*.$$

For any cones $\mathcal{K}$ and $\mathcal{K}'$, if $\mathcal{K} \subseteq \mathcal{K}'$, then $\mathcal{K}'^* \subseteq \mathcal{K}^*$. Finally, if $\mathcal{K}$ and $\mathcal{K}'$ are closed and convex, it holds that

$$(\mathcal{K} + \mathcal{K}')^* = \mathcal{K}^* \cap \mathcal{K}'^* \qquad \text{and} \qquad (\mathcal{K} \cap \mathcal{K}')^* = \mathcal{K}^* + \mathcal{K}'^*.$$

The interior of a cone will be important for exploring duality properties of convex cones in the next section. Here we state a useful characterization of the interior of closed convex cones.

Proposition 3.10 ([1])**.** *Let $\mathcal{V}$ be a real euclidean space and let $\mathcal{K} \subseteq \mathcal{V}$ be a closed convex cone. It holds that*

$$\operatorname{int}(\mathcal{K}) = \{x \in \mathcal{V} \mid \langle x, y \rangle > 0 \ \forall y \in \mathcal{K}^* \, s.t. \, y \neq 0\}. \tag{3.8}$$

One of the most common cones studied in convex analysis is the cone of positive semidefinite operators $\mathrm{H}(\mathcal{H})_+$ on a finite-dimensional Hilbert space. As before, if $\mathcal{H} = \mathbb{C}^d$ we can identify this cone with the space of $d \times d$ positive semidefinite matrices $\mathrm{H}(\mathbb{C}^d)_+ \simeq \mathrm{H}_{d,+}$. This cone is closed and convex and is self-dual, i.e., $(\mathrm{H}(\mathcal{H})_+)^* = \mathrm{H}(\mathcal{H})_+$. The interior of this cone is the cone of positive definite operators, i.e.,

$$\operatorname{int}(\mathrm{H}(\mathcal{H})_+) = \mathrm{H}(\mathcal{H})_{++} = \{X \in \mathrm{H}(\mathcal{H}) \mid X > 0\}.$$

3.4.2 Conic programming

Conic programming is a useful tool in convex analysis. A conic program expresses the optimization of a linear function over the intersection of an affine subspace and a closed convex cone in a finite-dimensional real inner product space. For the purposes of this thesis, it is sufficient to consider only cone programs defined over spaces of Hermitian operators. Given

[1] see Exercise 2.31(d) in[BV04]

a finite-dimensional Hilbert space $\mathcal{H}$ (with dimension d), the space of hermitian operators on this space $\mathrm{H}(\mathcal{H})$ is a finite-dimensional real inner product space (with dimension d^2) together with the Hilbert-Schmidt inner product.

Let $\mathcal{H}$ and $\mathcal{H}'$ be two finite dimensional Hilbert spaces, and let $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H})$ be a cone. A *conic program* is an ordered set denoted $(\Phi, A, B, \mathcal{K})$, where $A \in \mathrm{H}(\mathcal{H})$ and $B \in \mathrm{H}(\mathcal{H}')$ are operators, and $\Phi : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H}')$ is a hermiticity-preserving linear map. A conic program has the following associated pair of optimization problems:

$$\begin{array}{rlcrl} \multicolumn{2}{c}{\underline{\text{Primal problem}}} & & \multicolumn{2}{c}{\underline{\text{Dual problem}}} \\ \text{maximize:} & \langle A, X\rangle & \qquad & \text{minimize:} & \langle B, Y\rangle \\ \text{subject to:} & X \in \mathcal{K} & & \text{subject to:} & \Phi^*(Y) - A \in \mathcal{K}^* \\ & \Phi(X) = B & & & Y \in \mathrm{H}(\mathcal{H}'), \end{array} \tag{3.9}$$

These are the *primal* and *dual* problems respectively. The *primal feasible set* $\mathcal{A} \subset \mathrm{H}(\mathcal{H})$ and the *dual feasible set* $\mathcal{B} \subset \mathrm{H}(\mathcal{H}')$ of the problem are defined as

$$\begin{aligned} \mathcal{A} &:= \{X \in \mathcal{K} \mid \Phi(X) = B\} \\ \mathcal{B} &:= \{Y \in \mathrm{H}(\mathcal{H}') \mid \Phi^*(Y) - A \in \mathcal{K}^*\}. \end{aligned} \tag{3.10}$$

Elements $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ are said to be *primal feasible* and *dual feasible* respectively, and the conic program is said to be (*primal*) *feasible* if $\mathcal{A} \neq \emptyset$ and said to be *dual feasible* if $\mathcal{B} \neq \emptyset$. The *primal optimal* and *dual optimal* values are given by

$$\begin{aligned} \alpha &:= \sup\{\langle A, X\rangle \mid X \in \mathcal{A}\} \\ \beta &:= \inf\{\langle B, Y\rangle \mid Y \in \mathcal{B}\}. \end{aligned} \tag{3.11}$$

For any conic program, *weak duality* always holds. That is, the primal optimal value and the dual optimal value always satisfy $\alpha \leq \beta$. Indeed, if X and Y are primal feasible and dual feasible respectively, then $\langle A, X\rangle \leq \langle B, Y\rangle$ since

$$\begin{aligned}\langle B, Y\rangle = \langle \Phi(X), Y\rangle &= \langle X, \Phi^*(Y)\rangle \\ &= \underbrace{\langle X, \Phi^*(Y) - A\rangle}_{\geq 0} + \langle X, A\rangle \geq \langle X, A\rangle,\end{aligned}$$

where we use the fact that $\langle X, \Phi^*(Y) - A\rangle \geq 0$ since $X \in \mathcal{K}$ and $\Phi^*(Y) - A \in \mathcal{K}^*$. The form of the conic programs stated in (3.9) are not in the standard form as it appears in most literature on the subject, but this is nonetheless the most useful formation for applications in quantum information (see, e.g., [Wat16]).

The primal optimum and dual optimum of a conic program do not always agree. But for many conic programs that arise in applications (especially in quantum information theory[2]), the primal optimum and dual optimum will be equal. This situation is called *strong duality*. The following theorem provides a set of conditions for which strong duality is guaranteed to hold. Let $(\Phi, A, B, \mathcal{K})$ be a conic program. The associated primal problem is said to be *strictly feasible* if there exists a primal feasible X in the interior of the cone (i.e., if $\mathcal{A} \cap \text{int}(\mathcal{K}) \neq \emptyset$). Analogously, the dual problem is strictly feasible if there exists a dual feasible Y such that $\Phi^*(Y) - A$ is in the interior of the dual cone $\mathcal{K}^*$ (i.e., if $(\Phi^*(\mathcal{B}) - A) \cap \text{int}(\mathcal{K}^*) \neq \emptyset$).

Theorem 3.11. *(Strong Duality Theorem for Conic Programming[3]) Let $\mathcal{K} \subseteq \mathcal{V}$ be a cone and let $(\Phi, A, B, \mathcal{K})$ be a conic program. Suppose that the primal problem and dual problems are both feasible. If either the primal or dual problem is also strictly feasible, then $\alpha = \beta$.*

The most common type of conic program that appears while studying quantum information theory are *semidefinite programs*, in which the cone of interest is the cone of positive semidefinite operators $\mathcal{K} = \text{H}(\mathcal{H})_+ \subseteq \text{H}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$. Analogous to the notation for conic programs, a semidefinite program will often be given as a triple (Φ, A, B) with

[2] In particular, all of the conic programs considered in this thesis have strong duality.

[3] The more general version of Theorem 3.11 is originally due to Slater [Sla50]. See also [LMT15] and [GM12, Theorem 4.7.1] for the conic programming version used here.

associated optimization problems

$$
\begin{array}{rlcrl}
\multicolumn{2}{c}{\underline{\text{Primal problem}}} & \qquad & \multicolumn{2}{c}{\underline{\text{Dual problem}}} \\[1ex]
\text{maximize:} & \langle A, X\rangle & & \text{minimize:} & \langle B, Y\rangle \\[1ex]
\text{subject to:} & X \in \mathrm{H}(\mathcal{H})_+ & & \text{subject to:} & \Phi^*(Y) \geq A \\[1ex]
 & \Phi(X) = B & & & Y \in \mathrm{H}(\mathcal{H}'),
\end{array}
\tag{3.12}
$$

where $\Phi^*(Y) \geq A$ means that $\Phi^*(Y) - A \in \mathrm{H}(\mathcal{H})_+$. Semidefinite programs have the additional benefit that there are efficient algorithms for solving typical semidefinite programs in both theory and practice [VB96]. The `cvx` package [**?**] for MATLAB, for example, allows small-scale semidefinite programs to be solved efficiently in most applications. More generally, any conic program can be solved efficiently in practice (via interior-point methods) as long as an efficiently computable self-concordant barrier function for the cone $\mathcal{K}$ is available [NN94].

3.4.3 Conic version of Farkas' Lemma

Theorems of alternatives are common in convex analysis. Such theorems give conditions on pairs of problems under which exactly one of the problems is feasible, but not both. One such well known theorem of alternatives is Farkas' Lemma for linear programming [Far02, BV04, DJ14]. Here, we state the conic version of Farkas' Lemma. This is a standard result that can be found in many textbooks on convex analysis (see e.g. [LY08]), but because of the nonstandard notation used in this thesis, a complete proof is provided in Appendix B for completeness.

Theorem 3.12 (Farkas' Lemma for conic programs)**.** *Let $\mathcal{H}$ and $\mathcal{H}'$ be finite-dimensional Hilbert spaces, let $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H})$ be a closed convex cone, let $\Phi : \mathrm{H}(\mathcal{H}) \to \mathrm{H}(\mathcal{H}')$ be a linear map, and let $B \in \mathrm{H}(\mathcal{H}')$. Suppose further that $\Phi(\mathcal{K})$ is closed. Then* ***exactly one*** *of the* *following statements holds:*

(i) There exists $X \in \mathcal{K}$ such that $\Phi(X) = B$.

(ii) There exists $Y \in \mathrm{H}(\mathcal{H}')$ such that $\Phi^*(Y) \in \mathcal{K}^*$ and $\langle Y, B\rangle < 0$.

That both statements (i) and (ii) of 3.12 cannot hold simultaneously is clear, since if $\Phi(X) = B$ and $\Phi^*(Y) \in \mathcal{K}^*$ both hold, then

$$\begin{aligned}\langle Y, B\rangle &= \langle Y, \Phi(X)\rangle \\ &= \langle \Phi^*(Y), X\rangle \geq 0.\end{aligned}$$

The proof that at least one of them must always hold makes use of the separating hyperplane theorem (Theorem 3.7) and requires that $\Phi(\mathcal{K})$ be closed.

The conic version of Farkas' Lemma is essentially the following statement: For a conic program in which $A = 0$, if $\Phi(\mathcal{K})$ is closed, then the primal problem is feasible if and only if $\beta = 0$. This only holds, however, if the image of the cone $\Phi(\mathcal{K})$ is a closed convex cone. Since $\mathcal{K}$ is a convex cone and Φ is linear, its image will always be a convex cone. But closedness is not guaranteed. Lemma 3.13 below provides give a useful sufficient condition under which $\Phi(\mathcal{K})$ is closed. It is a standard result in convex analysis which can be found in many textbooks, but its proof is included in Appendix B for completeness due to our use of nonstandard notation. In particular, it states that if the dual problem is strictly feasible (i.e. if $\Phi^*(\mathrm{H}(\mathcal{H}')) \cap \mathrm{int}(\mathcal{K}^*) \neq \emptyset$), then $\Phi(\mathcal{K})$ is closed.

Lemma 3.13. *Let $\mathcal{H}$ and $\mathcal{H}'$ be finite-dimensional Hilbert spaces, let $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H})$ be a closed convex cone and let $\Phi : \mathrm{H}(\mathcal{H}) \to \mathrm{H}(\mathcal{H}')$ be a linear map. Suppose there exists $Z \in \mathrm{H}(\mathcal{H}')$ such that $\Phi^*(Z) \in \mathrm{int}(\mathcal{K}^*)$. Then $\Phi(\mathcal{K})$ is closed.*

The result of Lemma 3.13 will be used to allow us to employ Theorem 3.12 in analyzing convertibility of resources in convex resource theories in Chapter 6.

Chapter 4

Convex optimization problems in quantum information theory

There are numerous important quantities in quantum information are defined in terms of a convex optimization problem. In particular, entanglement is an important resource in quantum information [PV07, HHHH09] which can be quantified this way. Many other quantities in quantum information can be considered in terms of convex optimization problems. Such problems are usually given in terms of a convex function $f : \mathrm{H}_d \to \overline{\mathbb{R}}$ (where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$) and some convex subset $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ of positive semidefinite matrices. Given such a convex optimization problem, we can ask the question, "when does a matrix $\sigma \in \mathcal{C}$ minimize f?" That is, when does a matrix $\sigma \in \mathcal{C}$ satisfy

$$f(\sigma) = \min_{\sigma' \in \mathcal{C}} f(\sigma'),$$

assuming that $f(\sigma')$ is finite for at least one $\sigma' \in \mathcal{C}$? We will be most interested in the case when the optimal σ is on the boundary of $\mathcal{C}$. Since f is a convex function, it is sufficient to show that σ is a semi-local minimum of f. That is, all of the directional derivatives of f at σ are nonnegative. Since the directional derivative is usually a linear map, denoted by $\mathfrak{D}_{f,\sigma} : \mathrm{H}_d \to \mathbb{R}$, the condition in (3.4) reduces to the fact that $\mathfrak{D}_{f,\sigma}$ defines a supporting hyperplane of $\mathcal{C}$ at σ on the boundary of $\mathcal{C}$.

In general, finding a closed analytic formula for an optimal σ is difficult, if not impossible. From a computational point of view, however, the complexity of finding a good numerical approximation to the optimal σ can be relatively easy (i.e. polynomial in terms of computation of the function f and the membership in $\mathcal{C}$). Such numerical optimization for the relative entropy of entanglement, for example, has been studied in [ZFG10]. Rather than trying to directly solve these optimization problems, however, in this chapter, methods of solving the *converse* problem are primarily emphasized. That is, given a matrix $\sigma \in \mathcal{C}$, we instead ask "which functions f achieve their minimum over $\mathcal{C}$ at σ?" Although this may seem trivial at first, these kinds of results can yield meaningful statements about finding closed formulas for certain quantities in quantum information [FGR11].

Recent work has been done [Ish03, MI08, FGR11] that employs similar methods to determine an explicit expression for the relative entropy of entanglement for certain bipartite quantum states. Given a separable state $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$, one can find all of the entangled states ρ for which σ is the *closest* separable state, thus minimizing the relative entropy of entanglement. The work presented in this chapter is based on these previous analyses of the relative entropy of entanglement.

Here, a general framework for finding minimization criteria for convex optimization problems that arise in quantum information theory will be constructed by studying the converse problem. Previous results are extended to find an explicit expression for the Rains bound [Rai99b] (a quantity related to the relative entropy of entanglement) and it will be shown how these results can be generalized to other functions of interest in quantum information theory. Among other results, these methods will be used to show that the Rains bound is equal to the relative entropy of entanglement for states of bipartite systems for which at least one subsystem is a qubit. In addition to the relative entropy, we can also study other similar convex functions that arise in quantum information theory, including the Rényi relative entropies. As an application, this method will be used to compute the *Rényi*

α-relative entropies of entanglement for pure states for $\alpha \in [0, 2]$.

Lastly, a general algorithm for numerically estimating the solutions to some of these convex optimization problems in quantum information theory will also be constructed. This algorithm makes use of the "cutting plane" method for approximating the epigraph of a convex function and is implemented using the CVX modeling system for convex programs in MATLAB. The algorithm is based on the algorithm in [ZFG10] for estimating the relative entropy of entanglement (with respect to the PPT states). The algorithm is improved here and extended to compute Rényi relative entropies of entanglement as well as the Rains bound. To showcase the algorithms efficacy, numerical results for computing these quantities by implementing the algorithm are compared to states that have known values for the relative entropies of entanglement and the Rains bound.

The remainder of this chapter is outlined as follows. In Section 4.1, necessary and sufficient conditions needed for a matrix $\sigma \in \mathcal{C}$ to minimize a convex function f over an arbitrary convex subset $\mathcal{C} \subseteq \mathrm{H}_n$ are presented. The examples of this analysis applied to the relative entropy of entanglement and the Rains bound are given in Sections 4.3 and 4.4, respectively. In Section 4.4.3, these results are then used to prove some interesting facts about these two quantities, such as the fact that the Rains bound and the relative entropy of entanglement coincide for states in quantum systems in which at least one subsystem is a qubit. Analysis of the Rényi α-relative entropies of entanglement is presented in Section 4.5. The cutting plane method for numerical estimation and applications to the relative entropy of entanglement and the Rains bound are discussed in Section 4.6. Further applications to other convex functions are contained in Section 4.7.

4.1 Conditions for minimizing convex functions

In this chapter, we discuss minimizing convex functions over convex subsets of positive matrices. Let $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ be a convex set, let $X \in \mathcal{C}$, and let $f : \mathcal{C} \to \overline{\mathbb{R}}$ be a convex function.

From the results in Section 3.1.3, conditions for when X minimizes f can be given in terms of the directional derivatives of f. In particular, it holds that $f(X) \leq f(X')$ for all $X' \in \mathcal{C}$ if and only if the directional derivatives satisfy $f'(X; X' - X) \geq 0$ for all $X' \in \mathcal{C}$.

The analysis in this chapter will primarily be concerned with convex functions of positive matrices $f : \mathrm{H}_{d,+} \to \overline{\mathbb{R}}$ that are of the form

$$f(X) = h(\mathrm{Tr}(Ag(X))) \tag{4.1}$$

for some differentiable functions g and h of real numbers and some positive semidefinite matrix $A \in \mathrm{H}_{d,+}$, where g is well-defined on the interval $g : (0, +\infty) \to \mathbb{R}$. For example, given a fixed density matrix $\rho \in \mathrm{D}(\mathbb{C}^d)$, the relative entropy of ρ with respect to a positive operator X is given by $S(\rho \| X) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log(X))$. In this section, we will show how to compute directional derivatives of functions of the form defined in (4.1). Conditions for minimization of convex functions can be given in terms of the directional derivatives (see Section 3.1.3), and in this section it will be shown how these criteria can be given in terms of supporting hyperplanes. This will allow us to analyze minimization criteria for the relative entropy (as well the Rényi α-relative entropies) in addition to other similar quantities in quantum information.

Given a positive matrix $A \in \mathrm{H}_{d,+}$ and an analytic function $f : (0, \infty) \to \mathbb{R}$, we can consider the function on positive definite matrices $f_A : \mathrm{H}_{d,++} \to \mathbb{R}$ defined by

$$f_A(X) = \mathrm{Tr}(Af(X)). \tag{4.2}$$

If A and X have spectral decompositions given by $A = \sum_{i=1}^{d} a_i |v_i\rangle\langle v_i|$ and $X = \sum_{j=1}^{d} x_j |u_j\rangle\langle u_j|$, we can write this as

$$f_A(X) = \sum_{i,j=1}^{d} a_i f(x_j) |\langle v_i | u_j \rangle|^2.$$

As long as the limit $\lim_{t \to 0} f(t)$ exists (which could be $\pm\infty$), we can extend f to a function

$f : [0, +\infty) \to \overline{\mathbb{R}}$ by defining $f(0)$ as this limit. Then we can also extend f_A to semidefinite matrices in the following manner. If $f(0)$ is finite then $\operatorname{Tr}(Af(X))$ is always well defined and finite for all $X \geq 0$. If $f(0) = \pm\infty$, then $\operatorname{Tr}(Af(X))$ is finite if and only if $\operatorname{supp}(A) \subseteq \operatorname{supp}(X)$ and is $\pm\infty$ otherwise. For example, we define $\operatorname{Tr}(A \log(X)) = -\infty$ if $\operatorname{supp}(A) \not\subseteq \operatorname{supp}(X)$. The functions $f : (0, +\infty) \to \mathbb{R}$ considered here will always be either concave or convex, so the limiting value $f(0)$ will always exist.

The functions as defined in (4.2) are convex (concave) on $\mathrm{H}_{d,+}$ as long as f is operator convex (concave) on $(0, +\infty)$. Indeed, if

$$f(tX + (1-t)Y) \leq tf(X) + (1-t)f(Y)$$

holds as an operator inequality for $X, Y \in \mathrm{H}_{d,++}$ and $t \in [0, 1]$, then

$$\operatorname{Tr}(Af(tX + (1-t)Y)) \leq t \operatorname{Tr}(Af(X)) + (1-t) \operatorname{Tr}(Af(Y))$$

holds for all $A \geq 0$. This convexity is preserved for singular matrices $X \in \mathrm{H}_{d,+}$ by continuity. Many of these types of functions arise in quantum information theory, and finding conditions for minimizing such functions over convex sets is important. For example, analysis of the convex function $X \mapsto -\operatorname{Tr}(\rho \log(X))$ for a fixed ρ will allow us to determine conditions for when a state σ minimizes the relative entropy $S(\rho\|\sigma) = \operatorname{Tr}(\rho(\log\rho - \log\sigma))$ over all states in some convex set of density matrices $\mathcal{C}$.

As shown in Section 3.1.3, necessary and sufficient conditions for finding minima of convex functions can be given in terms of directional derivatives. Here, we show how to compute directional derivatives for functions of the form f_A. These functions are Fréchet differentiable at any positive definite X, so the conditions for minimization can be given in terms of linear functionals. Careful consideration must be taken in the case when $X \geq 0$ but $X \not> 0$ (i.e., when the matrix X has at least one zero eigenvalue).

4.1.1 Derivatives of matrix functions

The directional derivatives of f_A are given by $f'_A(X;Y) = \mathrm{Tr}(Af'(X;Y))$, as long as the directional derivative of matrices $f'(X;Y)$ exists. In particular if $f : (0,+\infty) \to \mathbb{R}$ and $X > 0$, then $f'(X;Y)$ is well-defined for any $Y \in \mathrm{H}_d$. However, if $A \neq 0$ then the directional derivative $f'_A(X;Y)$ can exist even if $f'(X;Y)$ does not. Here we show how to compute the directional derivatives $f'_A(X;Y)$ for all $X \in \mathrm{H}_{d,+}$ and $Y \in \mathrm{H}_d$. We first consider the case when $X > 0$. To do so, we first introduce some notation for differentiating functions of matrices. The case when $X \not> 0$ must be carefully considered separately.

Definition 4.1. Let $f : (0,+\infty) \to \mathbb{R}$ be differentiable, and let $x, y \in \mathbb{R}$ with $x, y > 0$. The *divided differences* of f are defined by

$$f^{[1]}(x,y) = \begin{cases} \dfrac{f(x) - f(y)}{x - y}, & x \neq y \\ f'(x), & x = y. \end{cases}$$

Given a diagonal matrix $\Lambda \in \mathrm{H}_{d,++}$ with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, the *matrix of divided differences of* Λ is denoted as $f^{[1]}(\Lambda)$ and has matrix elements given by

$$(f^{[1]}(\Lambda))_{i,j} = f^{[1]}(\lambda_i, \lambda_j).$$

The matrix of divided differences $f^{[1]}(\Lambda)$ is symmetric. We also make use of the *Schur* (or *entry-wise*) product of two $d \times d$ matrices. Given two $d \times d$ matrices $A, B \in \mathrm{H}_d$, this Schur product is denoted by $A \circ B$ which has matrix elements $(A \circ B)_{i,j} = A_{i,j}B_{i,j}$. For a fixed symmetric matrix $A \in \mathrm{H}_d$ (with $A^T = A$), Schur multiplication by A is a linear map that is self-adjoint with respect to the Hilbert-Schmidt inner product. That is, for all $X, Y \in \mathrm{H}_d$ it holds that[1]

$$\langle X, A \circ Y \rangle = \mathrm{Tr}(X(A \circ Y)) = \mathrm{Tr}((A \circ X)Y) = \langle A \circ X, Y \rangle$$

[1] In general, it holds that $\mathrm{Tr}(X(A \circ Y)) = \mathrm{Tr}((A^T \circ X)Y)$ for arbitrary matrices A, X, and Y (see e.g. Lemma 5.4 of [HJ94]). Here we reqire that $A^T = A$.

and the map $X \mapsto A \circ X$ is linear in X.

The following proposition tells us how to compute the derivatives of matrix functions. Proposition 4.2 shows that matrix functions derived from real-analytic functions are Fréchet differentiable, with the Fréchet derivative given in terms of the matrix of divided differences. A proof of this fact can be found in many textbooks on matrix analysis (see e.g., Lemma V.3.1 in [Bha97] or Theorem 6.6.3 in [HJ94]).

Proposition 4.2. *Let $f : (0, \infty) \to \mathbb{R}$ be analytic. For every diagonal matrix $\Lambda \in \mathrm{H}_{d,+}$ and every hermitian matrix $Y \in \mathrm{H}_d$, it holds that*

$$f'(\Lambda; Y) = f^{[1]}(\Lambda) \circ Y.$$

For an arbitrary $X \in \mathrm{H}_{d,++}$, let $X = U\Lambda U^\dagger$ be a spectral decomposition such that U is unitary and Λ is diagonal. Then

$$f'(X; Y) = U\Big(f^{[1]}(U^\dagger X U) \circ (U^\dagger Y U)\Big)U^\dagger.$$

We see that the directional derivative is linear in Y, so f is Fréchet differentiable at X with its Fréchet derivative given by $\mathfrak{D}_{f,X}(Y) = f^{[1]}(\Lambda) \circ Y$ if X is diagonal and

$$\begin{aligned}\mathfrak{D}_{f,X}(Y) &= U\Big(f^{[1]}(U^\dagger X U) \circ (U^\dagger Y U)\Big)U^\dagger \\ &= U\Big(\mathfrak{D}_{f,U^\dagger X U}(U^\dagger Y U)\Big)U^\dagger\end{aligned}$$

otherwise, where $\mathfrak{D}_{f,X}$ is a linear map of $d \times d$ matrices. For most of the remainder of this chapter, we may assume for simplicity (and without loss of generality) that X diagonal, and we may write the directional derivatives as $f'(X; Y) = \mathfrak{D}_{f,X}(Y) = f^{[1]}(X) \circ Y$. The following proposition shows how to use this fact to write the directional derivatives of functions of the form f_A for positive definite X. It was first considered in [FGR11], where it was shown in the case when f is the function $f(x) = \log(x)$.

Proposition 4.3. *Let $f : (0, \infty) \to \mathbb{R}$ be analytic, let $A \in \mathrm{H}_{d,+}$, and let $X \in \mathrm{H}_{d,++}$. For all $Y \in \mathrm{H}_d$, it holds that*

$$f'_A(X; Y) = \langle \mathfrak{D}_{f,X}(A), Y \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the Hilbert-Schmidt inner product.

Proof. Since f_A is defined as $f_A(X) = \mathrm{Tr}(Af(X))$, it is easy to see from the chain rule that $f'_A(X; Y) = \mathrm{Tr}(Af'(X; Y))$. We may assume without loss of generality that X is diagonal, hence $f'(X; Y) = \mathfrak{D}_{f,X}(Y) = f^{[1]}(X) \circ Y$. Note that the matrix $f^{[1]}(X)$ of divided differences is symmetric, and thus

$$\begin{aligned} f'_A(X; Y) &= \mathrm{Tr}(A\mathfrak{D}_{f,X}(Y)) \\ &= \langle A, \mathfrak{D}_{f,X}(Y) \rangle \\ &= \langle \mathfrak{D}_{f,X}(A), Y \rangle, \end{aligned}$$

as desired. □

In particular, if $X > 0$ we see that f_A is differentiable at X with the gradient given by $\nabla f_A(X) = \tilde{\mathfrak{D}}_{f,X}(A)$. We must carefully consider the case when X is not positive definite. Even if $\mathrm{Tr}(Af(X))$ is finite, the directional derivatives $f'(X; Y)$ are not necessarily finite if $Y \not> 0$. Let $A \in \mathrm{H}_{d,+}$ and $X \in \mathrm{H}_{d,+}$ such that $\mathrm{Tr}(Af(X))$ is finite. The directional derivatives $f'_A(X; Y)$ exists for any $Y \in \mathrm{H}_d$, but the value is possibly infinite if X is on the boundary of the domain of f_A. When $X > 0$, the directional derivatives are always finite since X is in the interior of the domain of f_A. If $X \not> 0$, then A must be zero outside of the support of X, i.e. $\mathrm{supp}(A) \subseteq \mathrm{supp}(X)$. That is, it must be the case that $A|v\rangle = 0$ for every nonzero $|v\rangle$ such that $\langle v|X|v\rangle = 0$. Since we can assume without loss of generality that X

is diagonal[2], this means that if $\mathrm{Tr}(Af(X))$ is finite then A must have the block form

$$A = \begin{pmatrix} \tilde{A} & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{where} \qquad X = \begin{pmatrix} \tilde{X} & 0 \\ 0 & 0 \end{pmatrix},$$

and $\tilde{X} = \mathrm{diag}(x_1, \ldots, x_r)$ is a diagonal matrix with eigenvalues $x_1, \ldots, x_r > 0$.

Here, we give the conditions for which the directional derivative $f_A(X;Y)$ exists and is finite. The definition of the Fréchet differential map $\mathfrak{D}_{f,X}$ can be extended to singular X as follows: If X is diagonal with $X = \mathrm{diag}(x_1, \ldots, x_d)$, we define the linear map $\tilde{\mathfrak{D}}_{f,X}$ on $d \times d$ matrices where $\tilde{\mathfrak{D}}_{f,X}(Y)$ has matrix elements given by

$$\left(\tilde{\mathfrak{D}}_{f,X}(Y)\right)_{i,j} = \begin{cases} f^{[1]}(x_i, x_j) Y_{i,j} & x_i, x_j \neq 0 \\ 0 & x_i = 0 \text{ or } x_j = 0. \end{cases}$$

If X is not diagonal, we may define $\tilde{\mathfrak{D}}_{f,X}$ analogously as $\tilde{\mathfrak{D}}_{f,X}(Y) = U\tilde{\mathfrak{D}}_{f,U^\dagger XU}(U^\dagger YU)U^\dagger$ where U is a unitary matrix such that $U^\dagger XU$ is diagonal.

Lemma 4.4. *Let $f : (0, \infty) \to \mathbb{R}$ be analytic. Let $X \in \mathrm{H}_{d,+}$ be singular, let $A \in \mathrm{H}_{d,+}$ with $\mathrm{supp}(A) \subseteq \mathrm{supp}(X)$, and let $Y \in \mathrm{H}_d$ such that $\langle v|Y|v\rangle > 0$ for every nonzero $|v\rangle$ with $\langle v|X|v\rangle = 0$. Then the directional derivative $f'_A(X;Y)$ exists and is given by*

$$f'_A(X;Y) = \langle \tilde{\mathfrak{D}}_{f,X}(A), Y\rangle.$$

This is a generalization of Lemma 6 in [FGR11], where it was proved for $f(x) = \log x$. A proof of this lemma may be found in the appendix. Lemma 4.4 leads to the following useful corollary.

Corollary 4.5. *Let $f : (0, \infty) \to \mathbb{R}$ be analytic, let $X, X' \in \mathrm{H}_{d,+}$ such that $X' > 0$, and let*

[2] Otherwise there is a unitary matrix U such that $UXU^\dagger$ is diagonal. Then $f(X) = U^\dagger f(UXU^\dagger)U$ and

$$\mathrm{Tr}(Af(X)) = \mathrm{Tr}(UAU^\dagger f(UXU^\dagger)).$$

$A \in \mathrm{H}_{d,+}$ such that $\operatorname{supp}(A) \subseteq \operatorname{supp}(X)$. Then

$$f'_A(X; X' - X) = \langle \tilde{\mathfrak{D}}_{f,X}(A), X' - X \rangle \tag{4.3}$$

and is finite.

Proof. This follows directly from Lemma 4.4. Indeed, the matrix $X' - X \in \mathrm{H}_d$ satisfies the condition that $\langle v|(X' - X)|v\rangle > 0$ for all nonzero $|v\rangle$ such that $\langle v|X|v\rangle = 0$, since $X' > 0$. □

Note that, if $X \not> 0$ then f_A is no longer differentiable at X and the gradient ∇f_A does not exist. Only if $X > 0$ can we replace $\tilde{\mathfrak{D}}_{f,X}(A)$ in (4.3) by the gradient $\nabla f_A(X)$. In fact, f_A is no longer Fréchet differentiable at X and the operator $\tilde{\mathfrak{D}}_{f,X}(A)$ can only be used to compute the directional derivatives for certain directions. Nonetheless, it is still useful for writing down necessary and sufficient conditions for minimization, as will be shown in the next section.

Since $\tilde{\mathfrak{D}}_{f,X} = \mathfrak{D}_{f,X}$ for positive definite X, for the rest of this thesis we will simply write $\mathfrak{D}_{f,X}$ even if X is singular.

4.1.2 Necessary and sufficient conditions

We can now present necessary and sufficient conditions for a matrix X in a convex set $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ to minimize f_A for a fixed analytic function f and a fixed $A \in \mathrm{H}_{d,+}$. These conditions are given in terms of the directional derivatives. We must assume that f is operator convex so that the function f_A is convex for any $A \in \mathrm{H}_{d,+}$. If $X > 0$, this condition is easy to verify. Otherwise, if $X \not> 0$, we must require that $X' > 0$ for all X' in the relative interior of $\mathcal{C}$.

Lemma 4.6. *Let $f : (0, +\infty) \to \mathbb{R}$ be an analytic, operator convex function. Let $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ be a convex subset, let $A \in \mathrm{H}_{d,+}$, and let $X \in \mathcal{C}$. Suppose one of the following holds:*

1. *$X > 0$, or*

2. $X \not> 0$ *but* $X' > 0$ *for all* $X' \in \operatorname{relint}(\mathcal{C})$*.*

Then $f_A(X') \geq f_A(X)$ *for all* $X' \in \mathcal{C}$ *if and only if*

$$\langle \mathfrak{D}_{f,X}(A), X' \rangle \geq \langle \mathfrak{D}_{f,X}(A), X \rangle \tag{4.4}$$

for all $X' \in \mathcal{C}$*.*

Proof. The function $f_A(X) = \operatorname{Tr}(Af(X))$ is convex since f is operator convex, so $X \in \mathcal{C}$ minimizes f_A over $\mathcal{C}$ if and only if $f_A'(X; X' - X) \geq 0$ for all X'. The result follows in case 1, since the directional derivatives can be given by $f_A'(X; X' - X) = \langle \mathfrak{D}_{f,X}(A), X' - X \rangle$ for all $X' \in \mathcal{C}$.

For case 2, suppose that $f_A'(X; X' - X) \geq 0$ holds for all $X' \in \mathcal{C}$, where the directional derivative is possibly infinite if $X' \not> 0$. We have that (4.4) holds for all $X' > 0$ in the relative interior of $\mathcal{C}$, so by continuity it must hold for all X' on the boundary of $\mathcal{C}$ as well. On the other hand, if (4.4) holds for all $X' \in \mathcal{C}$, then $f_A(X') \geq f_A(X)$ for all $X' > 0$ in $\mathcal{C}$. By continuity, it must hold for all X' on the boundary of $\mathcal{C}$ as well. □

Case 2 in Lemma 4.6 requires that the relative interior of $\mathcal{C}$ contains only positive definite matrices. This requires that $\mathcal{C}$ cannot consist entirely of singular matrices. For all applications of this theorem, this will indeed be the case. For example, the sets of separable and PPT operators on a bipartite Hilbert space satisfy this requirement, as do other sets of free states in most quantum resource theories. We will therefore assume that this holds in the remainder of this work.

Essentially, Lemma 4.6 shows that if $X \in \mathcal{C}$ minimizes the function $f_A(X') = \operatorname{Tr}(Af(X'))$ over $\mathcal{C}$, then the Fréchet derivative applied to A defines a supporting hyperplane of $\mathcal{C}$ at X. As the following theorem shows, the reverse is also true. Characterizing the supporting hyperplanes of a point $X \in \operatorname{bd}(\mathcal{C})$ on the boundary of a convex set allows us to find all $A \neq X$ such that the function f_A is minimized at X. In Theorem 4.7, we will show necessary and sufficient conditions for minimization given in terms of supporting hyperplanes.

Here we will work with functions $f : (0, +\infty) \to \mathbb{R}$ that are strictly monotone so that the map $\mathfrak{D}_{f,X}$ is invertible if $X > 0$. Strict monotonicity of f means that $f'(t) \neq 0$ for all $t > 0$, and that $f(t) \neq f(s)$ for all $t \neq s$. Thus all of the elements of the matrix $f^{[1]}(X)$ are nonzero. If X is full rank, then $\mathfrak{D}_{f,X}$ is fully invertible as a linear map on matrices. We may assume without loss of generality that X is diagonal. Then the inverse is given by

$$\mathfrak{D}_{f,X}^{-1}(Y) = \left(f^{[1]}(X)\right)^{-1} \circ Y,$$

where $(f^{[1]}(X))^{-1}$ is the entry-wise inverse of the matrix of divided differences with entries

$$\left(\left(f^{[1]}(X)\right)^{-1}\right)_{j,k} = \frac{1}{\left(f^{[1]}(X)\right)_{j,k}}$$

such that $\mathfrak{D}_{f,X}^{-1}(\mathfrak{D}_{f,X}(Y)) = \mathfrak{D}_{f,X}(\mathfrak{D}_{f,X}^{-1}(Y)) = Y$ for all matrices $Y \in \mathrm{H}_d$. In the case when X is not full rank, the Fréchet derivative operator $\mathfrak{D}_{f,X}$ is only invertible on matrices Y with $\mathrm{supp}(Y) \subseteq \mathrm{supp}(X)$. We can define the *pseudoinverse* of the map $\mathfrak{D}_{f,X}$ in this case by $\mathfrak{D}_{f,X}^{\ddagger}$, which has matrix elements given by

$$\left(\mathfrak{D}_{f,X}^{\ddagger}(Y)\right)_{j,k} = \left\{ \begin{array}{ll} 0, & x_j = 0 \text{ or } x_k = 0, \end{array}\right.$$

where we assume without loss of generality that X is diagonal with $X = \mathrm{diag}(x_1, \ldots, x_d)$. If X is not diagonal, then we define the pseudoinverse by $\mathfrak{D}_{f,X}^{\ddagger}(Y) = U\mathfrak{D}_{f,U^\dagger XU}^{\ddagger}(U^\dagger YU)U^\dagger$, where U is a unitary matrix such that $U^\dagger XU$ is diagonal. The pseudoinverse then satisfies

$$\mathfrak{D}_{f,X}^{\ddagger}(\mathfrak{D}_{f,X}(Y)) = \mathfrak{D}_{f,X}(\mathfrak{D}_{f,X}^{\ddagger}(Y)) = P_X Y P_X$$

for all $Y \in \mathrm{H}_d$, where P_X denotes the projection matrix onto the subspace that is image of X such that $P_X X = X P_X = X$. It holds that $P_X Y P_X = Y$ for any Y such that $\mathrm{supp}(Y) \subseteq \mathrm{supp}(X)$. If $X > 0$ then $\mathfrak{D}_{f,X}^{\ddagger} = \mathfrak{D}_{f,X}^{-1}$, and the projection matrix $P_X = \mathbb{1}_d$ is the identity matrix.

We are now ready to state the main result of this section. Given a convex subset of positive matrices $\mathcal{C}$, Theorem 4.7 states the necessary sufficient conditions for a matrix $X \in \mathcal{C}$ on the relative boundary to minimize functions of the form $f_A(X) = \mathrm{Tr}(Af(X))$. These conditions are given in terms of supporting hyperplanes that are related to the matrix A by the inverse of the Fréchet derivative operator.

Theorem 4.7. *Let $f : (0, \infty) \to \mathbb{R}$ be analytic such that f is operator convex and strictly monotonic. Let $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ be convex such that $X' > 0$ for all $X' \in \mathrm{relint}(\mathcal{C})$, let $X \in \mathrm{bd}(\mathcal{C})$, and let $A \in \mathrm{H}_{d,+}$ such that $A \neq X$. The following are equivalent:*

(i) It holds that $\mathrm{Tr}(Af(X')) \geq \mathrm{Tr}(Af(X))$ for all $X' \in \mathcal{C}$ (i.e., X minimizes f_A over $\mathcal{C}$).

(ii) There exists a nonzero matrix $\phi \in \mathrm{H}_d$, $\phi \neq P_X$, such that $\mathrm{supp}(\phi) \subseteq \mathrm{supp}(X)$ and $\langle \phi, X' \rangle \geq \langle \phi, X \rangle$ for all $X' \in \mathcal{C}$, and $A = \mathfrak{D}^{\ddagger}_{f,X}(\phi)$.

Proof. First suppose that $f_A(X') \geq f_A(X)$ holds for all $X' \in \mathcal{C}$. Then $\mathrm{supp}(A) \subseteq \mathrm{supp}(X)$, and Lemma 4.6 says that $\langle \mathfrak{D}_{f,X}(A), X' \rangle \geq \langle \mathfrak{D}_{f,X}(A), X \rangle$ holds for all $X' \in \mathcal{C}$. Hence, the matrix $\phi = \mathfrak{D}_{f,X}(A)$ satisfies the desired properties. Indeed, it holds that $\mathrm{supp}(\phi) \subseteq \mathrm{supp}(X)$ by definition of pseudo-inverse of $\mathfrak{D}_{f,X}$, and $\phi \neq P_X$ since $A \neq X$.

On the other hand, suppose that ϕ is nonzero with $\phi \neq P_X$ and $\mathrm{supp}(\phi) \subseteq \mathrm{supp}(X)$, and suppose ϕ defines a supporting hyperplane of $\mathcal{C}$ such that $\langle \phi, X' \rangle \geq \langle \phi, X \rangle$ for all $X' \in \mathcal{C}$. For $A = \mathfrak{D}^{\ddagger}_{f,X}(\phi)$, we have that $\mathfrak{D}_{f,X}(A) = \mathfrak{D}_{f,X}(\mathfrak{D}^{\ddagger}_{f,X}(\phi)) = P_X \phi P_X = \phi$, so by Lemma 4.6 we see that X does indeed minimize f_A over $\mathcal{C}$. Lastly, we see that $A \neq X$ since $\phi \neq P_X$ and $\mathfrak{D}_{f,X}$ is invertible on all all operators whose support is contained in $\mathrm{supp}(X)$. □

This theorem also allows us to examine the *converse* convex optimization problem. That is, given a point $X \in \mathrm{bd}(\mathcal{C})$ on the boundary of a convex set, we can find all matrices $A \geq 0$ such that the function f_A is minimized at X by characterizing all of the supporting hyperplanes of $\mathcal{C}$ at X. Indeed, if a matrix ϕ defines a supporting hyperplane of $\mathcal{C}$ at X such that $\langle \phi, X' \rangle \geq \langle \phi, X \rangle$ for all $X' \in \mathcal{C}$, then X minimizes f_A over $\mathcal{C}$ for the matrix $A = \mathfrak{D}^{\ddagger}_{f,X}(\phi)$.

We will use the notation $\sigma = X$ when X is a density matrix. For most applications of Theorem 4.7 in this thesis, the convex set $\mathcal{C}$ will be a subset of density matrices σ', and $\operatorname{Tr}\sigma' = \langle \mathbb{1}_d, \sigma' \rangle = 1$ for all $\sigma' \in \mathcal{C}$. Hence, if ϕ defines a supporting hyperplane of $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ at some point $\sigma \in \mathcal{C}$, then so does $t\phi + s\mathbb{1}_d$ for all $t > 0$ and $s \in \mathbb{R}$. This allows us to state the following corollary.

Corollary 4.8. *Suppose that $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ is a subset of density matrices and that $A > 0$. If $\sigma \in \mathcal{C}$ minimizes f_A over $\mathcal{C}$, then σ also minimizes $f_{A'}$ over $\mathcal{C}$ for all $A' \in \mathrm{H}_{d,+}$ of the form*

$$A' = tA + s\mathfrak{D}_{f,\sigma}^{-1}(\mathbb{1}_d)$$

with $t > 0$ and $s \in \mathbb{R}$.

Proof. Note that $\sigma > 0$, otherwise $f_A(\sigma)$ would not be finite since $A > 0$. From Theorem 4.7, there is a $\phi \in \mathrm{H}_d$ defining a supporting hyperplane of $\mathcal{C}$ such that $\langle \phi, \sigma' \rangle \geq \langle \phi, \sigma \rangle$ for all $\sigma' \in \mathcal{C}$. Let $t > 0$ and $s \in \mathbb{R}$, in which case $\phi' = t\phi + s\mathbb{1}_d$ also defines a supporting hyperplane of the desired form, since

$$\begin{aligned} \langle \phi', \sigma' \rangle = \langle t\phi + s\mathbb{1}_d, \sigma' \rangle &= t\langle \phi, \sigma' \rangle + s \\ &\geq t\langle \phi, \sigma \rangle + s \\ &= \langle t\phi + s\mathbb{1}_d, \sigma \rangle \\ &= \langle \phi', \sigma \rangle \end{aligned}$$

for all $\sigma' \in \mathcal{C}$. By Theorem 4.7, the matrix σ also minimizes $f_{A'}$ over $\mathcal{C}$ for

$$\begin{aligned} A' &= \mathfrak{D}_{f,\sigma}^{-1}(\phi') \\ &= t\mathfrak{D}_{f,\sigma}^{-1}(\phi) + s\mathfrak{D}_{f,\sigma}^{-1}(\mathbb{1}_d) \\ &= tA + s\mathfrak{D}_{f,\sigma}^{-1}(\mathbb{1}_d), \end{aligned}$$

as desired. □

Note that $\mathfrak{D}_{f,X}^{-1}(\mathbb{1}_d) = f'(X)^{-1}$ for any $X > 0$. This will be most applicable in the case when $f(x) = -\log x$ which has derivative given by $f'(x) = -x^{-1}$. In this case, we have $f'(X) = -X^{-1}$, and thus $\mathfrak{D}_{f,X}^{-1}(\mathbb{1}_d) = -X$ and $\mathfrak{D}_{f,X}^{-1}(X) = -\mathbb{1}_d$, two facts that will be very useful for analyzing the relative entropy.

Lastly, for a given $A > 0$ we note that the matrix $X \in \mathcal{C}$ which minimizes f_A over $\mathcal{C}$ will be unique in most cases. This only happens if the function f is not only operator convex, but also *strongly* operator convex, a notion that was discussed in Section 3.2. The functions $f(x) = -x^p$ are strongly operator convex for $p \in (0, \frac{1}{2}]$, as is $f(x) = -\log x$ [FGR11].

Theorem 4.9. *Let $\mathcal{C} \subset \mathrm{H}_{d,+}$ be convex, let $A > 0$, and let $f : (0, +\infty) \to \mathbb{R}$ be analytic and strongly operator convex. There exists a unique $X \in \mathcal{C}$ such that X minimizes f_A over $\mathcal{C}$.*

Proof. Let $X' \in \mathcal{C}$ and suppose that $f_A(X') = f_A(X)$, but $X' \neq X$. By strong operator convexity of f, for all $t \in (0,1)$ it holds that $(1-t)f(X) + tf(X') - f((1-t)X + tX') \geq 0$ and

$$(1-t)f(X) + tf(X') - f((1-t)X + tX') \neq 0,$$

and thus $\mathrm{Tr}\Big(A\Big((1-t)f(X) + tf(X') - f((1-t)X + tX')\Big)\Big) > 0$ for all $t \in (0,1)$, since $A > 0$. Hence $f_A((1-t)X + tX') < (1-t)f_A(X) + tf_A(X') = f_A(X)$, which is a contradiction since X no longer minimizes f_A over $\mathcal{C}$. □

To summarize, we have given necessary and sufficient conditions for a matrix $X \in \mathrm{H}_{d,+}$ on the boundary of some convex set $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ to minimize a convex functional over $\mathcal{C}$. For analytic functions $f : (0, +\infty) \to \mathbb{R}$, we may define the function $f_A : \mathrm{H}_{d,+} \to \overline{\mathbb{R}}$ by $f_A(X) = \mathrm{Tr}(Af(X))$. If f is operator convex, then f_A is convex as a function of matrices. Given a matrix $X \in \mathcal{C}$ on the boundary of a convex set $\mathcal{C}$, Theorem 4.7 states that finding all $A \in \mathrm{H}_{d,+}$ such that f_A is minimized at X is equivalent to finding all matrices $\phi \in \mathrm{H}_d$ that define supporting hyperplanes of $\mathcal{C}$ at X.

4.2 Relative entropy for convex sets

We can now use the analysis from the previous section to state necessary and sufficient conditions for minimizing the relative entropy with respect to a convex set of positive matrices. Recall that the relative entropy of two positive matrices $\rho, \sigma \in \mathrm{H}_{d,+}$ is defined as $S(\rho\|\sigma) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma)$. This quantity is finite as long as $\mathrm{supp}(\sigma) \subseteq \mathrm{supp}(\rho)$ (i.e., as long as $\rho|v\rangle = 0$ for all $|v\rangle$ such that $\sigma|v\rangle = 0$). Otherwise, we define $S(\rho\|\sigma) = +\infty$. The relative entropy is satisfies $S(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$ (This is known as *Klein's inequality*. See, e.g., [Car10]). It is also jointly convex. Convexity in the second argument is easy to see, since the function $-\log(x)$ is operator convex. The relative entropy is therefore a useful quantity to define the 'distance' from a matrix ρ to a convex set $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ of matrices.

Let $\rho \in \mathrm{H}_{d,+}$, and let $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ be convex and compact. We define the *relative entropy of ρ with respect to $\mathcal{C}$* as

$$S(\rho\|\mathcal{C}) := \min_{\sigma \in \mathcal{C}} S(\rho\|\sigma).$$

A matrix $\sigma \in \mathcal{C}$ is said to *minimize the relative entropy of ρ with respect to $\mathcal{C}$* if it holds that $S(\rho\|\mathcal{C}) = S(\rho\|\sigma)$, that is, if $S(\rho\|\sigma) \leq S(\rho\|\sigma')$ holds for all other $\sigma' \in \mathcal{C}$. For example, the relative entropy entanglement (with respect to the separable density matrices) for a bipartite state $\rho \in \mathrm{H}_{d,+,1}$ with $d = d_\mathsf{A} d_\mathsf{B}$ can be written as $E_R^{\mathrm{Sep}}(\rho) = S(\rho\|\mathrm{SepD}(\mathbb{C}^{d_\mathsf{A}} : \mathbb{C}^{d_\mathsf{B}}))$. The relative entropy also arises as an important resource measure in any other quantum resource theory in which the set of free states is a convex set. It is therefore important to find conditions for when a state $\sigma \in \mathcal{C}$ minimizes the relative entropy with respect to an arbitrary convex set $\mathcal{C}$.

Finding a matrix $\sigma \in \mathrm{C}$ that minimizes the relative entropy of ρ with respect to $\mathcal{C}$ is a difficult problem, and no closed-form solution can be found in general. Here, however, we will state useful conditions that determine when a matrix minimizes this relative entropy. As we shall see, this will allow us to find expressions for the relative entropy with respect to convex sets of density matrices.

The problem of computing $S(\rho\|\mathcal{C})$ for an arbitrary fixed $\rho \in \mathrm{H}_{d,+}$ and arbitrary convex sets $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ can be stated as follows: Find $\sigma \in \mathcal{C}$ such that

$$-\operatorname{Tr}(\rho \log \sigma) \leq -\operatorname{Tr}(\rho \log \sigma') \quad \text{for all } \sigma' \in \mathcal{C}. \tag{4.5}$$

Since $-\log$ is an operator convex function, we can use the analysis from the previous section to restate this in terms of the directional derivatives. In particular, $\sigma \in \mathcal{C}$ satisfies (4.5) if and only if

$$f'_\rho(\sigma; \sigma' - \sigma) \geq 0 \quad \text{for all } \sigma' \in \mathcal{C},$$

where f is the function $f(x) = -\log x$. This function is convex, analytic, and strictly monotonic on the interval $(0, \infty)$.

We can now present useful necessary sufficient conditions for a matrix X on the boundary of a convex set $\mathcal{C}$ of positive matrices to minimize the relative entropy of some matrix $\rho \notin \mathcal{C}$ with respect to $\mathcal{C}$. The cases when X is full-rank (positive definite) and singular (not strictly positive definite) must be considered separately. The case when of full-rank minimizers is considered in Section 4.2.1, while the case of singular minimizers is considered in Section 4.2.2.

4.2.1 Full-rank minimizers of relative entropy

We first consider the simpler case of full-rank minimizers. We will state the conditions for a matrix $\sigma \in \mathcal{C}$ to minimize $S(\rho\|\mathcal{C})$ in the case when σ is full-rank. If $\sigma \in \mathrm{H}_{d,++}$ is full rank, for any $X \in \mathrm{H}_d$ it holds that $\sigma + tX > 0$ for t small enough. The directional derivatives $f_\rho(\sigma; X)$ exist and are finite for all X as long as σ is full rank.

Let $\sigma \in \mathrm{H}_{d,++}$ be positive definite. For simplicity, the Fréchet derivative of the logarithm function at a positive definite X will be denoted by L_X, which we define as

$$L_X = \mathfrak{D}_{\log,X} = -\mathfrak{D}_{-\log,X},$$

and is a linear map on H_d. We may assume without loss of generality that σ is diagonal with $\sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ such that L_σ acts on any $X \in \mathrm{H}_d$ as

$$(L_\sigma(X))_{j,k} = \begin{cases} \frac{1}{\lambda_j} X_{j,k} & \lambda_j = \lambda_k \\ \frac{\log \lambda_j - \log \lambda_k}{\lambda_j - \lambda_k} X_{j,k} & \lambda_j \neq \lambda_k. \end{cases} \tag{4.6}$$

The Fréchet derivative of the logarithm function on matrices has the unique and useful property that $L_\sigma(\sigma) = \mathbb{1}_d$ for any $\sigma \in \mathrm{H}_d$ that is positive definite. This property is very useful for analyzing the relative entropy of entanglement. In Theorem 4.12, we state necessary and sufficient conditions for a full-rank matrix σ on the boundary of a convex set $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ of density operators to minimize $S(\rho\|\sigma)$ over $\sigma \in \mathcal{C}$. We first make use of Theorem 4.7 to state the following Lemma. This result was first used in the context of minimizing the relative entropy of entanglement for states of two-qubit systems [Ish03], and generalized to multipartite systems of any dimension [FGR11]. Here the statement will be made in relation to *any* convex subset positive matrices, not just in the context of entanglement theory.

Lemma 4.10. *Let $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ be a convex subset. Let $\sigma \in \mathcal{C}$ be full-rank and let $\rho \in \mathrm{H}_{d,+}$. Then σ minimizes the relative entropy of ρ with respect to $\mathcal{C}$ if and only if*

$$\langle L_\sigma(\rho), \sigma' \rangle \leq \langle L_\sigma(\rho), \sigma \rangle$$

for all $\sigma' \in \mathcal{C}$. Furthermore, $\langle L_\sigma(\rho), \sigma \rangle = \mathrm{Tr}(\rho)$.

Proof. This follows from Theorem 4.7 and the fact that $L_\sigma(\sigma) = \mathbb{1}_d$ for all $\sigma > 0$. □

For most of the applications of Lemma 4.10 in this thesis, the matrix $\rho \in \mathrm{H}_{d,+}$ will be a density matrix with $\mathrm{Tr}\,\rho = 1$. Since the trace of ρ may be written as $\mathrm{Tr}\,\rho = \langle \rho, \mathbb{1}_d \rangle$, and

$L_\sigma(\sigma) = \mathbb{1}_d$ when σ is full-rank, we have that $\langle L_\sigma(\rho), \sigma\rangle = \langle \rho, L_\sigma(\sigma)\rangle = 1$.

We will assume that $S(\rho\|\mathcal{C})$ is finite for all convex sets $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ considered here (i.e., there is at least one $\sigma \in \mathcal{C}$ such that $S(\rho\|\sigma) < +\infty$). Otherwise, $S(\rho\|\sigma) = +\infty$ for every $\sigma \in \mathcal{C}$ and it does not make sense to discuss which σ 'minimizes' the relative entropy with ρ. With this assumption, note that if ρ is full rank and $\sigma \in \mathcal{C}$ minimizes the relative entropy of ρ with respect to $\mathcal{C}$, then σ must also be full rank. Indeed if ρ is full rank then $S(\rho\|\sigma)$ is infinite for every σ that is not full rank.

A convex set $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ is said to *span* all of H_d if any hermitian $X \in \mathrm{H}_d$ can be written as a linear combination of elements in $\mathcal{C}$. For example, the sets of separable and PPT density matrices on a bipartite system spans the space of all hermitian matrices. Lemma 4.10 leads to the following intuitive corollary for such convex sets of density matrices. Suppose $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ is a subset of density matrices that spans H_d. For full-rank matrices $\sigma \in \mathcal{C}$, and a density matrix $\rho \notin \mathcal{C}$, if σ minimizes the relative entropy of ρ with $\mathcal{C}$, then σ must be on the relative boundary of $\mathcal{C}$.

Corollary 4.11. *Let $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ be convex and spanning. Let $\sigma, \rho \in \mathrm{H}_{d,+}$ with $\sigma \in \mathcal{C}$ and $\rho \notin \mathcal{C}$ such that $S(\rho\|\mathcal{C}) = S(\rho\|\sigma)$. If $\sigma > 0$ then $\sigma \in \mathrm{bd}(\mathcal{C})$.*

Proof. Suppose instead that σ is in the relative interior of $\mathcal{C}$ and let $\sigma' \in \mathcal{C}$. Since σ is in the interior, there is an $\epsilon > 0$ such that $(1-t)\sigma + t\sigma' \in \mathcal{C}$ for all $t \in (-\epsilon, \epsilon)$. From Lemma 4.10,

$$\begin{aligned}\langle L_\sigma(\rho), (1-t)\sigma + t\sigma'\rangle &= \langle L_\sigma(\rho), \sigma\rangle + t\langle L_\sigma(\rho), \sigma' - \sigma\rangle \\ &\leq \langle L_\sigma(\rho), \sigma\rangle\end{aligned}$$

holds for all $t \in (-\varepsilon, \varepsilon)$, so $\langle L_\sigma(\rho), \sigma' - \sigma\rangle = 0$ and thus $\langle L_\sigma(\rho), \sigma'\rangle = \langle L_\sigma(\rho), \sigma\rangle = 1$. Note that $\operatorname{Tr}\sigma' = \langle \mathbb{1}_d, \sigma'\rangle = 1$. Hence we see that $\langle L_\sigma(\rho), \sigma'\rangle = \langle \mathbb{1}_d, \sigma'\rangle$ for all $\sigma' \in \mathcal{C}$. Since $\mathcal{C}$ spans all of H_d, it follows that $L_\sigma(\rho) = \mathbb{1}_d$ and thus $\rho = \sigma$, a contradiction since $\rho \notin \mathcal{C}$. □

Recall that a matrix $\phi \in \mathrm{H}_d$ is said to define a supporting hyperplane of a convex set $\mathcal{C}$ at a point σ on the boundary of $\mathcal{C}$ if $\phi \neq 0$ and $\langle \phi, \sigma'\rangle \leq \langle \phi, \sigma\rangle$ for all $\sigma' \in \mathrm{C}$. We now state the necessary and sufficient conditions for a full-rank state on the boundary of a convex set to minimize the relative entropy of an operator ρ with respect to $\mathcal{C}$. This is a generalization of Theorem 3 in [FGR11], where it was stated in the case where the convex set $\mathcal{C}$ is the set of separable (or PPT) density operators on a bipartite space.

Theorem 4.12. *Let $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ be a closed, convex subset of density operators. Let σ be a full-rank density operator on the boundary of $\mathcal{C}$, and let $\rho \in \mathrm{H}_{d,+,1}$ with $\rho \notin \mathcal{C}$. The following are equivalent:*

(i) *$S(\rho\|\sigma) \leq S(\rho\|\sigma')$ for all $\sigma' \in \mathcal{C}$*

(ii) *There exists $\phi \in \mathrm{H}_d$ with $\|\phi - \mathbb{1}_d\|_1 = 1$ such that*

$$\langle \phi, \sigma' \rangle \leq \langle \phi, \sigma \rangle = 1 \quad \textit{for all } \sigma' \in \mathcal{C} \tag{4.7}$$

(i.e., ϕ defines a supporting hyperplane of $\mathcal{C}$ at σ), and ρ is of the form

$$\rho = (1-t)\sigma + tL_\sigma^{-1}(\phi) \tag{4.8}$$

for $t \in (0, t_{\max}]$, where $t_{\max} > 0$ is the largest value such that $(1-t)\sigma + tL_\sigma^{-1}(\phi) \geq 0$.

For a fixed σ and ϕ, note that $t_{\max}$ can be computed as the solution to the following semidefinite program:

$$\begin{aligned} &\text{maximize: } t \\ &\text{subject to: } t(L_\sigma^{-1}(\phi) - \sigma) \geq -\sigma. \end{aligned}$$

Proof. Suppose that $S(\rho\|\sigma) \leq S(\rho\|\sigma')$ for all $\sigma' \in \mathcal{C}$. By Lemma 4.10 it holds that $\langle L_\sigma(\rho), \sigma' \rangle \leq \langle L_\sigma(\rho), \sigma \rangle$ for all $\sigma' \in \mathcal{C}$. For any $t > 0$, the operator $\phi = \frac{1}{t}\left(L_\sigma(\rho) - \mathbb{1}_d\right) + \mathbb{1}_d$

defines a supporting hyperplane of the desired form, since for all $\sigma' \in \mathcal{C}$

$$\begin{aligned}\langle \phi, \sigma' \rangle = \frac{1}{t}\left(\langle L_\sigma(\rho), \sigma' \rangle - 1\rangle\right) + 1 &\leq \frac{1}{t}\left(\langle L_\sigma(\rho), \sigma \rangle - 1\rangle\right) + 1 \\ &= \frac{1}{t}\underbrace{\left(\langle \rho, \mathbb{1}_d \rangle - 1\rangle\right)}_{=0} + 1 \\ &= 1,\end{aligned}$$

where we note that $L_\sigma(\sigma) = \mathbb{1}_d$. Furthermore, we see that $\phi - \mathbb{1}_d = \frac{1}{t}L_\sigma(\rho - \sigma) \neq 0$ since $\rho \neq \sigma$ and L_σ is an invertible linear map. Hence we may choose $t = \|L_\sigma(\rho - \sigma)\|_1 > 0$ such that $\|\phi - \mathbb{1}_d\|_1 = 1$. We see that $L_\sigma(\rho) = (1-t)\mathbb{1}_d + t\phi$, and thus $\rho = (1-t)\sigma + tL_\sigma^{-1}(\phi)$.

On the other hand, let $\phi \in \mathrm{H}_d$ be an operator that defines a supporting hyperplane of the desired form and suppose that $\rho = (1-t)\sigma + tL_\sigma^{-1}(\phi)$ for some $t \in (0, t_{\max}]$. To show that σ minimizes the relative entropy with ρ, it suffices to check that $\langle L_\sigma(\rho), \sigma' - \sigma \rangle \leq 0$ for all $\sigma' \in \mathcal{C}$. Let $\sigma' \in \mathcal{C}$. It follows that

$$\begin{aligned}\langle L_\sigma(\rho), \sigma' - \sigma \rangle &= \langle (1-t)\mathbb{1}_d + t\phi, \sigma' - \sigma \rangle \\ &= (1-t)\underbrace{\langle \mathbb{1}_d, \sigma' - \sigma \rangle}_{=\mathrm{Tr}(\sigma' - \sigma) = 0} + t\langle \phi, \sigma' - \sigma \rangle \\ &\leq 0.\end{aligned}$$

This concludes the proof. □

The proof of Theorem 4.12 relies on the fact that $L_\sigma(\sigma) = \mathbb{1}_d$ if σ is full-rank. If σ is not full-rank, only a sufficient condition regarding when a state ρ has σ on the boundary on $\mathcal{C}$ that minimizes the relative entropy with ρ can be given (see [FGR11]). It is also important that $\mathcal{C}$ be a subset of density matrices (or at least contains only matrices with constant trace), otherwise the theorem no longer holds.

A depiction of the result of Theorem 4.12 can be seen in Figure 4.1. Given a convex subset of density matrices $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ and a point $\sigma \in \mathrm{bd}(\mathcal{C})$, we can characterize all states

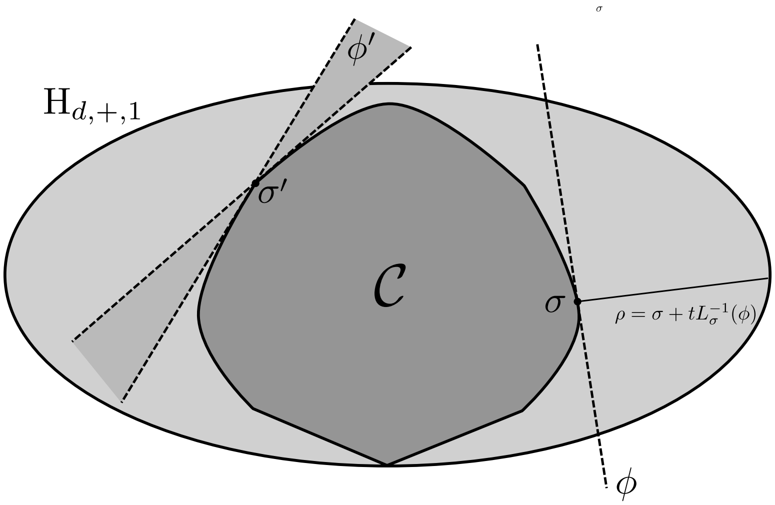


Figure 4.1: Depiction of minimizers for relative entropy with respect to convex sets. For a state $\sigma \in \mathrm{bd}(\mathcal{C})$ on the boundary of a convex set of density matrices $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$, consider the matrices $\phi \in \mathrm{H}_d$ that define supporting hyperplanes of $\mathcal{C}$ at σ. This allows us to find states ρ that have σ as a closest state in $\mathcal{C}$ (where 'closest' means 'minimizes the relative entropy'). Some states σ' on the boundary do not have a unique supporting hyperplane, but instead have a continuous family of supporting hyperplanes. Each such hyperplane defines a family of states for which σ' is the closest.

$\rho \in \mathrm{H}_{d,+,1}$ for which σ minimizes the relative entropy of entanglement of ρ to $\mathcal{C}$ by examining the supporting hyperplanes of $\mathcal{C}$ at σ. If ϕ defines a supporting hyperplane, then states of the form $\sigma + tL_\sigma^{-1}(\phi)$ have relative entropy that is minimized at σ. A generic σ will have a unique supporting hyperplane, but points on cusps of the boundary may have a family of supporting hyperplanes.

4.2.2 Singular minimizers of relative entropy

The preceding section is concerned only with the case when $\rho \in \mathrm{H}_{d,+}$ is full rank, in which case the closest σ in $\mathcal{C}$ must also be full rank. If $\rho \in \mathrm{H}_{d,+}$ is not full rank and $\sigma \in \mathcal{C}$ minimizes the relative entropy of ρ with respect to $\mathcal{C}$, then σ can also be singular. As before, we can extend the definition the Fréchet derivative map for singular $\sigma \in \mathrm{H}_{d,+}$ as in the general case as follows. If σ is singular and diagonal with $\sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, then $L_\sigma(X)$ has matrix elements given by

$$\Big(L_\sigma(X)\Big)_{j,k} = \begin{cases} \frac{1}{\lambda_j} X_{j,k} & x_j = x_k \text{ and } \lambda_j, \lambda_k \neq 0 \\ \frac{\log\lambda_j - \log\lambda_k}{\lambda_j - \lambda_k} X_{j,k} & \lambda_j \neq x_k \text{ and } \lambda_j, \lambda_k \neq 0 \\ 0 & \lambda_j = 0 \text{ or } \lambda_k = 0. \end{cases}$$

As long as X is strictly positive on the null space of σ, the directional derivatives $f'_\rho(\sigma; X)$ are well defined and given by

$$f'_\rho(\sigma; X) = \langle L_\sigma(\rho), X\rangle.$$

Since $f(x) = -\log(x)$ is strictly monotonically decreasing, this map has a pseudo-inverse given by

$$\Big(L^\ddagger_\sigma(X)\Big)_{j,k} = \begin{cases} \lambda_{j,j} X_{j,k} & \lambda_j = \lambda_k \text{ and } \lambda_j \neq 0 \\ \frac{\lambda_j - \lambda_k}{\log\lambda_j - \log\lambda_k} X_{j,k} & \lambda_j \neq \lambda_k \text{ and } \lambda_j, \lambda_k \neq 0 \\ 0 & \lambda_j = 0 \text{ or } \lambda_k = 0. \end{cases}$$

such that $L^\ddagger_\sigma(L_\sigma(X)) = L_\sigma(L^\ddagger_\sigma(X)) = P_\sigma X P_\sigma$ and $L^\ddagger_\sigma(\mathbb{1}_d) = L^\ddagger_\sigma(P_\sigma) = \sigma$, where P_σ is the projection matrix onto the support of σ.

Similar to Lemma 4.10, we can state the necessary and sufficient conditions for a state $\sigma \in \mathcal{C}$ to minimize the relative entropy of ρ with respect to $\mathcal{C}$. However, as in Lemma 4.6, we require that the interior of $\mathcal{C}$ contain only positive definite matrices.

Lemma 4.13. *Let $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ be a convex subset such that $\mathrm{relint}(\mathcal{C}) \subseteq \mathrm{H}_{d,++}$. Let $\sigma \in \mathcal{C}$ be singular and let $\rho \in \mathrm{H}_{d,+}$ such that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. Then σ minimizes the relative entropy of ρ with respect to $\mathcal{C}$ if and only if*

$$\langle L_\sigma(\rho), \sigma'\rangle \leq \langle L_\sigma(\rho), \sigma\rangle$$

for all $\sigma' \in \mathcal{C}$. Furthermore, $\langle L_\sigma(\rho), \sigma\rangle = \mathrm{Tr}(\rho)$.

Proof. This follows directly from the general case of Lemma 4.6. □

We may now make a similar statement to Theorem 4.12, except this time for singular matrices $\sigma \in \mathrm{H}_{d,+}$. However, since $L_\sigma(\sigma) = P_\sigma \neq \mathbb{1}_d$, the statement we can make is slightly weaker. In particular, we cannot guarantee that σ is minimizes the relative entropy for states of the form $\rho = (1-t)\sigma + tL_\sigma^\ddagger(\phi)$ with $t > 1$.

Theorem 4.14. *Let $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ be a closed, convex subset of density operators. Let σ be a singular density operator on the boundary of $\mathcal{C}$, and let $\rho \in \mathrm{H}_{d,+,1}$ with $\rho \notin \mathcal{C}$ and $\rho \neq 0$. The following are equivalent:*

1. *$S(\rho\|\sigma) \leq S(\rho\|\sigma')$ for all $\sigma' \in \mathcal{C}$*

2. *There exists a nonzero $\phi \in \mathrm{H}_d$ with $\phi \neq P_\sigma$ and $\mathrm{supp}(\phi) \subseteq \mathrm{supp}(\sigma)$ such that ϕ defines a supporting hyperplane of $\mathcal{C}$ at σ of the form*

$$\langle \phi, \sigma' \rangle \leq \langle \phi, \sigma \rangle = 1 \quad \textit{for all } \sigma' \in \mathcal{C}, \tag{4.9}$$

and ρ is of the form

$$\rho = (1-t)\sigma + tL_\sigma^\ddagger(\phi) \tag{4.10}$$

for some $t \in (0, t_{\max}]$, where $0 < t_{\max} \leq 1$ is the largest value not greater than 1 such that $(1-t)\sigma + tL_\sigma^\ddagger(\phi) \geq 0$.

Proof. Suppose that $S(\rho\|\sigma) \leq S(\rho\|\sigma')$ for all $\sigma' \in \mathcal{C}$. Note that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ since we can assume that $S(\rho\|\sigma) \neq +\infty$. Then by Lemma 4.13, it holds that $\langle L_\sigma(\rho), \sigma' \rangle \leq \langle L_\sigma(\rho), \sigma \rangle$ for all $\sigma' \in \mathcal{C}$. Hence $\phi = L_\sigma(\rho)$ defines a supporting hyperplane of the desired form, and $\phi \neq P_\sigma$ since $\sigma \neq \rho$ and $L_\sigma = P_\sigma$, and L_σ is invertible on matrices with support contained in $\mathrm{supp}(\sigma)$. Finally, we see that $\rho = L_\sigma^\ddagger(\phi)$, so (4.10) holds with $t = 1$.

On the other hand, suppose that $\phi \in \mathrm{H}_d$ defines a supporting hyperplane of $\mathcal{C}$ at σ of the form in (4.9), and that $\rho = (1-t)\sigma + tL_\sigma^\ddagger(\phi)$ for some $t \in (0, 1]$. Then $L_\sigma(\rho) = t\phi + (1-t)P_\sigma$.

For any $\sigma' \in \mathcal{C}$ we have

$$\begin{aligned}\langle L_\sigma(\rho), \sigma' \rangle &= t\langle \phi, \sigma' \rangle + (1-t)\langle P_\sigma, \sigma' \rangle \\ &\leq t\langle \phi, \sigma \rangle + (1-t)\langle P_\sigma, \sigma \rangle \\ &= \langle L_\sigma(\rho), \sigma \rangle \\ &= 1,\end{aligned}$$

since $\langle P_\sigma, \sigma' \rangle \leq 1$ for all $\sigma' \in \mathrm{H}_{d,+,1}$ and $\langle P_\sigma, \sigma \rangle = 1$. □

The key difference between the singular case and the full-rank case can be seen in the following observation. Since $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ is a subset of density matrices, we see that

$$t\phi + (1-t)\mathbb{1}_d$$

defines a supporting hyperplane for *all* positive t, while $t\phi + (1-t)P_\sigma$ only defines a supporting hyperplane for $0 < t \leq 1$ if $P_\sigma \neq \mathbb{1}_d$.

We also have the following natural corollary, which states that if $\rho \notin \mathcal{C}$ is any state for which $\sigma \in \mathrm{bd}(\mathcal{C})$ minimizes $S(\rho\|\sigma)$ over $\mathcal{C}$, then σ also minimizes $S(t\rho + (1-t)\sigma\|\sigma)$ for all $t \in (0,1]$ as well. If σ is non-singular, then this holds for $t > 1$ as well. Previously [VP98], it was only known that that this could be done if $t \leq 1$.

Corollary 4.15. *Let $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ be a closed convex subset of density operators. Let $\sigma \in \mathrm{bd}(\mathcal{C})$, and let $\rho \in \mathrm{H}_{d,+,1}$ with $\rho \notin \mathcal{C}$ and $\rho \neq 0$ such that $S(\rho\|\sigma) \leq S(\rho\|\sigma')$ for all $\sigma' \in \mathcal{C}$ (i.e., σ minimizes the relative entropy of ρ with respect to $\mathcal{C}$). The following statements hold.*

1. *If $\sigma > 0$ then σ also minimizes the relative entropy of all states of the form $t\rho+(1-t)\sigma$ with respect to $\mathcal{C}$ for all $t > 0$.*

2. *If σ is singular, then σ also minimizes the relative entropy of all states of the form $t\rho + (1-t)\sigma$ with respect to $\mathcal{C}$ for all $t \in (0, 1]$.*

4.3 Application to Relative Entropy of Entanglement

We now consider a bipartite system $\mathbb{C}^{d_\mathsf{A}} \otimes \mathbb{C}^{d_\mathsf{B}}$ with $d = d_\mathsf{A} d_\mathsf{B}$ and apply the previous results to the relative entropy of entanglement (with respect to the PPT states). If we let $\mathcal{P} = \mathrm{PPTD}(\mathbb{C}^{d_\mathsf{A}} : \mathbb{C}^{d_\mathsf{B}})$ denote the convex subset of PPT states, the relative entropy of entanglement can be written as $E_R^{\mathrm{PPT}}(\rho) = S(\rho \| \mathcal{P})$. Then all of the results in Section 4.2 apply to this circumstance. To make use of these results, it is useful to characterize the supporting hyperplanes of $\mathcal{P}$ at a PPT state on the boundary. Here, we only consider the case when σ is full rank.

Suppose $\sigma \in \mathrm{D}(\mathbb{C}^d) = \mathrm{H}_{d,+,1}$ is a full-rank bipartite density operator such that its partial transpose σ^{T_B} is positive but has at least one zero eigenvalue, (that is, $\sigma > 0$ and $\sigma^{T_\mathsf{B}} \geq 0$, but $\sigma^{T_\mathsf{B}} \not> 0$), then σ is on the boundary of $\mathcal{P}$. Indeed, we can define a supporting hyperplane $\mathcal{P}$ at σ as follows: Let $|\psi\rangle \in \mathbb{C}^d$ be a unit vector such that $\langle\psi|\sigma^{T_\mathsf{B}}|\psi\rangle = 0$ and define $\phi = \mathbb{1}_d - |\psi\rangle\langle\psi|^{T_\mathsf{B}}$. Then $\langle\phi, \sigma\rangle = \langle\psi|\sigma^{T_\mathsf{B}}|\psi\rangle = 0$ and

$$\langle\phi, \sigma'\rangle = 1 - \langle\psi|\sigma'^{T_\mathsf{B}}|\psi\rangle \leq 1 \tag{4.11}$$

for all $\sigma'^{T_\mathsf{B}} \geq 0$. We can use this result to find a 'closed form' expression for the relative entropy of entanglement for certain states.

Corollary 4.16. *Let $\sigma \in \mathrm{D}(\mathbb{C}^{d_\mathsf{A} d_\mathsf{B}})$ be a full-rank PPT density operator on the boundary of the PPT states (i.e., such that $\sigma > 0$ and $\sigma^{T_\mathsf{B}} \geq 0$ but $\sigma^{T_\mathsf{B}} \not> 0$), and let $|\psi\rangle \in \mathbb{C}^d$ be a unit vector such that $\langle\psi|\sigma^{T_\mathsf{B}}|\psi\rangle = 0$. For all density operators of the form $\rho = \sigma - tL_\sigma^{-1}(|\psi\rangle\langle\psi|^{T_\mathsf{B}})$, it holds that*

$$E_R^{\mathrm{PPT}}(\rho) = S(\sigma) - S(\rho) + t\langle\psi|(\sigma\log\sigma)^{T_\mathsf{B}}|\psi\rangle \tag{4.12}$$

where $S(\rho) = -\operatorname{Tr}(\rho\log\rho)$ is the von Neumann entropy.

Proof. By the previous explanation, for all ρ of this form we see that σ minimizes the relative entropy of ρ with respect to the PPT states. Then

$$\begin{aligned}E_R^{\mathrm{PPT}}(\rho) = S(\rho\|\sigma) &= \mathrm{Tr}\Big(\rho\log\rho - \rho\log\sigma\Big)\\ &= -S(\rho) - \mathrm{Tr}\Big(\Big(\sigma - tL_\sigma^{-1}(|\psi\rangle\langle\psi|^{T_{\mathsf{B}}})\Big)\log\sigma\Big)\\ &= S(\sigma) - S(\rho) + t\,\mathrm{Tr}\Big(|\psi\rangle\langle\psi|^{T_{\mathsf{B}}} L_\sigma^{-1}(\log\sigma)\Big)\\ &= S(\sigma) - S(\rho) + t\langle\psi|(\sigma\log\sigma)^{T_{\mathsf{B}}}|\psi\rangle,\end{aligned}$$

where $L_\sigma^{-1}(Y) = \sigma Y$ for any $Y \in \mathrm{H}_d$ that is diagonal in the eigenbasis of σ. □

We now use this method to explicitly compute the relative entropy of entanglement of some families of entangled states. We can generate states ρ with known value for the relative entropy of entanglement by generating PPT states that are on the boundary of the set of PPT states and finding supporting hyperplanes (i.e., witnesses) of the set of PPT states at that point. While these states cannot be parameterized, we can generate them by the following procedure. This will be done here for states on $\mathbb{C}^2 \otimes \mathbb{C}^2$ (i.e., states of systems of two qubits). For two-qubit states, the sets of PPT states and the set of separable states coincide, and thus $E_R^{\mathrm{PPT}} = E_R^{\mathrm{Sep}}$.

1. Generate a random 4×3 matrix X. Set $Y = \frac{XX^\dagger}{\mathrm{Tr}(XX^\dagger)}$. Note that Y is a 4×4 density matrix with rank at most 3. If X is generated sufficiently randomly, the matrix Y will have rank 3 with unit probability.

2. If $Y^{T_{\mathsf{B}}} \geq 0$ then set $\sigma = Y^{T_{\mathsf{B}}}$. Otherwise repeat step 1 until σ is generated this way. Note that σ generated this way lies on the boundary of the PPT states.

3. Let $|\psi\rangle$ be a normalized eigenvector of the null space of $\sigma^{T_{\mathsf{B}}} = Y$.

4. Consider the family of states

$$\rho(\sigma,\psi,t) = \sigma + tL_\sigma^{-1}((|\psi\rangle\langle\psi|)^{T_{\mathsf{B}}})$$

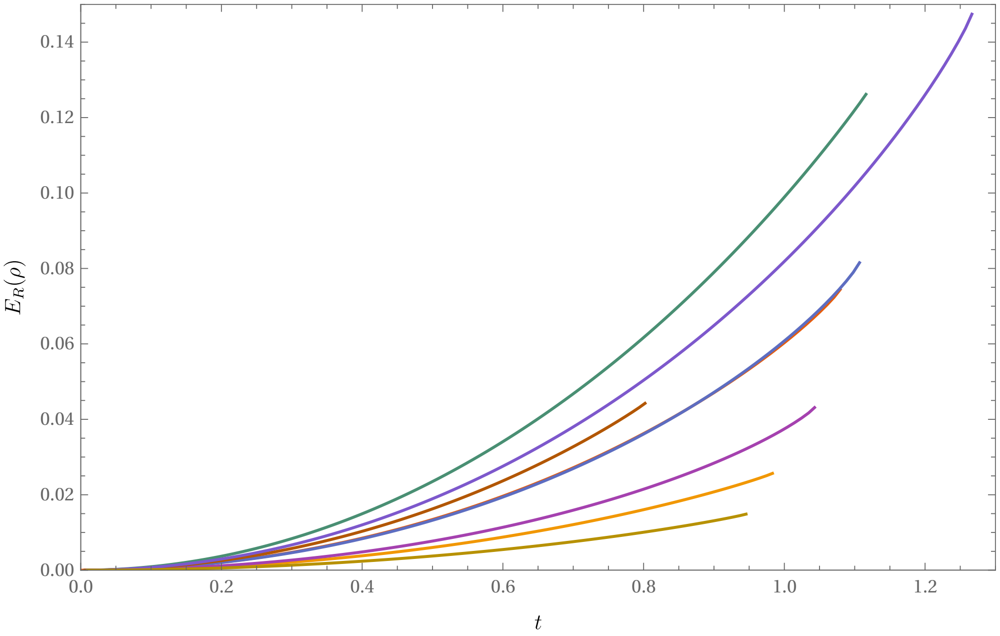


Figure 4.2: Relative entropy of entanglement ($E_R(\rho)$) is plotted for two-qubit states of the form $\rho = \sigma + tL_\sigma^{-1}((|\psi\rangle\langle\psi|)^{T_\mathsf{B}})$ for $t \in (0, t_{\max}]$. This is shown for eight differently randomly generated states σ on the boundary of the set of PPT states. Each curve depicts E_R for a different family of states derived from the different σ generated this way.

for $t \in (0, t_{\max}]$, where $t_{\max}$ is the largest value of t such that $\rho(\sigma, \psi, t)$ is positive semidefinite. Then each $\rho = \rho(\sigma, \psi, t)$ is an entangled two-qubit state whose closest separable state is σ. The relative entropy of entanglement of each ρ is given by

$$E_R(\rho) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma).$$

Using this procedure, we can now generate entangled two-qubit states with known values of relative entropy of entanglement. For eight randomly generated separable states σ that are on the boundary, the relative entropy of entanglement of $\rho = \sigma + tL_\sigma^{-1}((|\psi\rangle\langle\psi|)^{T_\mathsf{B}})$ is plotted for $t \in (0, t_{\max}]$ in Figure 4.2. Note that $t_{\max}$ is different for each of the different randomly generated σ on the boundary of the PPT states.

4.3.1 Additivity of the relative entropy of entanglement

It is of great importance to determine conditions for when the relative entropy of entanglement is weakly additive. These are states ρ for which the relative entropy of entanglement is equal to its regularized version, $E_R^{\mathrm{PPT}}(\rho) = E_{R,\infty}^{\mathrm{PPT}}(\rho)$. Indeed, the regularized version has been shown to be the unique measure of entanglement in a reversible theory of entanglement [BP10], although it is much more difficult to compute. Given a state ρ and a PPT state $\sigma \in \mathrm{bd}(\mathcal{P})$ that minimizes the relative entropy of ρ, it has been shown [Rai99b, Rai00] that $E_R^{\mathrm{PPT}}(\rho)$ is weakly additive if

$$[\rho, \sigma] = 0 \quad \text{and} \quad \left(\rho\sigma^{-1}\right)^{T_{\mathrm{B}}} \geq \mathbb{1}_d. \tag{4.13}$$

However, we have shown that the state ρ may be given by $\rho = (1-t)\sigma + tL_\sigma^{-1}(\phi)$ for some matrix ϕ that defines a supporting hyperplane, so the condition in (4.13) is equivalent to

$$[L_\sigma^{-1}(\phi), \sigma] = 0 \quad \text{and} \quad \left(L_\sigma^{-1}(\phi)\sigma^{-1}\right)^{T_{\mathrm{B}}} \geq \mathbb{1}_d.$$

But $[L_\sigma^{-1}(\phi), \sigma] = 0$ if and only if $[\phi, \sigma] = 0$, so the conditions for weak additivity can be stated as

$$[\phi, \sigma] = 0 \quad \text{and} \quad \left(L_\sigma^{-1}(\phi)\sigma^{-1}\right)^{T_{\mathrm{B}}} \geq \mathbb{1}_d. \tag{4.14}$$

Hence, we have reduced the task of finding states ρ for which $E_{\mathcal{P}}(\rho)$ is weakly additive to finding states σ on the boundary of $\mathcal{P}$ and a corresponding supporting hyperplane of $\mathcal{P}$ at σ defined by ϕ that satisfy the conditions in (4.14). In particular, it has been shown [MI08] that $E_R^{\mathrm{PPT}}(\rho)$ is weakly additive for all states ρ that commute with σ if σ minimizes the relative entropy of entanglement for ρ.

4.3.2 Relative entropy of entanglement for pure states

Here we prove the well known fact [VP98] that, for bipartite pure states, the relative entropy of entanglement (with respect to the separable states) reduces to the von Neumann entanglement entropy. The method uses the conditions for minimization presented earlier in this chapter and is similar to the one that we will use later to compute the Rényi relative entropies of entanglement in Section 4.5. It is also slightly different than the original method used in [VP98]. For these reasons, we will provide the full proof here. For simplicity, we can write $S(\psi\|\sigma) = S(\rho\|\sigma)$ in the case when $\rho = |\psi\rangle\langle\psi|$ is a pure state.

Theorem 4.17. *Let $|\psi\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ be a bipartite pure state. It holds that*

$$E_R^{\text{Sep}}(\psi) = E(\psi),$$

where $E(\psi)$ is the entropy of entanglement of $|\psi\rangle$.

Without loss of generality the pure state may be considered to be in the form $|\psi\rangle = \sum_i \sqrt{\lambda_i}|ii\rangle$, and the entropy of entanglement can be given by $E(\psi) = \sum_i \lambda_i \log \lambda_i$. We can use the formalism presented in this chapter for giving minimization criteria. Lemma 4.18 presents the other main tool that will be used to prove Theorem 4.17.

Lemma 4.18. *For all $p, q > 0$ it holds that $0 < \sqrt{pq}\log^{[1]}(p,q) \leq 1$.*

Proof. Since log is a strictly increasing function, it holds that $0 < \log^{[1]}(p,q)$ for all $p, q > 0$. To show the remaining inequality, note that we can write the divided differences of the logarithm function using an integral representation as

$$\log^{[1]}(p,q) = \int_0^\infty \frac{1}{(p+t)(q+t)} dt.$$

Since $p + q \geq 2\sqrt{pq}$ for any $t \geq 0$, it holds that

$$\begin{aligned}(p+t)(q+t) &= pq + (p+q)t + t^2 \\ &\geq pq + 2\sqrt{pq}t + t^2 \\ &= (\sqrt{pq} + t)^2.\end{aligned}$$

Hence, for any $p, q > 0$ we have

$$\begin{aligned}\sqrt{pq}\log^{[1]}(p,q) &= \sqrt{pq}\int_0^\infty \frac{1}{(p+t)(q+t)}dt \\ &\leq \int_0^\infty \frac{\sqrt{pq}}{(\sqrt{pq}+t)^2}dt \\ &= 1,\end{aligned}$$

as desired. □

Proof (*of Theorem 4.17*). Without loss of generality we may consider $|\psi\rangle$ to be of the form $|\psi\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i}|ii\rangle$, with all $\lambda_i > 0$ and $\sum_i \lambda_i = 1$. Consider the following density matrix

$$\sigma = \sum_{i=1}^{d} \lambda_i |ii\rangle\langle ii|,$$

which is clearly separable since it is a convex combination of separable states $|ii\rangle\langle ii|$. For $\rho = |\psi\rangle\langle\psi|$, note that $\mathrm{Tr}(\rho\log\rho) = 0$ since the eigenvalues of ρ are all either 1 or 0. Thus

$$\begin{aligned}S(\psi\|\sigma) &= -\mathrm{Tr}(|\psi\rangle\langle\psi|\log\sigma) \\ &= -\sum_{i=1}^{d}\lambda_i\log\lambda_i \\ &= E(\psi).\end{aligned}$$

We now need to prove that σ minimizes the relative entropy of $|\psi\rangle$ with respect to the separable states. By Theorem 4.7, this happens if and only if it holds that

$$\langle L_\sigma(\rho), \sigma' \rangle \leq 1 \qquad \text{for all } \sigma' \in \mathrm{SepD}(\mathcal{H}_\mathsf{A} : \mathcal{H}_\mathsf{B}).$$

It suffices to consider only separable pure states of the form $\sigma' = |u\rangle\langle u| \otimes |v\rangle\langle v|$ where

$$|u\rangle = \sum_{i=1}^{d} u_i |i\rangle \qquad \text{and} \qquad |v\rangle = \sum_{i=1}^{d} v_i |i\rangle$$

are pure state vectors with $\sum_i |u_i|^2 = \sum_i |v_i|^2 = 1$. Now

$$\begin{aligned}
\langle L_\sigma(\rho), |u\rangle\langle u| \otimes |v\rangle\langle v| \rangle &= \sum_{i,j} \sqrt{\lambda_i \lambda_j} \log^{[1]}(\lambda_i, \lambda_j) u_i v_i \overline{u_j v_j} \\
&\leq \left| \sum_{i,j} \sqrt{\lambda_i \lambda_j} \log^{[1]}(\lambda_i, \lambda_j) u_i v_i \overline{u_j v_j} \right| \\
&\leq \sum_{i,j} \underbrace{\sqrt{\lambda_i \lambda_j} \log^{[1]}(\lambda_i, \lambda_j)}_{\leq 1 \text{ by Lemma 4.18}} |u_i||v_i||u_j||v_j| \\
&\leq \Big(\sum_i |u_i||v_i| \Big)^2 \\
&\leq \sum_i |u_i|^2 \sum_i |v_i|^2 \\
&= 1,
\end{aligned}$$

as desired. □

4.4 The Rains bound

We may also use the results of Section 4.2 to analyze the Rains bound, whose value can be cast as the solution to a convex optimization problem. In this section, we examine properties of the Rains bound, show how it can be converted into a convex optimization problem, and characterize the supporting hyperplanes of the corresponding convex set. The necessary and sufficient conditions for minimizing the Rains bound is shown. This is used to prove the main result in this chapter, which is that the Rains bound is equal to the relative entropy of entanglement for states of systems that have at least one subsystem be $\mathbb{C}^2$ (i.e., one subsystem is a qubit).

The Rains bound $R(\rho)$ of a bipartite state ρ (see Section 2.4.2) is a quantity that is an upper bound to the PPT-distillable entanglement and a lower bound to the relative entropy of entanglement. It can be defined similarly to the relative entropy of entanglement as

$$R(\rho) = \min_{\sigma} S(\rho\|\sigma) + \log|\sigma^{T_{\mathsf{B}}}|, \tag{4.15}$$

where the minimum is taken over *all* normalized density operators $\sigma \in \mathrm{H}_{d,+,1}$ rather than just the PPT states. However, the function in (4.15) is not convex as a function of σ. To make use of the results from the previous section, we must first convert the problem of computing the Rains bound into a convex optimization problem, which can be done as follows: Consider the convex subset of bipartite positive operators $\mathcal{R} \subseteq \mathrm{H}_{d,+}$ defined by

$$\mathcal{R} = \{\tau \in \mathrm{H}_{d,+} \mid \|\tau^{T_{\mathsf{B}}}\|_1 \leq 1\} \tag{4.16}$$

with $d = d_A d_B$. Note that $\tau \in \mathcal{R}$ represent *subnormalized* states, since $\operatorname{Tr}\tau \leq 1$ for any $\tau \in \mathcal{R}$. Indeed,

$$\begin{aligned}\operatorname{Tr}\tau &= \operatorname{Tr}(\tau^{T_{\mathsf{B}}}) \\ &\leq \|\tau^{T_{\mathsf{B}}}\|_1 \\ &\leq 1\end{aligned}$$

and a matrix $\tau \in \mathcal{R}$ has $\operatorname{Tr}\tau = 1$ if and only if $\tau^{T_{\mathsf{B}}} \geq 0$. Furthermore, the set $\mathcal{R}$ is convex.

Indeed, for any $\tau, \tau' \in \mathcal{R}$ and $t \in [0,1]$ it is clear that $t\tau + (1-t)\tau' \geq 0$ and we have

$$\begin{aligned}\|(t\tau + (1-t)\tau')^{T_{\mathsf{B}}}\|_1 &= \|t\tau^{T_{\mathsf{B}}} + (1-t)\tau'^{T_{\mathsf{B}}}\|_1 \\ &\leq t\|\tau^{T_{\mathsf{B}}}\|_1 + (1-t)\|\tau'^{T_{\mathsf{B}}}\|_1 \\ &\leq 1,\end{aligned}$$

hence $t\tau + (1-t)\tau' \in \mathcal{R}$.

As the following proposition shows, the value of the Rains bound of a bipartite state $\rho \in \mathrm{H}_{d,+,1} = \mathrm{D}(\mathbb{C}^{d_{\mathsf{A}}} \otimes \mathbb{C}^{d_{\mathsf{B}}})$ can be computed as a convex optimization problem by minimizing the relative entropy of ρ with respect to this convex set $\mathcal{R}$.

Proposition 4.19 ([ADVW02]). *Let $\rho \in \mathrm{H}_{d,+,1}$ be a bipartite density matrix. The Rains bound can be given by the solution to the convex optimization problem*

$$R(\rho) = \min_{\tau \in \mathcal{R}} S(\rho\|\tau), \tag{4.17}$$

where $\mathcal{R} \subseteq \mathrm{H}_{d,+}$ is the set of matrices defined in (4.16).

Proof. Let $\sigma \in \mathrm{H}_{d,+,1}$ and suppose that σ minimizes the Rains bound defined in (4.15). Then $\frac{1}{\|\sigma^{T_{\mathsf{B}}}\|_1}\sigma \in \mathcal{R}$, since $\left\|\left(\frac{1}{\|\sigma^{T_{\mathsf{B}}}\|_1}\sigma\right)^{T_{\mathsf{B}}}\right\|_1 = 1$, and $R(\rho) \geq S(\rho\|\mathcal{R})$, as

$$\begin{aligned}R(\rho) &= S(\rho\|\sigma) + \log\|\sigma^{T_{\mathsf{B}}}\|_1 \\ &= S\left(\rho\middle\|\frac{1}{\|\sigma^{T_{\mathsf{B}}}\|_1}\sigma\right) \\ &\geq \min_{\tau \in \mathcal{R}} S(\rho\|\tau).\end{aligned}$$

On the other hand, suppose that $\tau \in \mathcal{R}$ satisfies $S(\rho\|\tau) \leq S(\rho\|\tau')$ for all other $\tau' \in \mathcal{R}$. Then $\|\tau^{T_{\mathsf{B}}}\| = 1$. Indeed, if it were the case that $\|\tau^{T_{\mathsf{B}}}\| < 1$, then $S(\rho\|\tau) > S\left(\rho\middle\|\frac{\tau}{\|\tau^{T_{\mathsf{B}}}\|_1}\right)$

where $\frac{\tau}{\|\tau^{T_{\mathsf{B}}}\|_1} \in \mathcal{R}$, so τ would not be the minimum. Then $\frac{1}{\operatorname{Tr}\tau}\tau \in \mathrm{H}_{d,+,1}$ and

$$\begin{aligned} S(\rho\|\tau) &= S\left(\rho\middle\|\tfrac{\tau}{\operatorname{Tr}(\tau)}\right) - \log\operatorname{Tr}(\tau) \\ &= S\left(\rho\middle\|\tfrac{\tau}{\operatorname{Tr}(\tau)}\right) + \log\left\|\left(\tfrac{\tau}{\operatorname{Tr}(\tau)}\right)^{T_{\mathsf{B}}}\right\|_1 \\ &\geq R(\rho), \end{aligned}$$

so $S(\rho\|\mathcal{R}) \geq R(\rho)$. Hence $R(\rho) = S(\rho\|\mathcal{R})$, as desired. □

For a given bipartite density matrix $\rho \in \mathrm{H}_{d,+,1}$, we can investigate the necessary and sufficient conditions for a matrix $\tau \in \mathcal{R}$ to minimize the Rains bound. If the state ρ itself is PPT then $\rho \in \mathcal{R}$. Thus the Rains bound vanishes on all PPT states since we may choose $\rho \in \mathcal{R}$ such that $S(\rho\|\mathcal{R}) = S(\rho\|\rho) = 0$. The interesting cases will be for states $\rho \notin \mathcal{R}$ (i.e., states for which $\rho^{T_{\mathsf{B}}} \ngeq 0$).

From the proof of Proposition 4.19, we see that τ minimizes the Rains bound only if $\|\tau^{T_{\mathsf{B}}}\|_1 = 1$, so τ must be on the boundary of $\mathcal{R}$. We discussed in Section 4.2 a method for determining the conditions for when a matrix σ on the boundary of a subset $\mathcal{C} \subseteq \mathrm{H}_{d,+,1}$ of density matrices minimizes some convex function. This is done by characterizing the supporting hyperplanes of the convex set at σ. Similar conditions can be found for minimizing the Rains bound, where the convex set that is minimized over $\mathcal{R}$. The requirements for a matrix in H_d to define a supporting hyperplane of $\mathcal{R}$ are given in 4.4.1, while the necessary and sufficient conditions for a matrix $\tau \in \mathcal{R}$ to minimize $R(\rho)$ for a given state $\rho \in \mathrm{H}_{d,+,1}$ are given in 4.4.2. We will use these results to compare the relative entropy of entanglement and the Rains bound in Section 4.4.3.

4.4.1 Supporting hyperplanes of $\mathcal{R}$

As in the case for arbitrary convex subsets $\mathcal{C} \subseteq \mathrm{H}_{d,+}$, for a matrix $\tau \in \mathcal{R}$ on the boundary, finding all $\rho \in \mathrm{H}_{d,+}$ such that τ minimizes $f_\rho(\tau) = \operatorname{Tr}(\rho f(\tau))$ over τ is reduced to characterizing the supporting hyperplanes of $\mathcal{R}$ at τ. These supporting hyperplanes are given by a matrix $\phi \in \mathrm{H}_d$ satisfying

$$\langle \phi, \tau' \rangle \leq \langle \phi, \tau \rangle \quad \text{for all } \tau' \in \mathcal{R}.$$

Note that the partial transpose of each $\tau \in \mathcal{R}$ is in the unit ball with respect to the trace norm $B_d^1 \subseteq \mathrm{H}_d$, which is defined as

$$B_d^1 = \{X \in \mathrm{H}_d \,|\, \|X\|_1 \leq 1\}.$$

The unit ball is also a convex set and characterizing the supporting functionals of B_d^1 will assist us in finding the supporting hyperplanes of $\mathcal{R}$. The 'dual norm' to the trace norm is the operator norm $\|\cdot\|_\infty$, which is equal to the largest singular value of a matrix. It satisfies $\|XY\|_1 \leq \|X\|_1 \|Y\|_\infty$ for any $X, Y \in \mathrm{H}_d$. The unit ball of the operator norm is

$$B_d^\infty = \{X \in \mathrm{H}_d \,|\, \|X\|_\infty \leq 1\}.$$

It is clear that $\phi \in B_d^\infty$ if and only if $-\mathbb{1}_d \leq \phi \leq \mathbb{1}_d$. Any matrix $\phi \in \mathrm{bd}(B_d^\infty)$ on the boundary of the ball in the operator norm defines a supporting hyperplane of B_d^1, and vice versa. Indeed, if $\phi \in \mathrm{bd}(B_d^\infty)$, then $\|\phi\|_\infty = 1$, and for any $X \in B_d^1$ it holds that

$$\begin{aligned} \langle \phi, X \rangle = \mathrm{Tr}(\phi X) &\leq \|\phi X\|_1 \\ &\leq \|\phi\|_\infty \|X\|_1 \\ &= \|X\|_1 \\ &\leq 1, \end{aligned}$$

with equality if and only if $\|X\|_1 = 1$ and $\phi X = |X|$. The following lemma shows what form ϕ must have for equality to hold. For an operator $X \in \mathrm{H}_d$, the projection operators $P_+, P_- \in \mathrm{H}_d$ onto the positive and negative eigenspaces of X are defined as the unique projection operators on the the supports of X_+ and X_-, respectively. These projections satisfy $P_+XP_+ = X_+$ and $P_-XP_- = -X_-$.

Lemma 4.20. *Let $X \in \mathrm{H}_d$ and let $\phi \in \mathrm{H}_d$ such that $-\mathbb{1}_d \leq \phi \leq \mathbb{1}_d$. It holds that*

$$\langle \phi, X \rangle \leq \|X\|_1 \tag{4.18}$$

with equality if and only if $\phi = P_+ - P_- + Q$, where P_+ and P_- are the projection operators onto the positive and negative eigenspaces of X, and $-\mathbb{1}_d \leq Q \leq \mathbb{1}_d$ is an operator orthogonal to both P_+ and P_- (i.e., $QP_+ = QP_- = 0$).

Proof. From the observations above, all that remains is to prove the conditions for equality. If $\phi = P_+ - P_- + Q$ then $\langle \phi, X \rangle = \mathrm{Tr}|X| = \|X\|_1$. On the other hand, if equality holds then it must be the case that $\phi X = |X|$. It follows that $(\phi - \mathbb{1}_d)X_+ = 0$ and $(\phi + \mathbb{1}_d)X_- = 0$, as desired. □

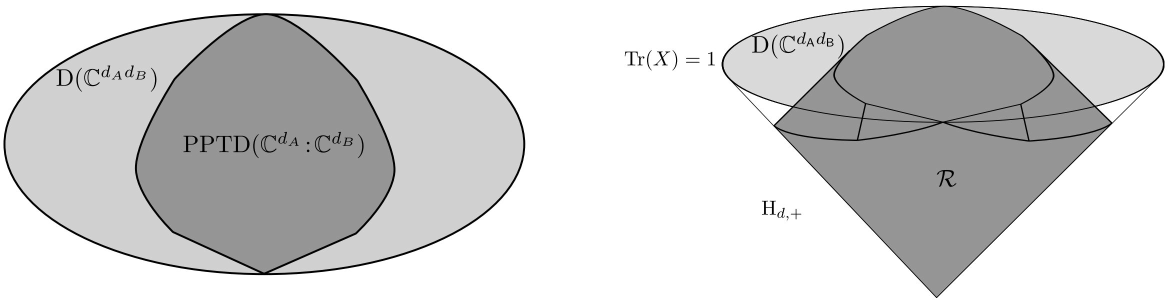
Theorem 4.21. *Let $X \in \mathrm{H}_d$ such that $\|X\|_1 = 1$ (i.e., X is on the boundary of $B_d^1 \subseteq \mathrm{H}_d$). An operator $\phi \in \mathrm{H}_d$ defines a supporting hyperplane of B_d^1 at X such that*

$$\langle \phi, Y \rangle \leq \langle \phi, X \rangle = 1 \quad \textit{for all } Y \in B_d^1 \tag{4.19}$$

if and only if $\phi = P_+ - P_- + Q$ where P_+ and P_- are the projection operators onto the positive and negative eigenspaces of X and $-\mathbb{1}_d \leq Q \leq \mathbb{1}_d$ is an operator orthogonal to both P_+ and P_-. In particular, the supporting hyperplane of this form at X is unique if and only if X has no zero eigenvalues.

Proof. This follows directly from Lemma 4.20. If X is non-singular, then $P_+ + P_- = \mathbb{1}_d$, and the only $\phi = P_+ - P_- + Q$ of this form has $Q = 0$; therefore the supporting hyperplane is unique. □

Let $\tau \in \mathcal{R}$ be on the boundary with $\|\tau^{T_\mathsf{B}}\|_1 = 1$. A matrix $\phi \in \mathrm{H}_d$ defines a supporting hyperplane of $\mathcal{R}$ at τ whenever ϕ^{T_B} defines a supporting hyperplane of B_d^1 at τ^{T_B}. Indeed,


D($\mathbb{C}^{d_A d_B}$)
PPTD($\mathbb{C}^{d_A}:\mathbb{C}^{d_B}$)
Tr$(X) = 1$
D($\mathbb{C}^{d_\mathsf{A} d_\mathsf{B}}$)
$\mathcal{R}$
$\mathrm{H}_{d,+}$

(a) The convex subset $\mathcal{P} = \mathrm{PPTD}(\mathbb{C}^{d_\mathsf{A}}:\mathbb{C}^{d_\mathsf{B}})$ of PPT density matrices as a subset of all density matrices.

(b) Convex subset $\mathcal{R} = \{\tau \geq 0 \,|\, \|\tau^{T_\mathsf{B}}\|_1 \leq 1\}$ as a subset of the positive cone. The intersection of $\mathcal{R}$ with the density matrices is exactly $\mathcal{P}$.

Figure 4.3: Depiction of $\mathcal{P}$ and $\mathcal{R}$ as convex subsets of positive matrices.

suppose that $\langle \phi^{T_\mathsf{B}}, X\rangle \leq \langle \phi^{T_\mathsf{B}}, \tau^{T_\mathsf{B}}\rangle$ for all $X \in B_d^1$. For all $\tau' \in \mathcal{R}$, it holds that

$$\begin{aligned} \langle \phi, \tau' \rangle &= \langle \phi^{T_\mathsf{B}}, \tau'^{T_\mathsf{B}} \rangle \\ &\leq \langle \phi^{T_\mathsf{B}}, \tau^{T_\mathsf{B}} \rangle, \end{aligned}$$

since $\tau'^{T_\mathsf{B}} \in B_d^1$. The converse is not necessarily true, but it does hold if τ is full rank. In the case when τ is full rank, a matrix $\phi \in \mathrm{H}_d$ defines a supporting hyperplane of $\mathcal{R}$ at τ *if and only if* ϕ^{T_B} defines a supporting hyperplane of B_d^1 at τ^{T_B}.

Depictions of the set $\mathcal{P}$ of PPT density matrices and the set $\mathcal{R}$ are shown in Figure 4.3. Note that this set can be written as the intersection $\mathcal{R} = \mathrm{H}_{d,+} \cap (B_d^1)^{T_\mathsf{B}}$ of the positive cone and the partial transpose of the unit ball B_d^1.

4.4.2 Criterion for minimization of the Rains bound

We can now state the necessary and sufficient conditions for a matrix $\tau \in \mathcal{R}$ to minimize the Rains bound of a state ρ. Let $\rho \in \mathrm{D}(\mathbb{C}^{d_\mathsf{A}} \otimes \mathbb{C}^{d_\mathsf{B}})$ be a bipartite state, and let $\tau \in \mathcal{R}$ such that $R(\rho) = S(\rho\|\tau)$ (i.e., τ minimizes the relative entropy of ρ with respect to $\mathcal{R}$). The Rains bound $R(\rho)$ vanishes exactly for PPT states ρ (in which case the matrix that minimizes $R(\rho)$ is $\tau = \rho$), so we only need to consider states ρ for which $\rho^{T_\mathsf{B}} \not\geq 0$. The following lemma is simply an application of the general case in Lemma 4.10 to the specific convex subset $\mathcal{R}$ used in the definition of the Rains bound.

Lemma 4.22. *Let $\rho \in \mathrm{H}_{d,+,1}$ be a non-PPT state and let $\tau \in \mathcal{R}$ be on the boundary with $\|\tau^{T_{\mathsf{B}}}\|_1 = 1$. Then $R(\rho) = S(\rho\|\tau)$ if and only if it holds that $\langle L_\tau(\rho), \tau\rangle = 1$ and*

$$\langle L_\tau(\rho), \tau'\rangle \leq 1$$

for all $\tau' \in \mathcal{R}$.

Proof. This follows directly from Lemma 4.10, since $\operatorname{Tr}\rho = 1$. □

Hence, as before, the conditions for minimizing the Rains bound are given in terms of supporting hyperplanes of $\mathcal{R}$. Given a matrix $\tau \in \mathcal{R}$ with $\|\tau^{T_{\mathsf{B}}}\|_1 = 1$, it is now possible to find all states ρ for which τ minimizes the Rains bound by finding all supporting hyperplanes of $\mathcal{R}$ at τ. All supporting hyperplanes considered here will have the form

$$\langle \phi, \tau'\rangle \leq \langle \phi, \tau\rangle = 1 \qquad \text{for all } \tau' \in \mathcal{R}. \tag{4.20}$$

The main result of this section states the conditions for which a matrix $\tau \in \mathcal{R}$ minimizes the Rains bound for some density matrix ρ. The proof of Theorem 4.25 will rely on a few facts about Schur products of matrices and monotone matrix functions, which are the Schur product theorem (see e.g., Theorem 7.5.3 in [HJ13]) and the fact that the matrix of divided differences $f^{[1]}(\Lambda)$ is positive if f is a monotone matrix function.

Theorem 4.23 (Schur product theorem)**.** *For any two matrices $X, Y \in \mathrm{H}_d$, if $X, Y \geq 0$, then $X \circ Y \geq 0$. Furthermore, if $Y > 0$ and X has no diagonal entry equal to 0, then $X \circ Y > 0$.*

Theorem 4.24 (see, e.g., Theorem 6.6.36 in [HJ94])**.** *Let $a, b \in \mathbb{R}$ and let $f : (a, b) \to \mathbb{R}$ be a continuously differentiable function. Let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ with all eigenvalues $\lambda_i \in (a, b)$. If f is operator monotone, then $f^{[1]}(\Lambda) \geq 0$.*

The function $\log : (0, \infty) \to \mathbb{R}$ is strictly operator monotone as a function on matrices. For diagonal $\tau > 0$, we have $L_\tau(X) = f^{[1]}(\tau) \circ X$, where f is the function $f(x) = \log x$. Since $f^{[1]}(\tau) > 0$ for any $\tau > 0$ by Theorem 4.24, it follows that $L_\tau(X) > 0$ for any $X > 0$. This generalizes easily to non-diagonal τ as well. We can now state the main result of this section:

Theorem 4.25. *Let $\tau \in \mathcal{R}$ with $\|\tau^{T_{\mathsf{B}}}\|_1 = 1$. There exists a state $\rho \in \mathrm{H}_{d,+,1}$ such that τ minimizes the Rains bound for ρ if and only if there exists a $\phi \in \mathrm{H}_d$ with $\operatorname{supp}(\phi) \subseteq \operatorname{supp}(\tau)$ that defines a supporting hyperplane of $\mathcal{R}$ at τ of the form in* (4.20)*, such that*

$$L_\tau^\ddagger(\phi) \geq 0$$

and $\rho = L_\tau^\ddagger(\phi)$. Furthermore $\phi \geq 0$, and if $\rho > 0$ then $\tau > 0$ and $\phi > 0$.

Proof. First assume there exists a state ρ such that τ minimizes the Rains bound for ρ. Define $\phi = L_\tau(\rho)$. Then ϕ defines a supporting hyperplane of the form in (4.20) and satisfies the desired conditions (i.e., $\operatorname{supp}(\phi) \subseteq \operatorname{supp}(\tau)$ and $L_\tau^\ddagger(\phi) \geq 0$ since $\rho = L_\tau^\ddagger(\phi)$ and ρ is a density matrix).

On the other hand, assume that there exists a matrix $\phi \in \mathrm{H}_d$ that defines a supporting hyperplane of $\mathcal{R}$ at τ of the form in (4.20) such that $\operatorname{supp}(\phi) \subseteq \operatorname{supp}(\tau)$ and $L_\tau^\ddagger(\phi) \geq 0$. Let $\rho = L_\tau^\ddagger(\phi)$. It follows that ρ is a valid density matrix, since $\rho \geq 0$ and

$$\begin{aligned} \operatorname{Tr} \rho &= \langle \rho, \mathbb{1}_d \rangle \\ &= \langle L_\tau^\ddagger(\phi), \mathbb{1}_d \rangle \\ &= \langle \phi, \tau \rangle = 1 \end{aligned}$$

where $L_\tau^\ddagger(\mathbb{1}_d) = \tau$. Therefore, τ minimizes the Rains bound for ρ since $L_\tau(\rho)$ defines a supporting hyperplane of the necessary form.

We now check the positivity of $\phi = L_\tau(\rho)$. We may assume without loss of generality that τ is diagonal. Let $f : (0, \infty) \to \mathbb{R}$ be the function $f(x) = \log x$. We first suppose that $\tau > 0$. Then $L_\tau(\rho)$ is given by

$$L_\tau(\rho) = f^{[1]}(\tau) \circ \rho,$$

where $f^{[1]}(\tau)$ is the matrix of divided differences of the eigenvalues of τ. Then $f^{[1]}(\tau) \geq 0$ and has no diagonal entry equal to 0, since f is strictly operator monotone, and thus $f^{[1]}(\tau) \circ \rho > 0$ by Theorem 4.23. Hence $\phi > 0$, since $\phi = L_\tau(\rho) = f^{[1]}(\tau) \circ \rho$. If τ is singular, we may assume τ and ρ have the form

$$\tau = \begin{pmatrix} \tilde{\tau} & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \rho = \begin{pmatrix} \tilde{\rho} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\tilde{\tau} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ is the upper block of τ such that $\tilde{\tau} > 0$, and $\tilde{\rho} \geq 0$. Then $L_\tau(\rho)$ has the form

$$L_\tau(\rho) = \begin{pmatrix} L_{\tilde{\tau}}(\tilde{\rho}) & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} f^{[1]}(\tilde{\tau}) \circ \tilde{\rho} & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that $L_{\tilde{\tau}}(\tilde{\rho}) \geq 0$, since $\rho \geq 0$ and $\tilde{\tau} > 0$. It follows that $\phi = L_\tau(\rho) \geq 0$ as desired. □

Hence, for a given $\tau \in \mathcal{R}$ with $\|\tau^{T_B}\|_1$, finding all density matrices ρ such that τ minimizes the Rains bound for ρ amounts to finding all supporting hyperplanes ϕ of $\mathcal{R}$ at τ satisfying

$$\langle \phi, \tau' \rangle \leq \langle \phi, \tau \rangle = 1 \qquad \text{for all } \tau' \in \mathcal{R}$$

and $L_\tau^\ddagger(\phi) \geq 0$. If ϕ is a supporting hyperplane of this form, $\rho = L_\tau^\ddagger(\phi)$ is a density matrix and a closed formula for the Rains bound of this state can be given by $R(\rho) = S(\rho \| \tau)$, which can be written as

$$\begin{aligned} R(\rho) = S(\rho\|\tau) &= S(L_\tau^\ddagger(\phi)\|\tau) \\ &= \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(L_\tau^\ddagger(\phi) \log \tau) \\ &= \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\phi L_\tau^\ddagger(\log \tau)) \\ &= \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\phi \tau \log \tau). \end{aligned}$$

Conversely, if there are no supporting hyperplanes of $\mathcal{R}$ at τ such that $L_\tau^\ddagger(\phi) \geq 0$ then τ cannot minimize the Rains bound for *any* state ρ. The following section applies this to states of bipartite systems where one of the systems is a qubit.

4.4.3 Bipartite systems where one system is a qubit

It has been shown in [MI08] that for a state of two qubits, the Rains bound is equal to the relative entropy of entanglement. This can be generalized to the bipartite case where at least one of the subsystems is a qubit. Here, we may assume that $\rho \in \mathrm{D}(\mathbb{C}^{d_\mathsf{A}} \otimes \mathbb{C}^{d_\mathsf{B}})$ with $d_\mathsf{A} = 2$ and $d_\mathsf{B} \geq 2$. For all states of such systems, the Rains bound is equal to the relative entropy of entanglement.

Theorem 4.26. *For states ρ of bipartite systems where at least one of the subsystems is a qubit, the Rains bound is equal to the relative entropy of entanglement (i.e., $R(\rho) = E_R^{\mathrm{PPT}}(\rho)$).*

Proof. Let ρ be a bipartite state for which at least one of the subsystems is a qubit. Without loss of generality, we may consider ρ to be of the form $\rho \in \mathrm{D}(\mathbb{C}^2 \otimes \mathbb{C}^{d_\mathsf{B}})$ and we define $d = 2d_\mathsf{B}$ (where $d_\mathsf{A} = 2$). We may assume that $\rho > 0$. Indeed, since E_R^{PPT} and R are continuous [Rai01], it will be satisfied for all $\rho \geq 0$ as well. Using an argument by contradiction, we will show that, if $\tau \in \mathcal{R}$ minimizes the Rains bound for ρ, then $\tau^{T_\mathsf{B}} \geq 0$. That is, τ is PPT and thus $E_R^{\mathrm{PPT}}(\rho) = R(\rho)$.

Suppose for the sake of obtaining a contradiction that there exists a $\tau \in \mathcal{R}$ with $\tau^{T_\mathsf{B}} \ngeq 0$

that minimizes the Rains bound for ρ. We will first consider the case where $\tau^{T_{\mathsf{B}}}$ has only one negative eigenvalue and the remaining eigenvalues are positive. The more general case (where $\tau^{T_{\mathsf{B}}}$ has non-trivial null space and at least one negative eigenvalue) will be considered later and similar techniques will be used.

First suppose that $\tau^{T_{\mathsf{B}}}$ is full-rank and that the dimension of the negative eigenspace is one. Let P_+ and P_- be the projection operators onto the positive and negative eigenspaces of τ. Then P_- is the rank-one projector onto the negative eigenspace of $\tau^{T_{\mathsf{B}}}$ and $P_+ + P_- = \mathbb{1}_d$. Thus $\phi = (P_+ - P_-)^{T_{\mathsf{B}}} = \mathbb{1}_d - 2P_-^{T_{\mathsf{B}}}$ is the unique operator defining a supporting hyperplane of $\mathcal{R}$ at τ of the form

$$\langle \phi, \tau' \rangle \leq \langle \phi, \tau \rangle = 1 \qquad \text{for all } \tau' \in \mathcal{R}. \tag{4.21}$$

We now show that the largest eigenvalue of $P_-^{T_{\mathsf{B}}}$ is greater than $\frac{1}{2}$, and thus $\phi = \mathbb{1}_d - 2P_-^{T_{\mathsf{B}}} \not\geq 0$. This is a contradiction to the last statement of the previous theorem (Theorem 4.25), namely that $\phi > 0$ whenever τ minimizes the Rains bound for ρ if $\rho > 0$.

Since P_- has rank one, we can write this as $P_- = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is the eigenvector corresponding the negative eigenvalue of $\tau^{T_{\mathsf{B}}}$ and can be written as

$$|\psi\rangle = |0\rangle \otimes |u\rangle + |1\rangle \otimes |v\rangle,$$

where $\langle\psi|\psi\rangle = \langle u|u\rangle + \langle v|v\rangle = 1$, and $|u\rangle, |v\rangle \in \mathbb{C}^{d_{\mathsf{B}}}$ are unnormalized vectors. We may write P_- and $P_-^{T_{\mathsf{B}}}$ in matrix form as

$$P_- = \begin{pmatrix} |u\rangle\langle u| & |u\rangle\langle v| \\ |v\rangle\langle u| & |v\rangle\langle v| \end{pmatrix} \qquad \text{and} \qquad P_-^{T_{\mathsf{B}}} = \begin{pmatrix} |u\rangle\langle u| & |v\rangle\langle u| \\ |u\rangle\langle v| & |v\rangle\langle v| \end{pmatrix}.$$

By Cauchy's interlacing theorem (see e.g., Theorem 4.3.28 in [HJ13]), the maximal eigenvalue of $P_-^{T_{\mathsf{B}}}$ is at least as great as the largest eigenvalue among the two submatrices $|u\rangle\langle u|$ and $|v\rangle\langle v|$. These are both rank-one positive matrices whose only nonzero eigenvalues are $\langle u|u\rangle$ and $\langle v|v\rangle$. Since these values must sum to unity, it holds that the largest eigenvalue of $P_-^{T_{\mathsf{B}}}$ is at least as large as $\lambda_{\max}(P_-^{T_{\mathsf{B}}}) \geq \max\{\langle u|u\rangle, \langle v|v\rangle\} \geq \frac{1}{2}$, and thus $\mathbb{1}_d \not\geq 2P_-^{T_{\mathsf{B}}}$. It follows that $\phi = \mathbb{1}_d - 2P_-^{T_{\mathsf{B}}} \not\geq 0$, and thus τ does not minimize the Rains bound for ρ, a contradiction.

We now consider the general case when $\tau^{T_{\mathsf{B}}}$ has at least one negative eigenvalue. Again, let P_+ and P_- be the projection matrices onto the positive and negative eigenspaces of $\tau^{T_{\mathsf{B}}}$, but it is no longer necessarily the case that $P_+ + P_- = \mathbb{1}_d$ or that P_- is rank one. By Lemma 4.20, any operator ϕ that defines a supporting hyperplane of $\mathcal{R}$ at τ of the form in (4.21) must be of the form $\phi = (P_+ - P_- + Q)^{T_{\mathsf{B}}}$ for some $-\mathbb{1}_d \leq Q \leq \mathbb{1}_d$ that is orthogonal to both P_+ and P_- (i.e., Q is supported on the null space of $\tau^{T_{\mathsf{B}}}$). Define P_0 as the projection matrix onto the null space of $\tau^{T_{\mathsf{B}}}$ such that $P_0 Q P_0 = Q$, and P_+, P_-, and P_0 form a complete set of projectors such that $P_+ + P_- + P_0 = \mathbb{1}_d$. Then we may write $\phi^{T_{\mathsf{B}}}$ as

$$\begin{aligned}
\phi^{T_{\mathsf{B}}} &= P_+ - P_- + Q \\
&= P_+ + P_- + P_0 - (2P_- + P_0 - Q) \\
&= \mathbb{1}_d - X,
\end{aligned}$$

where we define $X = 2P_- + P_0 - Q$. Note that $P_0 \geq Q$ and thus $X \geq 0$ since $P_- \geq 0$. Let $|\psi\rangle = |0\rangle \otimes |u\rangle + |1\rangle \otimes |v\rangle$ be a normalized eigenvector of $\tau^{T_{\mathsf{B}}}$ with negative eigenvalue. We may write X and $|\psi\rangle\langle\psi|$ in matrix form as

$$X = \begin{pmatrix} Y & W \\ W^\dagger & Z \end{pmatrix} \qquad \text{and} \qquad |\psi\rangle\langle\psi| = \begin{pmatrix} |u\rangle\langle u| & |u\rangle\langle v| \\ |v\rangle\langle u| & |v\rangle\langle v| \end{pmatrix}$$

for some matrices $W, Y, Z \in \mathrm{H}_{d_{\mathsf{B}}}$. Note that $P_- \geq |\psi\rangle\langle\psi|$ since $|\psi\rangle$ is in the negative eigenspace of $\tau^{T_{\mathsf{B}}}$. Thus $X \geq 2|\psi\rangle\langle\psi|$ since $X \geq 2P_-$. Hence it must hold that $Y \geq 2|u\rangle\langle u|$ and $Z \geq 2|v\rangle\langle v|$, and thus the largest eigenvalues of Y and Z are at least as large as $2\langle u|u\rangle$ and $2\langle v|v\rangle$ respectively. As in the previous case, we have $\langle u|u\rangle + \langle v|v\rangle = 1$ and thus the largest eigenvalue among all of the eigenvalues of Y and Z must be at least 1. Since

$$X^{T_{\mathsf{B}}} = \begin{pmatrix} Y & W^\dagger \\ W & Z \end{pmatrix},$$

by the Cauchy interlacing theorem, the largest eigenvalue of $X^{T_{\mathsf{B}}}$ is at least 1. It follows that $\phi = \mathbb{1}_d - X^{T_{\mathsf{B}}} \not\geq 0$, and thus τ cannot minimize the Rains bound for ρ, a contradiction.

Therefore any $\tau \in \mathcal{R}$ that minimizes the Rains bound for ρ must satisfy $\tau^{T_{\mathsf{B}}} \geq 0$. This completes the proof that $R(\rho) = E_R^{\mathrm{PPT}}(\rho)$ for any bipartite state ρ for which at least one subsystem is a qubit. □

The statement and proof of Theorem 4.26 generalizes to the multipartite case where at least one subsystem is a qubit.

4.5 Rényi relative entropies

Let $\rho, \sigma \in \mathrm{H}_{d,+}$ be positive semidefinite. For $\alpha \in (0, \infty)$ with $\alpha \neq 1$, recall the *quantum Rényi α-relative entropy* (or *Rényi α-divergence*) of ρ with respect to σ is defined as

$$S_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha}). \tag{4.22}$$

For $\alpha < 1$, this is always well-defined for all $\rho, \sigma \geq 0$ if $\mathrm{supp}(\rho) \not\perp \mathrm{supp}(\sigma)$. If $\alpha > 1$, then $\mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha})$ is finite as long as $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. Otherwise we may define $S_\alpha(\rho\|\sigma) = +\infty$. The standard relative entropy is recovered in the limit $\alpha \to 1$, so we simply define $S_1(\rho\|\sigma) = S(\rho\|\sigma)$. In the limit $\alpha \to 0$, we have $\rho^\alpha \to P_\rho$ (where P_ρ is just the projection operator onto the support of ρ), and thus $\mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha}) \to \langle P_\rho, \sigma \rangle$ in this limit, which is just the trace of σ projected onto the support of ρ. Hence the 0-Rényi divergence reduces to

$$S_0(\rho\|\sigma) = -\log\langle P_\rho, \sigma \rangle$$

which is finite as long as $\mathrm{supp}(\rho) \not\perp \mathrm{supp}(\sigma)$.

The α-relative entropies are convex in the second argument for all $\alpha \in [0, 2]$ [MD09], but are jointly convex only for $\alpha \in [0, 1]$ [Pet86]. Furthermore, for $\alpha \in [0, 2]$ they are also monotonic under completely positive and trace preserving maps [MD09], i.e.

$$S(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \leq S(\rho \| \sigma)$$

for any quantum channel $\mathcal{E}$. If ρ and σ are density matrices, then it holds that $S_\alpha(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$. These pseudo-distance properties of the Rényi α-relative entropies make it an interesting to study. In particular, it can be used to define a resource monotone when the set of free states is convex. Furthermore, since $S_\alpha(\rho\|\sigma)$ is convex in σ as long as $\alpha \in [0, 2]$, the methods presented in this chapter can be used for finding necessary and sufficient conditions for minimizing $S_\alpha(\rho\|\mathcal{C})$ over convex sets $\mathcal{C}$ of density matrices, which will be done in this section.

Analogous to the relative entropy of entanglement, we can also define the *α-relative entropy of entanglement* of a bipartite density matrix ρ as

$$E_{R,\alpha}^{\mathrm{Sep}}(\rho) := \inf_{\sigma \in \mathrm{SepD}} S_\alpha(\rho\|\sigma) = S_\alpha(\rho\|\mathrm{SepD}). \tag{4.23}$$

For $\alpha \in [0, 2]$, this quantity is a faithful entanglement measure. In this section, it will be shown that $E_{R,\alpha}^{\mathrm{Sep}}(\psi) = E_{1/\alpha}(\psi)$ for all bipartite pure states, where $E_\alpha(\psi) = \frac{1}{\alpha} \log \sum_i \lambda_i^\alpha$ is the Rényi α-entropy measure of entanglement for pure states (and $\vec{\lambda}$ is the vector of Schmidt coefficients of $|\psi\rangle$).

4.5.1 Conditions for minimizing α-relative entropy

We first analyze how the general case of Lemma 4.6 applies to the Rényi α-entropies. Given a fixed density matrix ρ, the function $\sigma \mapsto S_\alpha(\rho\|\sigma)$ is convex when $\alpha \in [0, 2]$. We may use the methods of the previous sections to derive the following necessary and sufficient conditions for a matrix σ in a convex subset $\mathcal{C}$ of positive semidefinite matrices to minimize $S_\alpha(\rho\|\mathcal{C})$. In all cases considered after this, $\mathcal{C}$ is a subset of density matrices, but Theorem 4.27 considers the full generality.

Theorem 4.27. *Let $\alpha \in [0,2]$ with $\alpha \neq 1$, let $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ be convex, let $\sigma \in \mathcal{C}$, and let $\rho \in \mathrm{H}_{d,+}$ such that* $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. *The following are equivalent.*

1. *It holds that $S_\alpha(\rho\|\mathcal{C}) = S_\alpha(\rho\|\sigma)$ (i.e., σ minimizes the α-relative entropy of ρ with respect to $\mathcal{C}$).*

2. *For all $\sigma' \in \mathcal{C}$ it holds that*

$$\frac{1}{\alpha-1}\langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma'\rangle \geq \frac{1}{\alpha-1}\langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma\rangle. \tag{4.24}$$

That is, the operator $\mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha)$ defines a supporting hyperplane of $\mathcal{C}$ at σ.

Furthermore, it holds that $\langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma\rangle = (1-\alpha)\,\mathrm{Tr}(\rho^\alpha\sigma^{1-\alpha})$.

Proof. By convexity, it holds that $S_\alpha(\rho\|\mathcal{C}) = S_\alpha(\rho\|\sigma)$ if and only if

$$\frac{d}{dt}S_\alpha\Big(\rho\Big\|(1-t)\sigma + t\sigma'\Big)\bigg|_{t=0^+} \geq 0$$

holds for all other $\sigma' \in \mathcal{C}$. For any $\sigma' \in \mathrm{H}_{d,+}$, we have

$$\begin{aligned}
\frac{d}{dt}S_\alpha\Big(\rho\Big\|(1-t)\sigma + t\sigma'\Big)\bigg|_{t=0^+} &= \frac{1}{\alpha-1}\,\frac{d}{dt}\log\mathrm{Tr}\Big[\rho^\alpha\Big(\sigma + t(\sigma'-\sigma)\Big)^{1-\alpha}\Big]\bigg|_{t=0^+} \\
&= \frac{1}{\alpha-1}\frac{1}{\mathrm{Tr}(\rho^\alpha\sigma^{1-\alpha})}\,\frac{d}{dt}\,\mathrm{Tr}\Big[\rho^\alpha\Big(\sigma + t(\sigma'-\sigma)\Big)^{1-\alpha}\Big]\bigg|_{t=0^+} \\
&= \frac{1}{\alpha-1}\frac{1}{\mathrm{Tr}(\rho^\alpha\sigma^{1-\alpha})}\,\mathrm{Tr}\left[\rho^\alpha\mathfrak{D}_{f_\alpha,\sigma}(\sigma'-\sigma)\right] \\
&= \frac{1}{\alpha-1}\frac{1}{\mathrm{Tr}(\rho^\alpha\sigma^{1-\alpha})}\Big(\langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma'\rangle - \langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma\rangle\Big),
\end{aligned}$$

which yields the desired result. Finally, note that $\mathfrak{D}_{f_\alpha,\sigma}(\sigma) = \sigma f'_\alpha(\sigma) = (1-\alpha)\sigma^{1-\alpha}$ and thus

$$\begin{aligned}\langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma\rangle &= \langle \rho^\alpha, \mathfrak{D}_{f_\alpha,\sigma}(\sigma)\rangle \\ &= (1-\alpha)\operatorname{Tr}(\rho^\alpha\sigma^{1-\alpha}).\end{aligned}$$

This completes the proof. □

If $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ is a convex subset of density matrices, then $\langle \mathbb{1}_d, \sigma\rangle = \operatorname{Tr}(\sigma') = 1$ for all $\sigma' \in \mathcal{C}$, and the condition in (4.24) is equivalent to the condition that $\langle \xi, \sigma'\rangle \geq 0$, where we define the matrix $\xi \in \mathrm{H}_d$ as

$$\xi = \mathbb{1}_d - \frac{1}{(1-\alpha)\operatorname{Tr}(\rho^\alpha\sigma^{1-\alpha})}\mathfrak{D}_{f_\alpha,\sigma}. \tag{4.25}$$

This is equivalent to saying that $\mathcal{C}$-witness. That is, if $\langle \xi, \sigma'\rangle < 0$ for some σ', then $\sigma' \notin \mathcal{C}$. Hence, the condition that $S_\alpha(\rho\|\mathcal{C}) = S_\alpha(\rho\|\sigma)$ is equivalent to the condition that

$$\langle \xi, \sigma'\rangle = 1 - \frac{1}{1-\alpha}\frac{1}{\operatorname{Tr}(\rho^\alpha\sigma^{1-\alpha})}\langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma'\rangle \geq 0 \tag{4.26}$$

hold for all $\sigma' \in \mathcal{C}$. As we shall see, this will be particularly useful for computing the α-relative entropy of entanglement of pure states.

4.5.2 Rényi relative entropy of entanglement for pure states

Here, we compute the Rényi relative entropy of entanglement $E_{R,\alpha}^{\mathrm{Sep}}$ for pure states. The case when $\alpha = 1$ was shown earlier, so only the cases when $\alpha \in [0,1) \cup (1,2]$ will be considered. Note that $|\psi\rangle\langle\psi|^\alpha = |\psi\rangle\langle\psi|$ for any $\alpha \geq 0$ and any pure state $|\psi\rangle$. Hence, for pure states $\rho = |\psi\rangle\langle\psi|$ and $\alpha \neq 1$ the divergence in (4.22) reduces to

$$S_\alpha(\psi\|\sigma) = \frac{1}{\alpha-1}\log\langle\psi|\sigma^{1-\alpha}|\psi\rangle. \tag{4.27}$$

For any $\alpha \in (0, \infty)$, the *Rényi α-entropy of entanglement* of a pure state $|\psi\rangle = \sum_i \sqrt{\lambda_i}|ii\rangle$ is defined by

$$\begin{aligned} E_\alpha(\psi) &= \frac{1}{1-\alpha} \log\Big(\sum_i \lambda_i^\alpha\Big) \\ &= \frac{1}{1-\alpha} \log\|\vec{\lambda}\|_\alpha^\alpha \end{aligned}$$

where $\vec{\lambda}$ is the vector of Schmidt coefficients of $|\psi\rangle$ and $\|\vec{\lambda}\|_\alpha = \Big(\sum_i \lambda_i^\alpha\Big)^{1/\alpha}$. We may define $E_\infty(\psi) = -\log(\max(\lambda_i))$

Theorem 4.28. *Let $\alpha \in [0, 2]$. For bipartite pure states $|\psi\rangle$ it holds that*

$$E_{R,\alpha}^{\text{Sep}}(\psi) = E_{1/\alpha}(\psi), \tag{4.28}$$

where we define $E_{1/0} = E_\infty$.

Here, $E_{1/\alpha}(\psi)$ is the α-entropy of entanglement of ψ of order $1/\alpha$. As before, it is helpful to first state the following technical lemma. For any $\alpha \in \mathbb{R}$, define the function $f_\alpha : (0, +\infty) \to \mathbb{R}$ by

$$f_\alpha(x) = x^{1-\alpha},$$

and we define $0^r = 0$ for any $r \in \mathbb{R}$. The statement and proof of Lemma 4.29 are analogous to the those of Lemma 4.18.

Lemma 4.29. *Let $\alpha \in (0, 1) \cup (1, 2]$.*

(a) For all $x, y > 0$ with $x \neq y$, it holds that

$$0 < \frac{1}{1-\alpha} \frac{x^{1-\alpha} - y^{1-\alpha}}{x-y} \leq \frac{1}{\sqrt{xy}^\alpha}. \tag{4.29}$$

(b) For all $p, q > 0$, it holds that

$$0 < \frac{\sqrt{pq}}{1-\alpha} f_\alpha^{[1]}(p^{1/\alpha}, q^{1/\alpha}) \leq 1. \tag{4.30}$$

Proof. We first prove part (a). Positivity of the expression in (4.29) is clear, since f_α is strictly decreasing for $\alpha < 1$ and strictly increasing for $\alpha > 1$ on $(0, +\infty)$. In the case when $\alpha \in (0,1) \cup (1,2)$, we may make use of the following integral representations. For $r \in (-1, 1)$ and any $x, y > 0$ with $x \neq y$, it holds that

$$rx^{r-1} = \frac{\sin(r\pi)}{\pi} \int_0^\infty \frac{t^r}{(x+t)^2} dt \tag{4.31}$$

$$\text{and} \quad \frac{x^r - y^r}{x - y} = \frac{\sin(r\pi)}{\pi} \int_0^\infty \frac{t^r}{(x+t)(y+t)} dt. \tag{4.32}$$

For completeness, a proof of these integral representations is included in the appendix. Since $\alpha - 1 \in (-1, 1)$, we may use these integral representations to find that

$$\begin{aligned}
\frac{1}{1-\alpha} \frac{x^{1-\alpha} - y^{1-\alpha}}{x - y} &= \frac{\sin((1-\alpha)\pi)}{(1-\alpha)\pi} \int_0^\infty \frac{t^{1-\alpha}}{(x+t)(y+1)} dt \\
&\leq \frac{\sin((1-\alpha)\pi)}{(1-\alpha)\pi} \int_0^\infty \frac{t^{1-\alpha}}{(\sqrt{xy}) + t)^2} dt \\
&= \frac{1}{\sqrt{xy}^\alpha}.
\end{aligned}$$

The inequality in (4.29) holds with equality in the case when $\alpha = 2$, since in this case

$$\begin{aligned}
\frac{1}{1-\alpha} \frac{x^{1-\alpha} - y^{1-\alpha}}{x - y} &= -\frac{\frac{1}{x} - \frac{1}{y}}{x - y} \\
&= \frac{1}{xy} \qquad = \frac{1}{\sqrt{xy}^\alpha},
\end{aligned}$$

which completes the proof of part (a).

To show (b), note that the divided differences of f_α are given by

$$f_\alpha^{[1]}(x,y) = \begin{cases} \frac{x^{1-\alpha}-y^{1-\alpha}}{x-y}, & x \neq y \\ (1-\alpha)x^{-\alpha}, & x = y. \end{cases}$$

As before, positivity of the expression in (4.30) is clear. To show that the expression in (4.30) is at most 1, we consider the cases when $p = q$ and $p \neq q$ separately. If $p = q$, we have $f_\alpha^{[1]}(p^{1/\alpha}, p^{1/\alpha}) = (1-\alpha)(p^{1/\alpha})^{-\alpha} = \frac{1-\alpha}{p}$, and the expression in (4.30) reduces to

$$\frac{\sqrt{pq}}{1-\alpha} f_\alpha^{[1]}(p^{1/\alpha}, q^{1/\alpha}) = \frac{p}{1-\alpha}\frac{1-\alpha}{p} = 1,$$

so the inequality holds with equality. In the case where $p \neq q$, the result follows directly from part (a), where we let $x = p^{1/\alpha}$ and $y = q^{1/\alpha}$. This completes the proof. □

We now present the proof of Theorem 4.28. We first find a separable state σ such that $S_\alpha(\psi\|\sigma) = E_{1/\alpha}(\psi)$, then show that $S_\alpha(\psi\|\sigma') \geq S_\alpha(\psi\|\sigma)$ for every other separable state σ'. The case $\alpha = 1$ has already been proven in Theorem 4.17, so we may suppose that $\alpha \neq 1$.

Proof (*of Theorem 4.28*). We may suppose without loss of generality that the pure state in Schmidt form. We first consider α in the range $\alpha \in (0,1) \cup (1,2]$. The case $\alpha = 0$ will be considered separately. For a given pure state $|\psi\rangle = \sum_i \sqrt{\lambda_i}|ii\rangle$, define the following state

$$\sigma = \frac{1}{\|\vec{\lambda}\|_{1/\alpha}^{1/\alpha}} \sum_i \lambda_i^{1/\alpha} |ii\rangle\langle ii|, \tag{4.33}$$

where $\|\vec{\lambda}\|_{1/\alpha}^{1/\alpha} = \sum_k \lambda_k^{1/\alpha}$. This state in (4.33) clearly separable, since it is a convex combination of separable states $|ii\rangle\langle ii|$. The fact that $S_\alpha(\psi\|\sigma) = E_{1/\alpha}(\psi)$ is straightforward, since

$$\begin{aligned}\langle\psi|\sigma^{1-\alpha}|\psi\rangle &= \sum_i \lambda_i \left(\frac{\lambda_i^{1/\alpha}}{\|\vec{\lambda}\|_{1/\alpha}^{1/\alpha}}\right)^{1-\alpha}\\ &= \|\vec{\lambda}\|_{1/\alpha}\frac{1}{\|\vec{\lambda}\|_{1/\alpha}^{1/\alpha}}\sum_i \lambda_i^{1/\alpha}\\ &= \|\vec{\lambda}\|_{1/\alpha},\end{aligned}$$

and thus

$$\begin{aligned}S_\alpha(\psi\|\sigma) &= \frac{1}{\alpha-1}\log\langle\psi|\sigma^{1-\alpha}|\psi\rangle\\ &= \frac{1}{\alpha-1}\log\|\vec{\lambda}\|_{1/\alpha}\\ &= \frac{1}{1-\frac{1}{\alpha}}\log\|\vec{\lambda}\|_{1/\alpha}^{1/\alpha}\\ &= E_{1/\alpha}(\psi).\end{aligned}$$

To prove $E_{R,\alpha}^{\mathrm{Sep}}(\psi) = S_\alpha(\rho\|\sigma)$, we need to show that σ is in fact a closest separable state to $|\psi\rangle\langle\psi|$ with respect the α-relative entropy for $\alpha \in (0,1)\cup(1,2]$. By Theorem 4.27 and the condition in (4.26), it suffices to show that

$$1-\frac{1}{1-\alpha}\frac{1}{\mathrm{Tr}(\rho^\alpha\sigma^{1-\alpha})}\langle\mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha),\sigma'\rangle \geq 0 \tag{4.34}$$

holds for all separable states σ', where $\rho = |\psi\rangle\langle\psi|$ and σ is the state defined in (4.33). Note that $\rho^\alpha = |\psi\rangle\langle\psi|$ for any $\alpha \geq 0$ and thus

$$\begin{aligned}\mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha) &= \mathfrak{D}_{f_\alpha,\sigma}(|\psi\rangle\langle\psi|)\\ &= \sum_{i,j} f_\alpha^{[1]}\left(\frac{\lambda_i^{1/\alpha}}{\|\vec{\lambda}\|_{1/\alpha}^{1/\alpha}},\frac{\lambda_j^{1/\alpha}}{\|\vec{\lambda}\|_{1/\alpha}^{1/\alpha}}\right)\sqrt{\lambda_i\lambda_j}|ii\rangle\langle jj|\\ &= \|\vec{\lambda}\|_{1/\alpha}\sum_{i,j} f_\alpha^{[1]}\left(\lambda_i^{1/\alpha},\lambda_j^{1/\alpha}\right)\sqrt{\lambda_i\lambda_j}|ii\rangle\langle jj|.\end{aligned}$$

To prove that (4.34) holds for all separable states σ', it will suffice to consider only pure product states. Let $\sigma' = |u\rangle\langle u| \otimes |v\rangle\langle v|$ be a pure product state where $|u\rangle = \sum_i u_i |i\rangle$ and $|v\rangle = \sum_i v_i |i\rangle$ are pure states on the individual subsystems such that $\sum_i |u_i|^2 = \sum_i |v_i|^2 = 1$. Now

$$\begin{aligned}\langle |ii\rangle\langle jj|, \sigma'\rangle &= \operatorname{Tr}(|ii\rangle\langle jj||u\rangle\langle u| \otimes |v\rangle\langle v|) \\ &= u_i v_i \overline{u_j v_j}\end{aligned}$$

for each i and j. Thus

$$\langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma'\rangle = \|\vec{\lambda}\|_{1/\alpha} \sum_{i,j} f_\alpha^{[1]}\left(\lambda_i^{1/\alpha}, \lambda_j^{1/\alpha}\right) \sqrt{\lambda_i \lambda_j} u_i v_i \overline{u_j v_j}.$$

Furthermore, $\operatorname{Tr}(\rho^\alpha \sigma^{1-\alpha}) = \langle \psi | \sigma^{1-\alpha} | \psi\rangle = \|\vec{\lambda}\|_{1/\alpha}$, and thus

$$\begin{aligned}\frac{1}{1-\alpha} \frac{1}{\operatorname{Tr}(\rho^\alpha \sigma^{1-\alpha})} \langle \mathfrak{D}_{f_\alpha,\sigma}(\rho^\alpha), \sigma'\rangle &= \sum_{i,j} \underbrace{\frac{1}{1-\alpha} f_\alpha^{[1]}\left(\lambda_i^{1/\alpha}, \lambda_j^{1/\alpha}\right) \sqrt{\lambda_i \lambda_j}}_{\leq 1 \text{ by Lemma 4.29(b)}} u_i v_i \overline{u_j v_j} \\ &\leq \sum_{i,j} u_i v_i \overline{u_j v_j} \\ &\leq \|u\|^2 \|v\|^2 \\ &= 1.\end{aligned}$$

Hence (4.34) holds for all separable states σ'. This completes the proof in the case $\alpha \in (0, 2]$. The proof of the case $\alpha = 0$ is proved in the following Lemma. □

Lemma 4.30. *For all bipartite pure states $|\psi\rangle$ it holds that $E_{R,0}^{\mathrm{Sep}}(\psi) = E_\infty(\psi)$.*

Note that, for any $\sigma \geq 0$, we have that

$$S_0(\psi \| \sigma) = -\log\langle \psi | \sigma | \psi\rangle.$$

Proof. Without loss of generality we may suppose $|\psi\rangle$ is in Schmidt form with $|\psi\rangle = \sum_i \sqrt{\lambda_i}|ii\rangle$ where the Schmidt coefficients are in decreasing order $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$. Hence $E_\infty(\psi) = -\log \lambda_1$. Let $\sigma = |11\rangle\langle 11|$ which is separable. Then

$$S_0(\psi\|\sigma) = -\log\langle\psi|\sigma|\psi\rangle = -\log \lambda_1$$

as desired. As above, to show that $S_0(\psi\|\sigma') \geq -\log \lambda_1$ for all other separable states σ', it will suffice to consider only pure product states of the form $\sigma' = |u\rangle\langle u| \otimes |v\rangle\langle v|$, where $|u\rangle = \sum_i u_i|i\rangle$ and $|v\rangle = \sum_i v_i|i\rangle$ are arbitrary pure states. For any separable state of this form, we have

$$\begin{aligned}
\langle\psi|\sigma'|\psi\rangle &= |\langle\psi|(|u\rangle \otimes |v\rangle)|^2 \\
&= \Big|\sum_i \sqrt{\lambda_i} u_i v_i\Big|^2 \\
&\leq \Big(\sum_i \sqrt{\lambda_i}|u_i||v_i|\Big)^2 \\
&\leq \lambda_1\Big(\sum_i |u_i||v_i|\Big)^2 \\
&\leq \lambda_1\Big(\sum_i |u_i|^2\Big)\Big(\sum_i |v_i|^2\Big) \\
&= \lambda_1,
\end{aligned}$$

since $\lambda_i \leq \lambda_1$ for all i. Thus $S_0(\psi\|\sigma') \geq -\log \lambda_1$ for any separable pure state σ'. This completes the proof. □

4.6 Cutting plane algorithm for convex optimization problems in quantum information

The optimal values for many of the convex optimization problems discussed in this chapter can be approximated using numerical methods for solving convex optimization problems.

This section outlines a practical algorithm that can be used for the calculation some of the quantities discussed here, such as E_R^{PPT} and $E_{R,\alpha}^{\mathrm{PPT}}$ (the standard and Rényi relative entropies of entanglement with respect to the PPT states), as well as the Rains bound. In particular, for low-dimensional examples, implementation of this algorithm in MATLAB will be shown to provide an estimation for these quantities with an absolute error smaller than 10^{-3}. This algorithm is based on the one first introduced in [ZFG10] used to numerically estimate the relative entropy of entanglement.

The main results of the previous sections of this chapter outline a method of solving *converse problems* of convex optimization problems in quantum information theory. That is, given a matrix σ on the boundary of a convex set $\mathcal{C} \subseteq \mathrm{H}_{d,+}$ of positive semidefinite matrices, we can find all states ρ for which σ minimizes the 'distance' of ρ to $\mathcal{C}$ with respect to some convex divergence measure. While this allows us to find closed-form solutions to the minimization of these quantities for some states, a general closed-form solution for arbitrary states remains elusive. This indicates that the solutions to these problems in general might not be analytical. Hence the need for methods to numerically estimate these quantities. While a similar algorithm has been previously implemented for computing the relative entropy of entanglement [ZFG10], this section presents new algorithms for the calculation of the Rains bound and the Rényi relative entropies of entanglement.

The algorithm presented here is based on a so-called "cutting plane" approach [Tuy08], along with practical techniques for solving convex programs. Namely, we construct successively refined approximations of the epigraph of the objective function using supporting hyperplanes, and use implementable optimization techniques (such as semidefinite programming) to characterize the convex set $\mathcal{C}$. The methods and main idea for the algorithm presented here are due to [ZFG10]. Efficient and practical methods exist [GB08] for solving certain types of convex optimization problems. Here, we implement the algorithm using CVX, a MATLAB-based modeling system for disciplined convex programming that can be used to construct and solve solving convex programs numerically [**?**]. This method cannot be used to directly solve the convex optimization problems that are of interest here, but together with cutting plane method we can implement a sequence of convex optimization problems that converge to the desired result.

This section first introduces the cutting plane method that will be used to implement the main algorithm. This is followed by an explicit description of the algorithm for estimating the optimal value of some of the convex optimization problems that have been introduced earlier in this chapter. Finally, this algorithm is implemented to compute $E_R^{\mathrm{PPT}}(\rho)$, $E_{R,\alpha}^{\mathrm{PPT}}(\rho)$, and $R(\rho)$ (the standard and Rényi relative entropies of entanglement and the Rains bound) for bipartite states of low-dimensional systems. The closed-form expressions for these quantities that can be obtained for some states using the methods of the previous sections are compared to the numerical results here.

4.6.1 Cutting plane method

The *cutting plane method* [Tuy08] is a standard tool for solving convex optimization programs by constructing a successively refined piecewise linear approximation to the epigraph of a convex function in order to estimate the minimum value of some convex function over a convex set. The core of this approach lies in the classical geometric idea that a convex set may be approximated by supporting hyperplanes. The ideas for this section are derived from the algorithm presented in [ZFG10] for computing the relative entropy of entanglement for bipartite quantum states.

Given a convex function f on a convex set $\mathcal{C}$, it is not always possible to implement numerical methods for directly solving the convex optimization problem. In problems where the usual methods for solving such problems fail, the cutting plane method can be implemented to solve such problems if the convex objective function and its gradient can be evaluated efficiently. For simplicity, only differentiable convex objective functions will be considered.

It will first be shown how to construct supporting hyperplanes of the epigraph of a differentiable convex function by making use of the gradient. Let $\mathcal{C}$ be a convex set and let

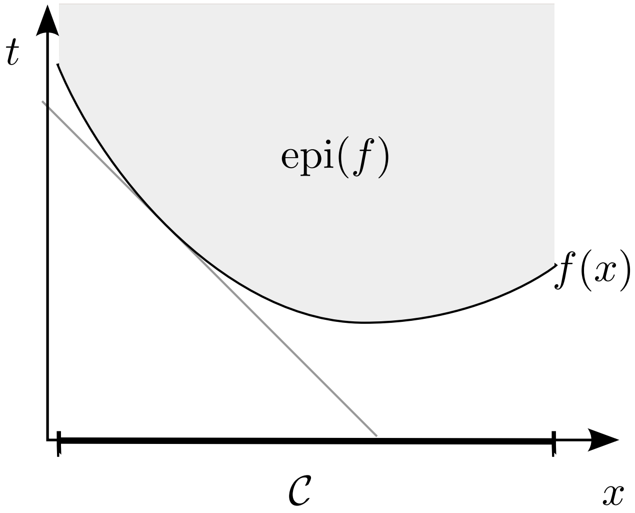


Figure 4.4: Example of an epigraph of a convex function in one dimension. The shaded region is the epigraph. A supporting hyperplane of the epigraph can be given at any point of $\mathcal{C}$ by the gradient of f.

$f : \mathcal{C} \to \mathbb{R}$ be a (differentiable) convex function. Recall that the *epigraph* of f is the set (see Figure 4.4 for an example)

$$\mathrm{epi}(f) = \{(x, t) \,|\, f(x) \leq t\} \subset \mathcal{C} \times \mathbb{R}.$$

Given a point $x \in \mathcal{C}$, we can construct a linear approximation of f at x by examining the supporting hyperplanes of the epigraph of f. Since f is differentiable at x, the function f has a gradient $\nabla f(x)$ at x such that the directional derivatives of f at x are given by

$$f'(x; y - x) = \langle \nabla f(x), y - x \rangle$$

for all $y \in \mathcal{C}$. By convexity of f, it holds that

$$f(y) \geq f(x) + f'(x; y - x) \tag{4.35}$$

for all $y \in \mathcal{C}$. Hence, for a fixed $x \in \mathcal{C}$, it holds that

$$f(y) - \langle \nabla f(x), y \rangle \geq f(x) - \langle \nabla f(x), x \rangle \tag{4.36}$$

for all $y \in \mathcal{C}$.

More generally, given a collection of points $\{x_0, x_1, \ldots, x_n\}$ in $\mathcal{C}$, any $y \in \mathcal{C}$ must satisfy

$$f(y) - \langle \nabla f(x_i), y \rangle \geq f(x_i) - \langle \nabla f(x_i), x_i \rangle \tag{4.37}$$

for all $i = 0, \ldots n$. Hence, any point in the epigraph $(y, t) \in \text{epi}(f)$ must satisfy

$$t - \langle B_i, y \rangle \geq c_i \qquad \text{for all } i, \tag{4.38}$$

where $B_i = \nabla f(x_i)$ and $c_i = f(x_i) - \langle B_i, x_i \rangle$. Hence, if the value of the function f and its gradient can be computed at any points in the interior of $\mathcal{C}$, given any collection of points $\{x_0, \ldots, x_n\} \subset \mathcal{C}$ we can construct a piecewise linear approximation of the epigraph that is tangent to the epigraph at these points.

If minimization of f over $\mathcal{C}$ is impossible or impractical to perform directly, a this type of approximation allows us to implement a cutting plane algorithm to approximate the minimum value of f over $\mathcal{C}$. This assumes that f is differentiable at each point $x \in \mathcal{C}$, and that the value of f and its gradient can be computed. We must also assume that membership in $\mathcal{C}$ can be easily determined. Such an algorithm makes use of the constraints in (4.38) by successively choosing the points $x_0, x_1, \ldots$ such that the constraints in (4.38) allow one to approximate the minimum value of f over $\mathcal{C}$ more and more closely with each iteration.

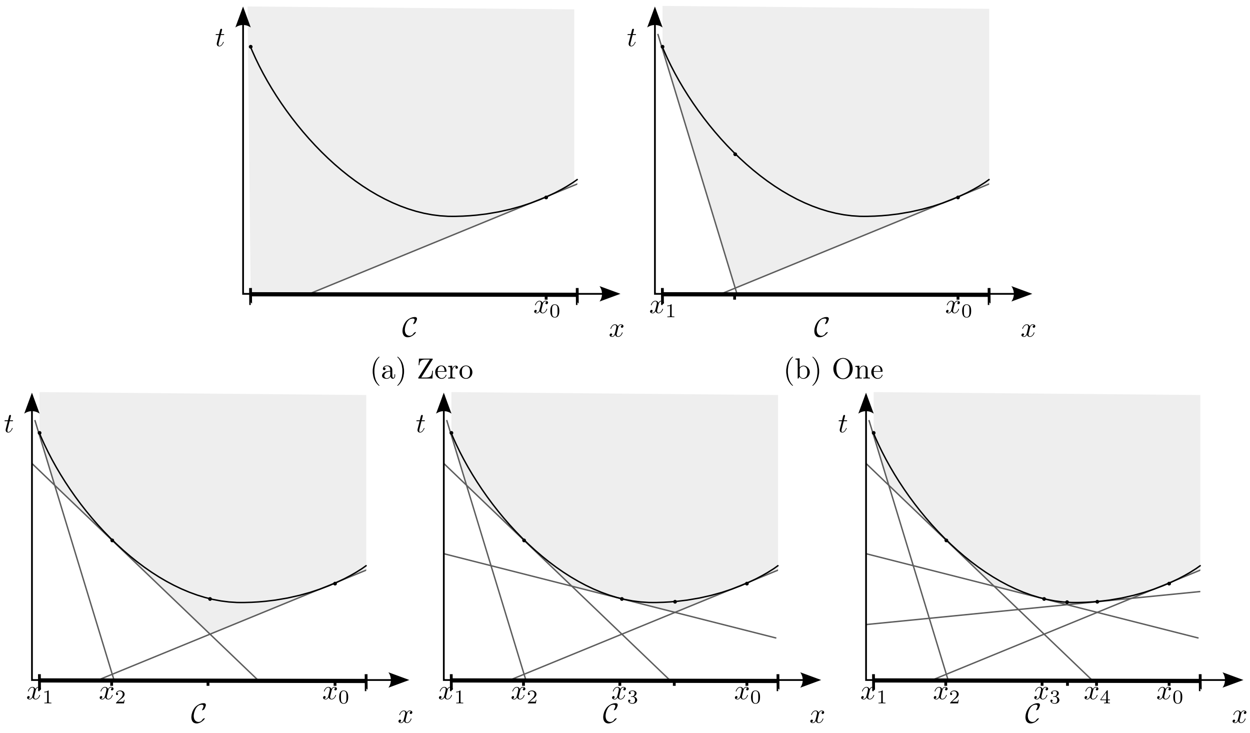
The implementation of the standard cutting-plane method that will be used here (which is based on a method used in [ZFG10] and improved in [GZFG15] for computing the relative entropy of entanglement) performs the following procedure(see Figure 4.5): Create successively improving piecewise-linear approximations to the epigraph of the function restricted to $\mathcal{C}$ and construct a sequence of points $\{x_0, x_1, \dots\}$ that converge to a point $x^\star$ in $\mathcal{C}$ that minimizes f over $\mathcal{C}$. At each iteration, we perform the optimization

$$\begin{aligned} \text{minimize:}\ & t \\ \text{subject to:}\ & y \in \mathcal{C} \\ & t - \langle B_i, y\rangle \geq c_i \,\text{for all}\ i = 0, \dots, n, \end{aligned} \tag{4.39}$$

where the minimization is performed over all (y, t) in the piecewise linear approximation to the epigraph. The optimal point $(y_{\text{opt}}, t_{\text{opt}})$ of this optimization problem is used to construct the next x_n that approximates the optimal point of the primary convex optimization problem $\min_{x\in\mathcal{C}} f(x)$. Efficient implementation of this algorithm requires only that membership in $\mathcal{C}$ be programmable as a constraint in the modeling system.

The main concepts of this type of algorithm are as follows (see Figure 4.5 for a visualization). An initial point $x_0 \in \mathcal{C}$ is chosen, then $B_0 = \nabla f(x_0)$ and $c_0 = f(x_0) - \langle B_0, x_0\rangle$ are computed. The minimization in (4.39) with $n = 0$ is performed and the optimal $(x_{\text{opt}}, t_{\text{opt}})$ are found. The next point x_1 is chosen to be the optimal point of this minimization, and B_1 and c_1 are computed. The optimal value t_{opt} is the new lower bound to $\min_{x\in\mathcal{C}} f(x)$, and $f(x_1)$ is a new upper bound. The minimization in (4.39) with $n = 1$ is now performed, and so on until the resulting upper bound and lower bound differ by less than $\varepsilon > 0$, where ε is the stopping criterion.

The resulting convex optimization relaxations in (4.39) may be efficiently solved with freely available numerical software, assuming that membership in the convex set $\mathcal{C}$ can be programmed as a constraint in the implementation. In the practical implementations of this algorithm presented in this section, the convex set will always either be the set of PPT density matrices $\mathcal{C} = \text{PPTD}(\mathbb{C}^{d_{\mathsf{A}}} : \mathbb{C}^{d_{\mathsf{B}}})$ (in which case the optimization in (4.39) can be performed with semidefinite programming), or the set $\mathcal{C} = \mathcal{R}$ defined in (4.16) that is used for computing the Rains bound. No attempt to prove convergence of the proposed approach will



(a) Zero

(b) One

(c) Two

(d) Three

(e) Four

Figure 4.5: Depiction of the cutting plane algorithm in one dimension. Each iteration constructs a piecewise linear approximation of the epigraph that more and more closely approximates the minimum of a convex function f over a convex set $\mathcal{C}$.

be done here, nor will the theoretical efficiency be discussed. Instead, numerical evidence of its efficiency will be provided by applying it to compute these quantities for low-dimensional states.

4.6.2 Cutting plane method for problems in quantum information

The types of convex optimization problems that we are concerned here with arise from quantum information theory, where the convex set $\mathcal{C}$ is a subset of positive hermitian matrices $\mathrm{H}_{d,+}$ and the functions that we want to minimize are functions $f : \mathrm{H}_{d,+} \to \mathbb{R}$ of matrices of the form

$$f(X) = h(\mathrm{Tr}(Ag(X))), \tag{4.40}$$

where g and h are differentiable functions of real numbers (so that $g : (0, +\infty) \to \mathbb{R}$ can act as a function on positive hermitian matrices) and $A \in \mathrm{H}_{d,+}$ is a fixed positive matrix. As we have seen, for most $X \geq 0$ and $Y \in \mathrm{H}_d$ the directional derivatives of f can be computed as

$$f'(X;Y) = h'(\mathrm{Tr}(Ag(X)))\langle \mathfrak{D}_{g,X}(A), Y\rangle. \tag{4.41}$$

If $X > 0$ is positive definite then f is differentiable at X with gradient given by

$$\nabla f(X) = h'(\mathrm{Tr}(Ag(X)))\mathfrak{D}_{g,X}(A)$$

and the directional derivatives can be written as $f'(X;Y) = \langle \nabla f(X), Y\rangle$, where f is a function of the form in (4.40).

As long as membership $\sigma \in \mathcal{C}$ in the convex set of interest can be written as a constraint that can be handled by the convex optimization modeling software, we can use the cutting plane method described earlier to numerically estimate the optimal value of these convex optimization problems. The algorithm that is presented here closely follows this cutting plane procedure, which was first introduced in [ZFG10]. To improve efficiency of this algorithm, an extra step is added at each iteration to select a the next point X_{n+1} in the approximation of the epigraph. This step involves performing a line search between the approximation point X_n from the previous iteration and the point Y_{opt} from the convex optimization problem at that iteration. Detailed pseudocode for this algorithm is now presented. Explanations of some of the variables used in the code follows below.

Cutting plane algorithm (minimizes convex function f over convex set $\mathcal{C}$)

```
requires $X_{\text{init}} \in \mathcal{C}$
requires $n_{\max}, \varepsilon, \varepsilon_{\text{line}} > 0$
initialize $X_0 := X_{\text{init}}$, $n := 0$, $b_{\text{up}} = f(X_0)$, $b_{\text{lo}} = -\infty$
while ($n \leq n_{\max}$ and $b_{\text{up}} - b_{\text{lo}} > \varepsilon$)
    $B_n := \nabla f(X_n)$
    $c_n := f(X_n) - \langle B_n, X_n \rangle$
    minimize: $t$
    over $(Y, t)$ subject to:
        $t \geq b_{\text{lo}}$
        $Y \in \mathcal{C}$
        $t + \langle B_i, Y \rangle \geq c_i$ for all $i \in \{1, \ldots, n\}$
    $(Y_{\text{opt}}, t_{\text{opt}}) := (Y, t)$
    $s_0 := 0$, $s_1 := 1$
    while $|s_1 - s_0| \|Y_{\text{opt}} - X_n\| > \varepsilon_{\text{line}}$
        $s := \frac{s_1 - s_0}{2}$
        $z := \langle \nabla f((1 - s)Y_{\text{opt}} + sX_n), X_n - Y_{\text{opt}} \rangle$
        if $z \leq 0$
            $s_1 := s$
        else if $z > 0$
            $s_0 := s$
        end if
    end while
    $X_{n+1} := (1 - s)Y_{\text{opt}} + sX_n$
    $b_{\text{up}} := f(X_{n+1})$
    $b_{\text{lo}} := t_{\text{opt}}$
    $n := n + 1$
end while
return $(b_{\text{up}}, b_{\text{lo}}, X_n)$
```

Some of the variables used in this pseudocode are explained below.

- $n_{\max}$ is the maximum number of cutting plane iterations to perform before stopping.
- b_{up} and b_{lo} are the upper and lower bounds for the estimation of the true minimum value of $f(X)$ over $\mathcal{C}$. These values are updated at each iteration.
- ε is the desired accuracy for the estimation. The algorithm halts if $b_{\text{up}} - b_{\text{lo}} < \varepsilon$.
- $X_0 = X_{\text{init}}$ is the initial guess for the optimal minimum. In practical implementations of this algorithm, the initial guess $X_{\text{init}} = \frac{1}{d}\mathbb{1}_d$ is always chosen.
- (Y, t) are the variables over which the internal convex program is performed at each cutting-plane iteration, while $(Y_{\text{opt}}, t_{\text{opt}})$ is the optimal output.
- s, s_0, and s_1 are the variables used in the line search used to find the next point X_{n+1}.
- z is the value of the directional derivative $f'((1-s)Y_{\text{opt}} + sX_n; X_n - Y_{\text{opt}})$ for a given s used in the line search.
- $\varepsilon_{\text{line}}$ denotes the stopping criterion for the line search. The line search at each iteration halt when $|s_1 - s_0|\|Y_{\text{opt}} - X_n\| > \varepsilon_{\text{line}}$.

The code outputs the upper and lower bounds to the desired quantity (b_{up} and b_{lo}) as well as the estimate X_n to the closest point in $\mathcal{C}$ to the actual minimum.

4.6.3 Applications to relative entropies of entanglement and the Rains bound

We can now implement the above algorithm in CVX to compute the relative entropy of entanglement with respect to the PPT states as well as the Rains bound. Full code for these algorithms can be found online at [?]. The numerical results of each algorithm are compared for states that have known values of these quantities.

Relative entropy of entanglement

The relative entropy of entanglement of a bipartite state ρ is computed by minimizing the function

$$f(\sigma) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma)$$

over all PPT states σ. As before, the gradient of this function at a matrix $\sigma > 0$ is given by

$$\nabla f(\sigma) = -L_\sigma(\rho)$$

and membership in the set of PPT density matrices can be given by the constraints

$$Y \geq 0, \quad Y^{T_\mathsf{B}} \geq 0, \quad \text{and} \quad \mathrm{Tr}(Y) = 1.$$

These constraints can be directly coded into CVX as constraints in semidefinite programming mode, and we can implement the cutting plane algorithm discussed earlier. The code for this optimization program in CVX in MATLAB looks like this:

```
minimize t
subject to
    t >= lBound;
    Y >= 0;
    PartialTranspose(Y,1,[dA,dB]) >= 0;
    trace(Y) == 1;
    t + trace(B{0}*Y) >= c{0};
    t + trace(B{1}*Y) >= c{1};
    .
    .
    .
    t + trace(B{n}*Y) >= c{n};
```

where `PartialTranspose` is the routine for taking the partial transpose of a bipartite density matrix and we have the following variables:

- `dA` and `dB` are the dimensions d_A and d_B.

- `B{0}`, $\ldots$, `B{n}` are the gradients $B_i = \nabla f(X_i)$

- `c{0}`, ..., `c{n}` are the constants $c_i = f(X_i) - \langle B_i, X_i \rangle$.

The complete algorithm can be found in [**?**].

To demonstrate the effectiveness this algorithm, we can compare its output for computing $E_R^{\mathrm{PPT}}(\rho)$ with states for which this value is known. Using the methods of Section 4.3, we can generate random states ρ with known value of relative entropy of entanglement. Figure 4.6a shows the actual values of the relative entropy of entanglement and estimated values determined by the numerical estimation algorithm for 100 randomly generated density matrices of two qubits where the matrices are generated according the procedure outlined above. In all cases, the estimated value was found to be within 10^{-5} of the actual value.

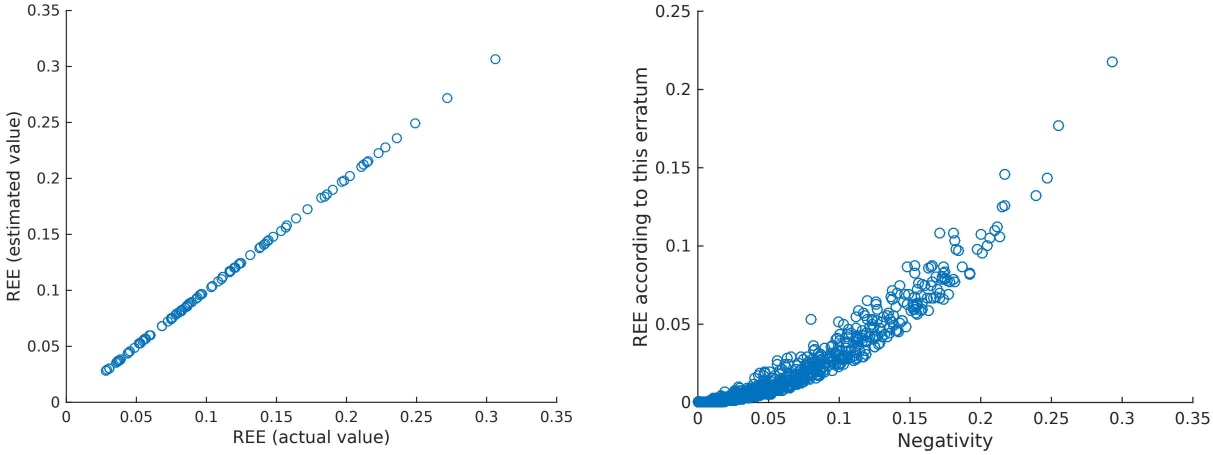
Another way to check the validity of the numerical estimates for the relative entropy of entanglement from this algorithm is to compare the computed value of E_R to the negativity, another well-known entanglement monotone. This is defined as

$$N(\rho) := \frac{\mathrm{Tr}|\rho^{T_{\mathrm{B}}}| - 1}{2}.$$

In Figure 4.6b, the values of the negativity and relative entropy of entanglement are compared for 1000 randomly generated density matrices of two qubits. Here, the random density matrices were generated according to the distribution induced by the Hilbert-Schmidt measure, as shown in [ZPNC11, Mis11]. The plot in Figure 4.6b is compatible with previous works (such as in [MI08]).

The Rains bound

The Rains bound $R(\rho)$ for a state ρ is also computed by minimizing the relative entropy of ρ, but the minimization is performed over all $\sigma \geq 0$ with $\sigma \in \mathcal{R}$ (i.e., $\|\sigma^{T_{\mathrm{B}}}\|_1 \leq 1$). We can use the same algorithm for computing $E_R^{\mathrm{PPT}}(\rho)$ to compute $R(\rho)$ by changing the constraints



(a) Comparison of actual and estimated values of the relative entropy of entanglement (REE) for 100 randomly generated density matrices of two qubits. Random density matrices with known E_R are generated according to the procedure outlined above.

(b) Comparison of negativity and E_R (as estimated using the algorithm presented here) for 1000 random density matrices of two qubits.

Figure 4.6: Comparison of numerical estimation of the relative entropy of entanglement (REE) $E_R(\rho)$ for randomly generated states.

in the CVX optimization of the cutting plane algorithm to the constraint that $Y \in \mathcal{R}$, i.e.,

$$Y \geq 0 \quad \text{and} \quad \|Y^{T_\mathrm{B}}\|_1 \leq 1.$$

This constraint can be encoded as a constraint in the semidefinite programming mode of CVX as follows:

```
Y >= 0;
TraceNorm(PartialTranspose(Y,1,[dA,dB])) <= 1;
```

where `TraceNorm(X) = sum(svd(X))` is the sum of the singular values of a matrix `X`. The algorithm for computing the Rains bound $R(\rho)$ is otherwise identical to the algorithm for computing the relative entropy of entanglement $E_R^{\mathrm{PPT}}(\rho)$ for bipartite states ρ.

To show the effectiveness of this algorithm, we may use this algorithm to compute the Rains bound for randomly generated states of bipartite systems for which at least one system is a qubit. In particular, random states of systems $\mathbb{C}^2 \otimes \mathbb{C}^d$ for $d = 2$, 3, and 4 are generated and are examined here. The result for each randomly generated state was compared to the output from the computation for the relative entropy of entanglement $E_R^{\mathrm{PPT}}(\rho)$ using the algorithm above. Since each state generated this way has at least one subsystem of dimension 2, the result of Theorem 4.26 indicates that these values must coincide. Indeed, we find that the estimated values for $R(\rho)$ and $E_R^{\mathrm{PPT}}(\rho)$ for these states were always between 10^{-5} of each other.

Rényi α-relative entropies of entanglement

This algorithm can also be used to numerically estimate the α-relative entropies of entanglement for $\alpha \in [0,1) \cup (1,2]$. For α in this range, the relative entropy of entanglement of a bipartite state ρ is computed by minimizing the function

$$f(\sigma) = \frac{1}{\alpha - 1} \log \operatorname{Tr}(\rho^\alpha \sigma^{1-\alpha})$$

over all PPT states σ. The gradient of this function at a matrix $\sigma > 0$ is given by

$$\nabla f(\sigma) = \frac{1}{\alpha - 1} \frac{1}{\operatorname{Tr}(\rho^\alpha \sigma^{1-\alpha})} \mathfrak{D}_{f_\alpha, \sigma}(\rho^\alpha),$$

where f_α is the function defined by $f_\alpha(x) = x^{1-\alpha}$, and membership in the set of PPT density matrices can be given by the same semidefinite constraints as in the computation of E_R^{PPT} above. This gradient can be easily computed for any $\sigma > 0$, so the algorithm can be implemented. The efficacy of this algorithm can be tested on pure states. Indeed, in Section 4.5.2, a closed formula for the Rényi relative entropy of entanglement for pure states was given. This analytical result can be compared to the numerical estimate for this value provided by running this algorithm. For randomly generated pure states $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ for $d = 2$, 3, and 4, we find that the result from the numerical estimation to be within 10^{-5} of the known value for variously chosen values of α.

4.7 Other applications

Instead of using $f(x) = -\log(x)$ and only considering ρ to be an arbitrary quantum state as we did for the relative entropy of entanglement and the Rains bound, we may analyze other more general operator convex functions f and positive matrices ρ and analyze when a matrix σ minimizes the function $f_\rho(\sigma) = \mathrm{Tr}(\rho f(\sigma))$. Since the criterion for optimality in Theorem 4.7 only supposes an arbitrary convex function $g : \mathrm{H}_d \to \overline{\mathbb{R}}$, we may use this analysis to produce similar hyperplane criteria for other quantities. In the following section, we examine this criterion as it applies to other quantities of interest in quantum information, such as h_{Sep} and entropy-like quantities, e.g. the relative Rényi entropies and arbitrary quasi-entropies.

4.7.1 On $h_{\mathbf{Sep}}(M)$ and similar quantities

Let $d = d_{\mathsf{A}} d_{\mathsf{B}}$ be the dimension of a bipartite system. Let $M \in \mathrm{H}_{d,+}$ be some positive semidefinite matrix satisfying $0 \leq M \leq \mathbb{1}_d$. Using the identity function $f(x) = x$, and optimizing over the set of separable states, we arrive at the function $f : \mathrm{H}_{d,+} \to \mathbb{R}$ defined by $f_M(\sigma) = \mathrm{Tr}(M\sigma)$. Maximizing this function over all separable density matrices $\sigma \in \mathrm{SepD}(\mathbb{C}^{d_{\mathsf{A}}} : \mathbb{C}^{d_{\mathsf{B}}})$, we define

$$h_{\mathrm{Sep}}(M) = \max_{\sigma \in \mathrm{SepD}} \mathrm{Tr}(M\sigma).$$

This quantity is related to the problem of finding the maximum output ∞-norm of a quantum channel, which is important for proving additivity and multiplicativity for random channels [HM13]. Any quantum channel from a d_1-dimensional quantum system to an d_2-dimensional quantum system can be written as $\mathcal{E}(\rho) = \mathrm{Tr}_{env}(V\rho V^\dagger)$ for some isometry $V : \mathbb{C}^{d_1} \to \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_{env}}$ [Cho75]. This quantum channel can be identified with the operator $M = VV^\dagger$. The maximal output p-norm of the channel $\mathcal{E}$ is defined by

$$\|\mathcal{E}\|_{1\to p} := \max_\rho \|\mathcal{E}(\rho)\|_p$$

where $\|A\|_p = (\mathrm{Tr}|A|^p)^{1/p}$ is the Schatten p-norm and the minimization is taken over all density matrices $\rho \in \mathrm{D}(\mathbb{C}^{d_1})$. It turns out that the maximal output ∞-norm can be given by $\|\mathcal{E}\|_{1\to\infty} = h_{\mathrm{Sep}}(M)$ [HM13]. In addition, h_{Sep} has a natural interpretation in terms of determining the maximum probabilities of success in QMA(2) protocols [Mon13].

Whereas numerically calculating this quantity within an accuracy of $1/\mathrm{poly}(n)$ is known to be an NP-hard problem [Gur03], it is useful to study the converse problem. That is, given a state σ on the boundary of $\mathrm{SepD}(\mathbb{C}^{d_\mathsf{A}} : \mathbb{C}^{d_\mathsf{B}})$, characterizing all matrices $0 \leq \phi \leq \mathbb{1}$ that define supporting functionals of the form

$$\mathrm{Tr}(\phi\sigma) \leq \mathrm{Tr}(\phi\sigma) \quad \text{for all } \sigma \in \mathcal{D}$$

allows us to characterize all of the matrices M for which σ maximizes $\mathrm{Tr}[M\sigma]$.

Since the set of PPT states is much easier to characterize than the set of separable states, maximizing $\mathrm{Tr}[M\sigma]$ over $\mathcal{P}$ instead gives an approachable upper bound to $h_{\mathrm{Sep}}(M)$. That is, we may analyze the quantity

$$h_{\mathrm{PPT}}(M) = \max_{\sigma\in\mathcal{P}} \mathrm{Tr}(M\sigma)$$

where $\mathcal{P} = \mathrm{PPT}$ is the set of states with positive partial transpose. Using this as an upper bound of $\|\mathcal{E}\|_{1\to\infty}$ still gives meaningful useful results [HM13]. Furthermore, the supporting functionals of $\mathcal{P}$ that are maximized by a state σ on the boundary are simple to find (see (4.11) in Section 4.3).

4.7.2 Quasi f-relative entropies

Many important properties of the relative entropy (such as the convexity and monotonicity) are only due to the convexity of the function $f(x) = -\log(x)$, so at first glance there is nothing special about this choice in terms analyzing divergence of two quantum states. It is important to understand more general entropy-like functions for quantum states to glean a better understanding of generalized pseudo-distance measures on the space of quantum states. Many such functions have already been introduced and analyzed [Pet86, OP93, Pet10, Sha10, Sha12, HMPB11].

The most general form of the so-called quasi f-relative entropies first appeared in [Pet86]. A more simplified form that we analyze here, supposing that ρ is a strictly positive matrix, is given by [Sha10, Sha12]

$$S_f(\rho\|\sigma) := \sum_i p_i \left\langle \psi_i \middle| f\left(\frac{\sigma}{p_i}\right) \middle| \psi_i \right\rangle, \tag{4.42}$$

where $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ is the spectral decomposition of ρ. As long as the function $f : (0, \infty) \to \mathbb{R}$ is operator convex, the f-relative entropy S_f satisfies some very important criteria that make it a useful quantity to study. For example, it has been shown [Pet86] that $S_f(\rho\|\sigma)$ is jointly convex and satisfies the data-processing inequality (i.e.

$$S_f(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq S_f(\rho\|\sigma)$$

for any completely positive trace-preserving map $\mathcal{E}$). These quasi-entropies are generally not additive or subadditive, however, so their physical significance is limited.

Although it may be possible to extend the definitions of the quasi-entropies to singular matrices, for the time being we assume that $f : (0, \infty) \to \mathbb{R}$ is is well-defined and that both σ and ρ are strictly positive definite. Considering a convex subset $\mathcal{C} \subseteq \mathrm{H}_{d,+}$, we can define quantities analogous to the relative entropy of of ρ with respect to the convex set $\mathcal{C}$ in the following manner. The *f-relative entropy with respect to $\mathcal{C}$* is defined as

$$S_f(\rho\|\mathcal{C}) := \min_{\sigma\in\mathcal{C}} S_f(\rho\|\sigma).$$

We may use our analysis from Section 4.1 to determine necessary and sufficient conditions for when a matrix $\sigma \in \mathcal{C}$ optimizes the f-relative entropy for a matrix ρ, i.e. when $S_f(\rho\|\mathcal{C}) = S_f(\rho\|\sigma)$. According to Theorem 4.7, this occurs if and only if, for all $\sigma' \in \mathcal{C}$, the directional derivatives satisfy $\frac{d}{dt}S_f(\rho\|\sigma + t(\sigma' - \sigma))\Big|_{t=0^+} \geq 0$.

We can evaluate this derivative by making use of the Fréchet derivative of the function f and find

$$\begin{aligned} \frac{d}{dt}S_f(\rho\|\sigma + t(\sigma - \sigma))\Big|_{t=0^+} &= \sum_i \left\langle \psi_i \left| \mathfrak{D}_{f,\frac{\sigma}{p_i}} (\sigma - \sigma) \right| \psi_i \right\rangle \\ &= \sum_i \left\langle \mathfrak{D}_{f,\frac{\sigma}{p_i}} \left(|\psi_i\rangle\langle\psi_i|\right), \sigma' - \sigma \right\rangle. \end{aligned}$$

Thus, a matrix $\sigma \in \mathcal{C}$ minimizes the f-relative entropy with respect to a matrix ρ if and only if the matrix

$$\phi = -\sum_i \mathfrak{D}_{f,\frac{\sigma}{p_i}} \left(|\psi_i\rangle\langle\psi_i|\right) \tag{4.43}$$

defines a supporting functional of $\mathcal{C}$ at $\sigma \in \mathcal{C}$ of the form

$$\langle \phi, \sigma' \rangle \leq \langle \phi, \sigma \rangle \quad \text{for all } \sigma' \in \mathcal{C}.$$

This characterization of the supporting functionals of a convex set yields interesting information about the original $S_f(\rho\|\sigma)$ quantity, as our analysis of $S(\rho\|\sigma)$ has shown.

Indeed, with the choice $f(x) = -\log(x)$ the standard definition of the relative entropy is recovered. Furthermore, the desired supporting functionals in (4.43) reduce to

$$\begin{aligned} \phi &= -\sum_i \mathfrak{D}_{f,\frac{\sigma}{p_i}} \left(|\psi_i\rangle\langle\psi_i|\right) \\ &= \sum_i L_{\frac{\sigma}{p_i}} \left(|\psi_i\rangle\langle\psi_i|\right) \\ &= L_\sigma \left(\sum_i p_i |\psi_i\rangle\langle\psi_i| \right) \\ &= L_\sigma(\rho), \end{aligned}$$

which is exactly the form of the hyperplanes found in the analysis of the relative entropy in Section 4.2.

Other standard choices of f yield additional well-known entropy-like quantities. For example, the choice $f_\alpha(x) = x^\alpha$ produces a quantity that is related to the relative Rényi entropy, which was studied in the previous section.

4.7.3 Sandwiched relative Rényi entropy

A generalization of the relative Rényi entropy that was recently proposed is another quantity that can be studied. For $\alpha \in (0,1) \cup (1,\infty)$ and matrices $\rho, \sigma \in \mathrm{H}_{d,+}$, the *order α quantum Rényi divergence* (or also called the *"sandwiched" α-relative Rényi entropy*) is defined as [MLDS+13]

$$\tilde{S}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\left(\mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]\right) \tag{4.44}$$

and reduces to the standard α-relative Rényi entropy $\tilde{S}_\alpha(\rho\|\sigma)$ when ρ and σ commute. This quantity has been shown to be jointly convex [FL13] when $\alpha \in [\frac{1}{2}, 1)$ and when the argument ρ is restricted to matrices with unit trace. It also satisfies the data processing inequality for $\alpha \geq \frac{1}{2}$ [Bei13]. In the limit $\alpha \to 1$, it reduces to the standard quantum relative entropy $S(\rho\|\sigma)$. For $\alpha = \frac{1}{2}$, the quantity $S_{1/2}(\rho\|\sigma) = -2\log\|\sqrt{\rho}\sqrt{\sigma}\|$ is closely related to the quantum fidelity [DL14]. It is also positive $S_\alpha(\rho\|\sigma) \geq 0$ for positive matrices ρ and σ and vanishes if and only if $\rho = \sigma$.

As in the previous examples, we can use the conditions in Theorem 4.7 to determine when a matrix σ minimizes the Rényi divergence of ρ over a convex set $\mathcal{C}$. This occurs when

$$\frac{d}{dt}S_\alpha\Big(\rho\Big\|\sigma + t(\sigma - \sigma)\Big)\Bigg|_{t=0^+} \geq 0 \quad \text{for all } \sigma \in \mathcal{C}.$$

Analogous to the many of the other cases studied here, this can be converted to a supporting functional criterion of the form

$$\langle\phi, \sigma'\rangle \leq \langle\phi, \sigma\rangle$$

for all $\sigma \in \mathcal{C}$. Here, ϕ is the matrix

$$\phi = -\mathfrak{D}_{f_\beta,\sigma}\left(\left\{\sigma^{-\beta}, \left(\sigma^\beta \rho \sigma^\beta\right)^\alpha\right\}\right), \tag{4.45}$$

where $\{A, B\} = AB + BA$ is the anti-commutator and $\beta = \frac{1-\alpha}{2\alpha}$. Thus, the Rényi divergence of order $\alpha \in [\frac{1}{2}, 1)$ of a matrix $\rho \in \mathrm{H}_{d,+}$ with respect to $\mathcal{C}$ is minimized by σ on the boundary of $\mathcal{C}$,

$$\min_{\sigma \in \mathcal{C}} S_\alpha(\rho\|\sigma) = S_\alpha(\rho\|\sigma),$$

if any only if the matrix ϕ in (4.45) defines a supporting hyperplane of $\mathcal{C}$ at σ.

4.8 Summary

For a convex function $g : \mathrm{H}_d \to \overline{\mathbb{R}}$ and a convex subset $\mathcal{C} \subseteq \mathrm{H}_d$, we have found a criterion to solve the converse convex optimization problem of determining when a matrix $\sigma \in \mathcal{C}$ minimizes f over $\mathcal{C}$ (i.e. such that $g(\sigma) = \min_{\sigma \in \mathcal{C}} g(\sigma)$). This criterion can usually be given in terms of a supporting functional of $\mathcal{C}$ at σ. Given a convex analytic function $f : (0, \infty) \to \mathbb{R}$, this approach allows us to determine all matrices ρ such that σ minimizes the function $f_\rho = \mathrm{Tr}(\rho f(\sigma))$ over $\mathcal{C}$ by characterizing the supporting hyperplanes of $\mathcal{C}$ at σ. In particular, given a matrix σ, we use this analysis to produce closed formulas for the relative entropy of entanglement and the Rains bound for *all* states for which σ minimizes this quantity, which hold regardless of the dimensionality of the system. Moreover, this allows us to show that the Rains bound and the relative entropy of entanglement coincide for all states for which at least one subsystem is a qubit. We also use this method to find a closed formula for the Rényi α-relative entropy of entanglement for pure states when $\alpha \in [0, 2]$.

Using the Fréchet derivatives, we also make use of the cutting plane method for solving convex optimization problems to create an algorithm for numerically estimating the quantities that arise here. Lastly, we also find supporting functional criteria to determine when a state σ minimizes other various important quantities in quantum information, such as the generalized relative entropies.

Chapter 5

Entanglement of bipartite symmetric states

Entanglement monotones on pure states can be extended to arbitrary mixed states by a convex roof construction [Vid00a, Uhl10, BL13]. Given an entanglement monotone E on pure states, its convex roof on mixed states is defined as

$$\widehat{E}(\rho) = \inf_{\{p_i, |\psi_i\rangle\}} \sum_i p_i E(\psi_i),$$

where the infimum is taken over all pure state decompositions of $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. While there are many known entanglement monotones for bipartite pure states, evaluating the entanglement of arbitrary mixed states is in general not possible. This is due to the fact that the weak membership problem for the set of separable states is known to be NP-hard [Gur03]. Instead, this chapter investigates the computation of convex roof entanglement monotones on certain restricted classes of *symmetric* entangled states, rather than on all states. In particular, convex roofs of the Rényi entropies and the Vidal monotones are computed on Werner and isotropic states [VW01, TV00].

Symmetry plays a very important role in many quantum information tasks. Restricting our attention to highly symmetric states not only simplifies many computations, but yields valuable information about the structure of bipartite entanglement. There is strong evidence that certain symmetric states may provide an example of bound entangled states that have negative partial transposes [DSS+00]. We can restrict our attention only to states that are symmetric in some manner, for example the well known Werner and isotropic states, and exploit that symmetry to compute the convex roof of certain entanglement monotones on those families of states. For example, the entanglement of formation has been computed for Werner states [VW01] and isotropic states [TV00]. Convex roofs of some generalizations of the concurrence [Gou05a] have been computed for isotropic states as well [ETS15, SEG+16].

This chapter expands on existing methods [VW01, TV00] to compute the convex roofs of many more entanglement monotones for these classes of symmetric states and more. In particular, it will be shown that this method can be used to compute the convex roof on Werner states for *all* possible entanglement monotones on pure states. The convex roofs of the Vidal monotones in (2.11) and certain other entanglement monotones for isotropic states will also be computed. These methods can also be extended to compute the convex roof on larger classes of symmetric states as well as on certain non-symmetric states.

While entanglement monotones are important for quantifying entanglement within states, it is also important to characterize which transformations between states can be performed via LOCC *deterministically*. For bipartite pure states, this is completely characterized by majorization of the vectors of Schmidt coefficients [Nie99], or equivalently by the Vidal monotones [Vid00b]. Only a finite number of entanglement measures are needed to determine convertibility of bipartite pure states, but an infinite number of entanglement measures are needed to completely determine convertibility of mixed states [Gou05b]. To characterize convertibility of mixed states, we can instead make use of entanglement *conversion witnesses* [GG15, GF13] (see also Section 2.5.1). An entanglement conversion witness is a function of two bipartite quantum states whose value 'detects' when one state can be converted into another. For example, a *no-go* entanglement conversion witness is a function $W(\rho, \sigma)$ such that $W(\rho, \sigma) < 0$ implies that ρ cannot be converted to σ with a deterministic LOCC operation. Similarly, a *go* entanglement conversion witness is a function $W(\rho, \sigma)$ such that $W(\rho, \sigma) \geq 0$ implies the existence of a deterministic LOCC protocol that converts ρ into σ. A witness is *complete* if it is both a go and a no-go witness.

In [GG08], it was shown that a bipartite pure state $|\psi\rangle$ can be converted into a bipartite mixed state ρ if and only if

$$E_k(\psi) \geq \sum_i p_i E_k(\varphi_i)$$

holds for all k and all decompositions $\rho = \sum_i p_i |\varphi_i\rangle\langle\varphi_i|$. This necessary and sufficient condition for LOCC transformation can be encoded into the following complete conversion witness:

$$W(\psi, \rho) = \max_{\{p_i, \varphi_i\}} \min_k \left(E_k(\psi) - \sum_i p_i E_k(\varphi_i) \right).$$

It holds that $W(\psi, \rho) \geq 0$ if and only if $|\psi\rangle$ can be converted into ρ via LOCC. Although this function cannot be computed for arbitrary mixed states, we can make extensive use of symmetry to compute it in the case when ρ is highly symmetric (e.g. Werner or isotropic). In the final section of this chapter, a class of entanglement conversion witnesses for pure to mixed bipartite state conversion are computed in the case when the target mixed state is symmetric.

The remainder of this chapter is structured as follows. The necessary background for constructing convex roof functions, the definition of the Werner and isotropic states, and other preliminary matter is presented in section 5.1. Convex roofs of certain entanglement monotones are evaluated on Werner and isotropic states in section 5.3. An entanglement conversion witness for pure to mixed state conversion is presented in section 5.4, where it is also shown how to evaluate this conversion witness when the target state is Werner or isotropic.

5.1 Convex roofs and symmetry

In the following, we use the notation $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. Let K be a compact set. Consider a subset $M \subset K$ and let $f \colon M \to \mathbb{R}$. The *convex roof* of f over K is the function $\widehat{f} : K \to \overline{\mathbb{R}}$ defined as

$$\widehat{f}(x) = \inf\left\{ \sum_i p_i f(y_i) \;\middle|\; y_i \in M,\, \sum_i p_i y_i = x \right\}, \tag{5.1}$$

for any $x \in \operatorname{conv}(M)$ in the convex hull of M. The infimum in (5.1) is taken over all convex combinations with $p_i \geq 0$ and $\sum_i p_i = 1$. Note that $\widehat{f}(x) = \infty$ if $x \notin \operatorname{conv}(M)$.

Definition 5.1. Let $\mathcal{G}$ be a compact group with a $\mathcal{G}$-action $g \cdot x$ on K that preserves convex combinations, i.e.,

$$g \cdot (tx + (1-t)y) = tg \cdot x + (1-t)g \cdot y$$

for any $x, y \in K$ and any $t \in [0,1]$. Then the *$\mathcal{G}$-twirling* operator $\mathcal{T}_\mathcal{G} : K \to K$ is defined as

$$\mathcal{T}_\mathcal{G}(x) = \int_\mathcal{G} dg\, g \cdot x, \tag{5.2}$$

for all $x \in K$, where the integral is taken over the Haar measure of the group. If $\mathcal{T}_\mathcal{G}(y) = x$ then we say that *y twirls to x* under $\mathcal{G}$.

In our applications of this theory of convex roofs to computing convex roof entanglement measures, the convex set in question will always be $K = \mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ the density matrices of a $d \times d$ bipartite quantum system, and $M \subseteq K$ is always the extremal set of pure states of the system. The groups will always be subgroups of the group of local unitaries, and the functions f will always be entanglement measures on pure states. However, for full generality we will only assume the necessary properties that must hold, rather than considering the specific case for convex roofs of entanglement measures.

For a group $\mathcal{G}$, the $\mathcal{G}$-invariant elements $x \in K$ are exactly those that satisfy $\mathcal{T}_\mathcal{G}(x) = x$. The subset of $\mathcal{G}$-invariant elements of K will be denoted as $\mathcal{T}_\mathcal{G}(K)$. We will only consider group actions such that $\mathcal{T}_\mathcal{G}(M) = M$. This indeed holds when M is the set of all pure states of a bipartite system and $\mathcal{G}$ is a group of local unitaries.

If f is $\mathcal{G}$-invariant, i.e., $f(g{\cdot}y) = f(y)$ for all $y \in M$ and $g \in \mathcal{G}$, then its convex roof cannot increase under $\mathcal{G}$-twirling. This is indeed the case when f is an entanglement measure on pure states and $\mathcal{G}$ is a subgroup of local unitaries. This statement is proved in the following lemma.

Lemma 5.2. *Let K be a convex set and let $f\colon M \to \overline{\mathbb{R}}$ be a function on a subset $M \subset K$. Let $\mathcal{G}$ be a compact group and K be a compact convex set with a $\mathcal{G}$-action that preserves convex combinations and such that $\mathcal{T}_{\mathcal{G}}(M) \subseteq M$. If f is $\mathcal{G}$-invariant then*

$$\widehat{f}(\mathcal{T}_{\mathcal{G}}(x)) \leq f(x)$$

for all $x \in M$.

Proof. We can consider the ensemble $\{(dg, g \cdot x)\}$ such that $\int_{\mathcal{G}} g \cdot x \, dg = \mathcal{T}_{\mathcal{G}}(x)$. By definition of the convex roof, it follows that

$$\begin{aligned} \widehat{f}(\mathcal{T}_{\mathcal{G}}(x)) &\leq \int_{\mathcal{G}} f(g \cdot x)\, dg \\ &= \int_{\mathcal{G}} f(x)\, dg \\ &= f(x), \end{aligned}$$

as desired. □

Given any function $f\colon M \to \mathbb{R}$ on a subset $M \subset K$, we define the function

$$f_{\mathcal{G}}\colon \mathcal{T}_{\mathcal{G}}(K) \to \overline{\mathbb{R}} \tag{5.3}$$

on the set of $\mathcal{G}$-invariant elements of K as

$$f_{\mathcal{G}}(x) := \inf\{f(y) \,|\, y \in M,\, \mathcal{T}_{\mathcal{G}}(y) = x\} \tag{5.4}$$

for all $x \in \mathcal{T}_\mathcal{G}(K)$. As the following theorem shows, this definition allows us to find a different expression for the convex roof of a function $f\colon M \to \overline{\mathbb{R}}$ evaluated on $\mathcal{G}$-invariant elements of K. This is the primary tool that we will use to compute convex roof entanglement monotones on the Werner and isotropic states.

Here the function $f_\mathcal{G}$ is defined only over the set of $\mathcal{G}$-invariant elements $\mathcal{T}_\mathcal{G}(K)$, and the its convex roof $\widehat{f_\mathcal{G}}$ is defined by minimizing only over convex combinations of elements in the domain $\mathcal{T}_\mathcal{G}(K)$. That is,

$$\widehat{f_\mathcal{G}}(x) := \inf\{\sum_i t_i f(y_i) \,|\, y_i \in \mathcal{T}_\mathcal{G}(K),\, t_i \geq \sum_i t_i = 1,\, \sum_i t_i y_i = x\}.$$

Theorem 5.3 (Sec. IV.A in [VW01])**.** *Let K be a convex set and let $f\colon M \to \overline{\mathbb{R}}$ be a function on a subset $M \subset K$. Let $\mathcal{G}$ be a compact group and K be a compact convex set with a $\mathcal{G}$-action that preserves convex combinations and such that $\mathcal{T}_\mathcal{G}(M) \subseteq M$. Suppose that f is $\mathcal{G}$-invariant. It holds that*

$$\widehat{f}(x) = \widehat{f_\mathcal{G}}(x) \tag{5.5}$$

for all $x \in \mathcal{T}_\mathcal{G}(K)$.

Proof. Let $x \in \mathcal{T}_\mathcal{G}(K)$ be a $\mathcal{G}$ invariant element and suppose that $\{(t_i, y_i)\}$ is an ensemble of elements $y_i \in M$ such that $t_i \geq 0$, $\sum_i t_i = 1$, and $\sum\limits_i t_i y_i = x$. Furthermore suppose that $\{(t_i, y_i)\}$ is the optimal convex combination of elements in M such that

$$\widehat{f}(x) = \sum_i t_i f(y_i).$$

Then $\mathcal{T}_\mathcal{G}(y_i)$ is $\mathcal{G}$-invariant for each y_i and thus $\{(t_i, \mathcal{T}_\mathcal{G}(y_i))\}$ is also an ensemble for x, since

$$\begin{aligned}\sum_i t_i \mathcal{T}_\mathcal{G}(y_i)) &= \mathcal{T}_\mathcal{G}\Big(t_i y_i\Big) \\ &= \mathcal{T}_\mathcal{G}(x) \\ &= x.\end{aligned}$$

It follows that $\widehat{f_{\mathcal{G}}}(x) \le \widehat{f}(x)$. On the other hand, we have

$$\begin{aligned}\widehat{f_{\mathcal{G}}}(x) &= \inf\Big\{\sum_i t_i f_{\mathcal{G}}(x_i)\,\Big|\, x_i \in \mathcal{T}_{\mathcal{G}}(K),\, \sum_i t_i x_i = x\Big\}\\ &= \inf\Big\{\sum_i t_i f(y_i)\,\Big|\, y_i \in M,\, \sum_i t_i \mathcal{T}_{\mathcal{G}}(y_i) = x\Big\}\\ &\ge \inf\Big\{\sum_i t_i f(y_i)\,\Big|\, y_i \in M,\, \sum_i t_i y_i = x\Big\}\\ &= \widehat{f}(x).\end{aligned}$$

The inequality arises due to the fact that $\sum_i t_i y_i = x$ implies that $\sum_i t_i \mathcal{T}_{\mathcal{G}}(y_i) = x$ since x is $\mathcal{G}$-invariant, but not vice versa. Hence $\widehat{f}(x) \le \widehat{f_{\mathcal{G}}}(x)$. This concludes the proof. □

To compute the convex roof $\widehat{f}$ of a function f on the $\mathcal{G}$-invariant elements of K, the result of Theorem 5.3 implies that we can simplify the computation by first minimizing f over all $y \in M$ that twirl to x. Computing the convex roof of the resulting function yields the desired result. This computation is simplified greatly if $f_{\mathcal{G}}$ is already convex as a function of $\mathcal{G}$-invariant elements, in which case $\widehat{f}(x)$ reduces to $f_{\mathcal{G}}(x)$. Note that both $f_{\mathcal{G}}$ and $\widehat{f_{\mathcal{G}}}$ are functions on the convex subset $\mathcal{T}_{\mathcal{G}}(K) \subset K$ of elements that are invariant under the action of $\mathcal{G}$.

One basic feature of convex roof functions is the existence of ‘linear sections’ in the roof function whenever the infimum in (5.1) is found at a non-trivial convex combination. The result of Lemma 5.4 (which is proven in [VW01]) allows us to compute convex roof functions on some elements that are not necessarily symmetric with respect to the group action.

Lemma 5.4. *Suppose $x = \sum_i t_i x_i \in K$ is a convex combination of elements $x_i \in M$ with $t_i > 0$ for each i such that $\widehat{f}(x) = \sum_i t_i f(x_i)$ is minimized. Then $\widehat{f}$ is linear on the convex hull of $\{x_i\}$. That is, it holds that*

$$\widehat{f}\Big(\sum_i s_i x_i\Big) = \sum_i s_i f(x_i) \tag{5.6}$$

for all $s_i \in [0,1]$ *satisfying* $\sum_i s_i = 1$.

Proof. Suppose that $\sum_i t_i x_i$ (with each $t_i > 0$ and $x_i \in M$) is a convex decomposition of x that minimizes $\widehat{f}(x)$. That is,

$$\sum_i t_i x_i = x \qquad \text{and} \qquad \sum_i t_i f(x_i) = \widehat{f}(x).$$

Let $\{s_i\}$ be some other values such that $s_i \geq 0$ and $\sum_i s_i = 1$. Then $y = \sum_i s_i x_i$ is some other convex combination of the points $\{x_i\}$. We will show that

$$\widehat{f}(y) = \sum_i s_i f(x_i). \tag{5.7}$$

That is, this is also an optimal decomposition for y. Let $\{(r_j, y_j)\}$ be any other optimal decomposition for y such that

$$\sum_j r_j y_j = s \qquad \text{and} \qquad \sum_j r_j f(y_j) = \widehat{f}(y).$$

Then there is a small number $\varepsilon > 0$ such that $t_i - \varepsilon r_i \geq 0$ for all i. Hence

$$\begin{aligned} x &= x + \varepsilon y - \varepsilon y \\ &= \sum_i (t_i - \varepsilon s_i) x_i + \varepsilon \sum_j r_j y_j \end{aligned}$$

is a convex combination of elements from M representing x. But we assumed that the decomposition $\{(t_i, x_i)\}$ is optimal for x, so

$$\sum_i t_i f(x_i) \leq \sum_i (t_i - \varepsilon s_i) f(x_i) + \varepsilon \sum_j r_j f(y_j).$$

Since $\varepsilon > 0$, this clearly implies that

$$\sum_i s_i f(x_i) \leq \sum_j r_j f(y_j). \tag{5.8}$$

But $\{(r_j, y_j)\}$ was assumed to be optimal, so $\{(s_i, x_i)\}$ must be optimal for $\widehat{f}(y)$ as well. □

In this chapter, we compute the convex roof of entanglement monotones on pure states for Werner and isotropic states. The minimizing sets will usually be an entire orbit of some pure state under the local-unitary group action. Every pure state in these orbits has the same amount of entanglement under any entanglement monotone, since they differ only by a local unitary. Hence the convex roof of any entanglement monotone will be constant on the convex hull of these orbits. This gives a fairly large class of non-symmetric states for which exact value of many different entanglement monotones can be computed.

5.2 Symmetric states

This section introduces some examples of groups that are used in the study of symmetric bipartite quantum entanglement. Let $d \geq 2$ be an integer and consider bipartite states on $\mathbb{C}^d \otimes \mathbb{C}^d$. The convex set of interest here is the set of normalized density operators $\mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d) = \{\rho \,|\, \rho \geq 0,\, \mathrm{Tr}\,\rho = 1\}$. We are interested in computing the convex roof of entanglement monotones that are defined on the pure states

$$\Big\{ |\psi\rangle\langle\psi| \,\Big|\, |\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d,\ \| |\psi\rangle \| = 1 \Big\} \subset \mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d).$$

It is well known that any entanglement monotone on pure states must be a symmetric concave function of the Schmidt coefficients of the pure states. The primary examples of symmetric states that we will study in this paper are the well known Werner states [Wer89] and isotropic states [HH99].

For the remainder of this paper we assume that $d \geq 2$ and we only consider bipartite states on $\mathbb{C}^d \otimes \mathbb{C}^d$. The symmetric states of interest are those that are symmetric with respect to different subgroups of the group of $d \times d$ local unitaries $\mathrm{LU} \subseteq \mathrm{U}(\mathbb{C}^d \otimes \mathbb{C}^d)$ defined by

$$\mathrm{LU} = \{U \otimes V \,|\, U, V \in \mathrm{U}(d)\}.$$

Given a subgroup $\mathcal{G} \subset \mathrm{LU}$, determining which states are invariant under $\mathcal{G}$ amounts to computing the commutant of $\mathcal{G}$,

$$\mathrm{comm}(\mathcal{G}) = \{A \in \mathrm{L}(\mathbb{C}^d \otimes \mathbb{C}^d) \,|\, [A, g] = 0 \text{ for all } g \in \mathcal{G}\},$$

where $\mathrm{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ denotes the space of linear operators on the tensor product space (i.e. the set of $d^2 \times d^2$ matrices). The commutant $\mathrm{comm}(\mathcal{G})$ is the subspace of operators that commute with every element of $\mathcal{G}$. The twirling operator $\mathcal{T}_{\mathcal{G}}$ can be viewed as the projection operator onto the commutant of $\mathcal{G}$. To determine $\mathrm{comm}(\mathcal{G}) \cap \mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$, i.e. the family of states that are invariant under this action, it is useful to find an orthogonal basis of operators for $\mathrm{comm}(\mathcal{G})$ and express the states as combinations of those basis elements. Finally, note that for any $\mathcal{G} \subseteq \mathrm{LU}$ the twirling operation $\mathcal{T}_{\mathcal{G}}$ is an LOCC operation, since it consists of a convex mixture of local unitary channels.

5.2.1 Werner states

The $d \times d$ *Werner states* [Wer89] are those that commute with all unitaries of the form $U \otimes U$ for some $U \in \mathrm{U}(d)$. That is, Werner states are those which are invariant under the subgroup $\{U \otimes U \,|\, U \in \mathrm{U}(d)\}$. The corresponding twirling operator is

$$\mathcal{T}_{\mathrm{wer}}(\rho) = \int_{\mathrm{U}(d)} dU \, U \otimes U \rho (U \otimes U)^{\dagger},$$

where the integral is taken over the Haar measure of the group $\mathrm{U}(d)$ of $d\times d$ unitary matrices. The commutant of this group is spanned by $\{\mathbb{1}, W\}$, where $\mathbb{1}$ is the identity operator and W is the swap operator defined by $W = \sum_{i,j=1}^{d} |ij\rangle\langle ji|$. The swap operator is both unitary and Hermitian, having eigenvalues 1 and -1 and satisfying $W^2 = \mathbb{1}$. Let W_+ and W_- denote the projectors onto the subspaces spanned by the positive and negative eigenvectors of W, respectively, such that $W = W_+ - W_-$. The Werner states can then be parameterized by

$$\rho_{\mathrm{wer}}(a) = a\tfrac{1}{\binom{d}{2}} W_- + (1-a)\tfrac{1}{\binom{d+1}{2}} W_+ \tag{5.9}$$

for $a \in [0,1]$. These states are entangled for $a \in [\frac{1}{2}, 1]$ and separable otherwise [VW01, Wat16]. Furthermore, it holds that $\mathcal{T}_{\mathrm{wer}}(\sigma) = \rho_{\mathrm{wer}}(\langle \sigma, W_- \rangle)$ for all states σ.

5.2.2 Isotropic states

The $d\times d$ *isotropic states* [HH99] are those invariant under the subgroup $\{U \otimes \overline{U} \,|\, U \in \mathrm{U}(d)\}$. The corresponding twirling operator is

$$\mathcal{T}_{\mathrm{iso}}(\rho) = \int_{\mathrm{U}(d)} dU\, U \otimes \overline{U} \rho (U \otimes \overline{U})^{\dagger}.$$

The commutant of this group is spanned by $\{\mathbb{1}, \Phi_d\}$, where $\Phi_d = \frac{1}{d}\sum_{i,j=1}^{d} |ii\rangle\langle jj|$ is the projection operator onto the maximally entangled pure state $\frac{1}{\sqrt{d}}\sum_{i=1}^{d} |ii\rangle$ of two qudits. This commutant is exactly the partial transpose of the space from the Werner states [ADVW02]. The isotropic states can be parameterized by

$$\rho_{\mathrm{iso}}(b) = b\Phi_d + (1-b)\frac{\mathbb{1} - \Phi_d}{d^2 - 1} \tag{5.10}$$

for $b \in [0,1]$. The isotropic states are entangled for $b \in [\frac{1}{d}, 1]$ and separable otherwise [VW01, Wat16]. Furthermore, it holds that $\mathcal{T}_{\mathrm{iso}}(\sigma) = \rho_{\mathrm{iso}}(\langle \sigma, \Phi_d \rangle)$ for all states σ.

5.2.3 *OO*-invariant states

One way to generalize the isotropic and Werner states to larger classes of symmetric states is to consider the *OO*-invariant states [VW01]. These are the states that are invariant under $\{U \otimes U \,|\, U \in \mathrm{O}(d)\}$, where $\mathrm{O}(d) \subset \mathrm{U}(d)$ is the group of orthogonal operators. Since the orthogonal matrices are the unitaries that satisfy $\overline{U} = U$, this group is a subgroup of both the isotropic and Werner group of local unitaries. The corresponding *OO*-twirling operator is defined as

$$\mathcal{T}_{\mathrm{OO}}(\rho) = \int_{\mathrm{O}(d)} dU \, U \otimes U \rho (U \otimes U)^\dagger .$$

The commutant of this group is spanned by $\{\mathbb{1}, W, \Phi_d\}$ [VW01, section II D]. The *OO*-invariant states can be parameterized as

$$\rho_{\mathrm{O}}(a,b) = a \frac{1}{\binom{d}{2}} W_- + b\Phi_d + (1-a-b)\frac{1}{\binom{d+1}{2}-1}(\mathbb{1} - \Phi_d - W_-) \tag{5.11}$$

for $a, b \in [0,1]$ satisfying $a + b \leq 1$. The *OO*-invariant states that are separable (and also positive under partial transposition) [VW01] are those in the rectangle $(a,b) \in [0, \frac{1}{2}] \times [0, \frac{1}{d}]$. The Werner states are *OO*-invariant states for which $b = \frac{2(1-a)}{d(d+1)}$ and the isotropic states are those for which $b = 1 - \frac{2(d+1)}{d} a$. A schematic of the *OO*-invariant states is shown in Fig 5.1.

The entanglement of formation and asymptotic relative entropy of entanglement of *OO*-invariant states have been computed [VW01, ADVW02]. In section 5.3, it will be shown how to compute almost any convex roof monotone on the *OO*-invariant states.

5.2.4 Phase-permutation-invariant states

Other subgroups of $\mathrm{U}(d)$ lead to further generalizations of the Werner and isotropic states. One possible subgroup that leads to two-parameter families of symmetric states is the fol-

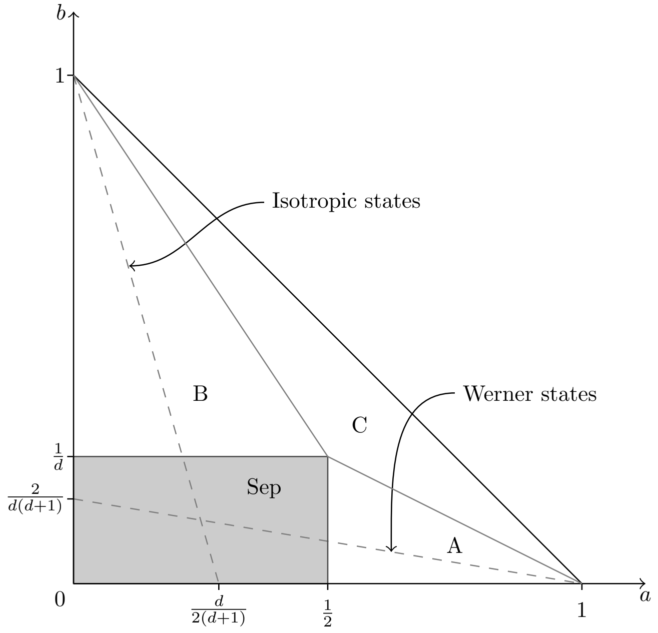


Figure 5.1: Schematic of the OO-invariant states $\rho_{\mathrm{O}}(a,b)$, as defined in (5.11). The shaded region represents the separable (and PPT) states. The one-dimensional subfamilies of Werner and isotropic states are also shown. Convex roof entanglement monotones can be computed for states in regions A and B, as discussed in Section 5.3.4. It remains unknown how to compute convex roofs on states in region C for an arbitrary entanglement monotone.

lowing. Consider the subgroup of 'phase-permutation' unitary matrices defined by

$$G = \{P_\pi U \mid \pi \in \mathcal{S}_d,\, U \in \mathrm{U}(d) \text{ is diagonal}\}, \tag{5.12}$$

where $\mathcal{S}_d$ is the symmetric group and $P_\pi = \sum_{i=1}^d |\pi(i)\rangle\langle i|$ is the permutation matrix for $\pi \in \mathcal{S}_d$. If we denote the group of diagonal unitary matrices by $N \simeq \mathrm{U}(1)^{\times d}$, we see that N is a normal subgroup of G. The group G of phase-permutation unitaries can be viewed as the semi-direct product $G = N \rtimes \mathcal{P}_d$, where $\mathcal{P}_d = \{P_\pi \mid \pi \in \mathcal{S}_d\}$ denotes the group of $d \times d$ permutation matrices. This is also the subgroup of unitaries that have exactly one nonzero entry in each row and column.

Phase-permutation Werner states

Consider the family of Werner-type states which are invariant under $\{U \otimes U \,|\, U \in G\}$, where G is the group of phase-permutation matrices defined in (5.12). Such states are referred to in this thesis as *phase-permutation Werner states*. This class of states was first introduced in [DSS+00] and was used in [GG15]. The corresponding twirling operation is

$$\mathcal{T}^G_{\text{wer}}(\rho) = \int_G dU\, U \otimes U \rho (U \otimes U)^\dagger.$$

The commutant of this group is spanned by $\{\mathbb{1}, W, Q\}$ [DSS+00, section II], where Q is the projection operator

$$Q = \sum_{i=1}^{d} |ii\rangle\langle ii| \tag{5.13}$$

that satisfies $QW_- = W_-Q = 0$ and $QW_+ = W_+Q = Q$. This family of states can be parameterized by

$$\rho^G_{\text{wer}}(a,b) = a\frac{1}{\binom{d}{2}}W_- + b\frac{1}{\binom{d}{2}}(W_+ - Q) + (1-a-b)\frac{1}{d}Q \tag{5.14}$$

for $a, b \in [0,1]$ satisfying $a + b \leq 1$. For all states ρ, it holds that $\mathcal{T}^G_{\text{wer}}(\rho) = \rho^G_{\text{wer}}(a,b)$, where $a = \langle \rho, W_- \rangle$ and $b = \langle \rho, W_+ - Q \rangle$. The Werner states form a subfamily of this class. A schematic of the phase-permutation Werner states is depicted in Fig 5.2.

Phase-permutation isotropic states

Similarly, we can consider the family of isotropic-type states which are invariant under $\{U \otimes \overline{U} \,|\, U \in G\}$. We refer to these as the *phase-permutation isotropic states*. These states have been studied by others [ES13, ETS15, SEG+16], where they have been called the

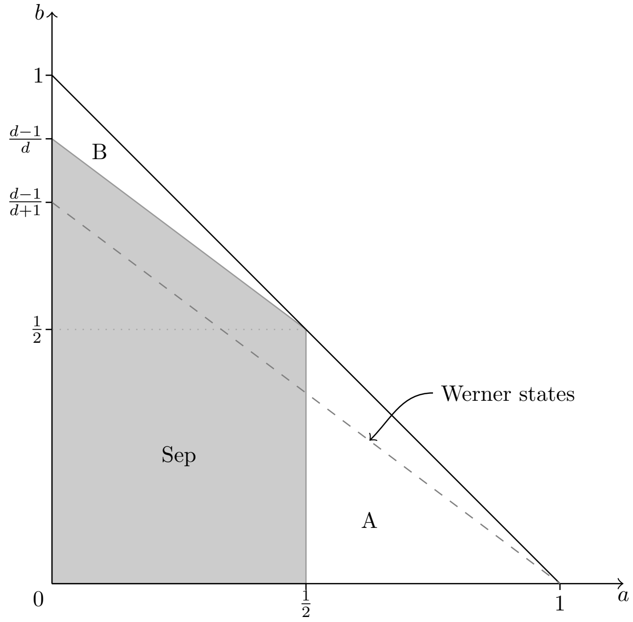


Figure 5.2: Schematic of the phase-permutation Werner states. The separable (PPT) region is shown in gray. The one-dimensional family of states with $b = \frac{d-1}{d+1}(1-a)$ is the well known family of Werner states. As shown in Sec. 5.3.4, the convex roof of any entanglement monotone can be computed for any state in region A. It remains unknown how to compute convex roofs on states in region B for an arbitrary entanglement monotone.

axisymmetric states. The corresponding twirling operation is

$$\mathcal{T}^G_{\text{iso}}(\rho) = \int_G dU\, U \otimes \overline{U} \rho (U \otimes \overline{U})^\dagger.$$

The commutant of this group is spanned by $\{\mathbb{1}, \Phi_d, Q\}$. The elements of this commutant are exactly obtained from the partial transposes of the elements of the commutant of the phase-permutation Werner group presented in (5.14). The family of phase-permutation isotropic states can be parameterized as

$$\rho^G_{\text{iso}}(a,b) = b\Phi_d + a\tfrac{1}{d-1}(Q - \Phi_d) + (1-a-b)\tfrac{1}{d(d-1)}(\mathbb{1} - Q) \tag{5.15}$$

for $a, b \in [0,1]$ satisfying $a + b \leq 1$. For all states ρ, it holds that $\mathcal{T}^G_{\text{iso}}(\rho) = \rho^G_{\text{iso}}(a,b)$, where $b = \langle \rho, \Phi_d \rangle$ and $a = \langle \rho, Q - \Phi_d \rangle$. The isotropic states form a subfamily of this class. A schematic of the phase-permutation isotropic states is depicted in Fig 5.3.

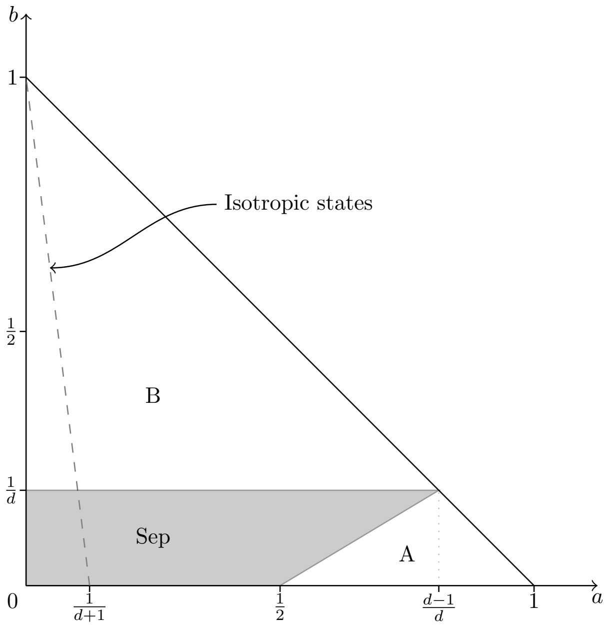


Figure 5.3: Schematic of the phase-permutation isotropic states. The separable (PPT) region is shown in gray. The one-dimensional family of states with $b = 1-(d+1)a$ is the family of isotropic states. As shown in Sec. 5.3.3, the convex roof of any entanglement monotone can be computed for any state in region B. It remains unknown how to compute convex roofs on states in region A for arbitrary entanglement monotones.

5.3 Convex roof entanglement monotones for symmetric states

In this section, we compute the convex roofs of entanglement monotones evaluated on Werner and isotropic states. For Werner states, we compute this for any monotone. For isotropic states, we compute the convex roofs of the Vidal monotones and generalize the computation to certain classes of other monotones.

5.3.1 Werner states

In this subsection we present a general method for computing convex roofs of entanglement monotones evaluated on the Werner states of a $d \times d$ bipartite system. For any $a \in [0,1]$, consider the minimum entanglement of all pure states that twirl to $\rho_{\mathrm{wer}}(a)$ under this action, as in (5.4). Given an arbitrary entanglement monotone E on pure states, we define the function $E_{\mathrm{wer}} \colon [0,1] \to \mathbb{R}$ as

$$E_{\mathrm{wer}}(a) = \min\{E(\psi) \mid \langle\psi|W_{-}|\psi\rangle = a\}. \tag{5.16}$$

If we can evaluate (5.16) for a given entanglement monotone E, then we may make use of Theorem 5.3 to compute the convex roof of E on Werner states by computing $\widehat{E_{\mathrm{wer}}}$. This result is greatly simplified if E_{wer} is already convex as a function of a.

Theorem 5.5. *Let E be an entanglement monotone on pure states. For all $a \in [0,1]$, it holds that*

$$E_{\mathrm{wer}}(a) = E(\psi_a), \tag{5.17}$$

where E_{wer} is the function as defined in (5.16)*, and $|\psi_a\rangle$ are the pure states defined by*

$$|\psi_a\rangle = \left(\sqrt{1-2a}\,|1\rangle + \sqrt{2a}\,|2\rangle\right) \otimes |2\rangle \tag{5.18}$$

whenever $a \in [0, \frac{1}{2}]$, and

$$|\psi_a\rangle = \sqrt{\tfrac{1}{2} + \sqrt{a(1-a)}}\,|12\rangle - \sqrt{\tfrac{1}{2} - \sqrt{a(1-a)}}\,|21\rangle \tag{5.19}$$

whenever $a \in [\frac{1}{2}, 1]$.

Note that the pure states $|\psi_a\rangle$ twirl to the Werner state $\rho_{\mathrm{wer}}(a)$. Indeed, a straightforward calculation shows that $\langle\psi_a|W_{-}|\psi_a\rangle = a$ for all a. In particular, Theorem 5.5 states that the pure states $|\psi_a\rangle$ are in fact optimal in the computation in (5.16) for *every* possible entanglement monotone. This is a generalization of the statement in [VW01, Sec. IV.C], where the convex roof of the entanglement of formation was computed for Werner states. The proof of Theorem 5.5 follows the method used in [VW01].

Proof (*of theorem 5.5*)*.* If $a \in [0, \frac{1}{2}]$ then $E(\psi_a) = 0$ since $|\psi_a\rangle$ is separable, so the conclusion is trivially true. Suppose that $a \in [\frac{1}{2}, 1]$ and let $|\psi\rangle$ be another pure state satisfying

$\langle\psi|W_-|\psi\rangle = a$. Let $\boldsymbol{\lambda}, \boldsymbol{\lambda}^a \in \mathbb{R}^d$ denote the Schmidt vectors of $|\psi\rangle$ and $|\psi_a\rangle$ respectively. We will show that $\boldsymbol{\lambda} \prec \boldsymbol{\lambda}^a$. Since

$$\boldsymbol{\lambda}^a = \left(\tfrac{1}{2} + \sqrt{a(1-a)}, \tfrac{1}{2} - \sqrt{a(1-a)}, 0, \ldots, 0\right)$$

has only two nonzero elements, it suffices to show that $\max(\boldsymbol{\lambda}) \leq \frac{1}{2} + \sqrt{a(1-a)}$.

Without loss of generality we may suppose that $|\psi\rangle$ is of the form

$$|\psi\rangle = U \otimes I \sum_{i=1}^{d} \sqrt{\lambda_i}\, |ii\rangle \tag{5.20}$$

for some unitary operator U. Then

$$\begin{aligned} a = \langle\psi|W_-|\psi\rangle &= \frac{1}{2}\left(1 - \sum_{i,j=1}^{d} \sqrt{\lambda_i\lambda_j}\langle i|U|j\rangle\langle i|U^\dagger|j\rangle\right) \\ &= \frac{1}{2}\left(1 - \sum_{i,j=1}^{d} \sqrt{\lambda_i\lambda_j} U_{ij}\overline{U}_{ji}\right) \\ &= \frac{1}{4}\sum_{i,j=1}^{d} |\sqrt{\lambda_i}U_{ij} - \sqrt{\lambda_j}U_{ji}|^2, \end{aligned}$$

where $U_{ij} = \langle i|U|j\rangle$ are the matrix elements of U. Since U is unitary, it holds that $\sum_j |U_{ij}|^2 = 1$ for each i and thus $\sum_{i,j} \lambda_i |U_{ij}|^2 = 1$. For each $i, j \in \{1, \ldots, d\}$, define the probabilities

$$p_{ij} = \frac{\lambda_i|U_{ij}|^2 + \lambda_j|U_{ji}|^2}{2}$$

such that $p_{ij} \geq 0$ and $\sum_{i,j} p_{ij} = 1$. Note that $p_{ij} = p_{ji}$. For all i and j such that $p_{ij} \neq 0$, define the quantities

$$z_{ij} = \frac{\sqrt{\lambda_i}U_{ij}}{\sqrt{\lambda_i|U_{ij}|^2 + \lambda_j|U_{ji}|^2}} \quad \text{and} \quad a_{ij} = \frac{|z_{ij} - z_{ji}|^2}{2}$$

such that $|z_{ij}|^2 + |z_{ji}|^2 = 1$ and $a_{ij} \in [0,1]$. Define the Schmidt vectors

$$\boldsymbol{\mu}^{(ij)} = |z_{ij}|^2 \boldsymbol{e}_i + |z_{ji}|^2 \boldsymbol{e}_j$$

where $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ are the standard basis vectors of $\mathbb{R}^d$. It follows that

$$\sum_{i,j=1}^{d} p_{ij} a_{ij} = a \qquad \text{and} \qquad \sum_{i,j=1}^{d} p_{ij} \boldsymbol{\mu}^{(ij)} = \boldsymbol{\lambda}.$$

That is, the quantity a and the Schmidt vector $\boldsymbol{\lambda}$ can be written as convex combinations of quantities $a_{ij} \in [0,1]$ and Schmidt vectors $\boldsymbol{\mu}^{(ij)}$ using the same weights p_{ij}. Since $|z_{ij}|^2 + |z_{ji}|^2 = 1$ and $||z_{ij}|^2 - |z_{ji}|^2| \leq |{z_{ij}}^2 - {z_{ji}}^2|$, we see that

$$\begin{aligned} 2\max(\boldsymbol{\mu}^{(ij)}) &= 2\max\{|z_{ij}|^2, |z_{ji}|^2\} \\ &= |z_{ij}|^2 + |z_{ji}|^2 + \Big| |z_{ij}|^2 - |z_{ji}|^2 \Big| \\ &\leq 1 + |{z_{ij}}^2 - {z_{ji}}^2| \\ &= 1 + |z_{ij} - z_{ji}||z_{ij} + z_{ji}|. \end{aligned} \tag{5.21}$$

Furthermore note that $1 - a_{ij} = \frac{1}{2}|z_{ij} + z_{ji}|^2$, and thus

$$|z_{ij} - z_{ji}||z_{ij} + z_{ji}| = 2\sqrt{a_{ij}(1 - a_{ij})}.$$

From (5.21) it follows that $\max(\boldsymbol{\mu}^{(ij)}) \leq \frac{1}{2} + \sqrt{a_{ij}(1 - a_{ij})}$. Since max is a convex function on $\mathbb{R}^d$, it follows that

$$\begin{aligned}\max(\boldsymbol{\lambda}) &= \max\Bigg(\sum_{i,j=1}^{d} p_{ij}\boldsymbol{\mu}^{(ij)}\Bigg)\\ &\leq \sum_{i,j=1}^{d} p_{ij}\max(\boldsymbol{\mu}^{(ij)})\\ &\leq \sum_{i,j=1}^{d} p_{ij}\left(\tfrac{1}{2}+\sqrt{a_{ij}(1-a_{ij})}\right)\\ &\leq \tfrac{1}{2}+\sqrt{a(1-a)}, \qquad (5.22)\end{aligned}$$

where the final inequality in (5.22) follows from concavity of the function $f(t)=\sqrt{t(1-t)}$. This yields the desired result that $\boldsymbol{\lambda} \prec \boldsymbol{\lambda}^a$.

By Nielsen's majorization theorem [Nie99], it follows that $E(\psi) \geq E(\psi_a)$. □

From Theorems 5.3 and 5.5, it follows that $\widehat{E}(\rho_{\text{wer}}(a)) = \widehat{E_{\text{wer}}}(a)$. The family of Werner states is convex, and

$$t\rho_{\text{wer}}(a_1) + (1-t)\rho_{\text{wer}}(a_2) = \rho_{\text{wer}}\Big(ta_1 + (1-t)a_2\Big).$$

Hence the computation of $\widehat{E_{\text{wer}}}(a)$ is greatly simplified if E_{wer} is already convex as a function of a, as it is for the entanglement of formation. Otherwise, there are simple procedures for computing the convex roof of a function of a single variable. Even if the convex roof of E_{wer} as a function of a cannot be computed for a particular entanglement monotone E, the formula in (5.17) still gives an upper bound for $\widehat{E}$ on Werner states, since $\widehat{E}(\rho_{\text{wer}}(a)) = \widehat{E_{\text{wer}}}(a) \leq E_{\text{wer}}(a)$ always holds.

Entanglement of formation

The entanglement of formation [BDSW96] is one well-known convex roof entanglement monotone. This is defined as $E_F(\rho) = \widehat{E}(\rho)$ for mixed states ρ, where E is the entropy of entanglement on pure states $E(\psi) = H(\boldsymbol{\lambda})$, H is the Shannon entropy, and $\boldsymbol{\lambda}$ is the vector of

Schmidt coefficients of $|\psi\rangle$. When $a \in [\frac{1}{2}, 1]$, the entropy of entanglement of $|\psi_a\rangle$ is given by

$$E(\psi_a) = h\Big(\tfrac{1}{2} - \sqrt{a(1-a)}\Big), \tag{5.23}$$

where $h(t) = -t \log t - (1-t)\log(1-t)$ is the binary entropy function. Note that the function in (5.23) is convex as a function of a, so it follows that

$$E_F(\rho_{\text{wer}}(a)) = \begin{cases} 0, & a \in [0, \frac{1}{2}] \\ h\Big(\frac{1}{2} - \sqrt{a(1-a)}\Big), & a \in [\frac{1}{2}, 1]. \end{cases} \tag{5.24}$$

This matches the result found in [VW01].

Vidal monotones

Consider now the Vidal monotones E_k on pure states. Evaluating the convex roof of these monotones on the Werner states can be done easily, because $E_{k,\text{wer}}(a)$ is already convex as a function of a.

Theorem 5.6. *Consider the convex roof of the Vidal monotones E_k on Werner states. The first Vidal monotone reduces to*

$$\widehat{E_1}(\rho_{\text{wer}}(a)) = \begin{cases} 0, & a \in [0, \frac{1}{2}] \\ \frac{1}{2} - \sqrt{a(1-a)}, & a \in [\frac{1}{2}, 1]. \end{cases} \tag{5.25}$$

For $k > 1$, $\widehat{E_k}(\rho_{\text{wer}}(a)) = 0$ for all a.

In particular, the convex roof of the k^{th} Vidal monotone vanishes for all Werner states when $k \neq 1$. Indeed, it holds that $E_k(\psi_a) = 0$ for all a if $k > 1$, since the Schmidt vector of $|\psi_a\rangle$ has at most two nonzero components. For $a \in [\frac{1}{2}, 1]$, note that

$$E_1(\psi_a) = \frac{1}{2} - \sqrt{a(1-a)},$$

which is already convex as a function of a.

Rényi entropies

The result of Theorem 5.5 can also be used to compute the convex roofs of Rényi entropies [HGKM10] of entanglement evaluated on Werner states. For $\alpha > 0$ with $\alpha \neq 1$, the Rényi-α entropy of entanglement is defined as $E_\alpha(\boldsymbol{\lambda}) = \frac{1}{1-\alpha}\log\Bigl(\sum_i \lambda_i^\alpha\Bigr)$ for pure states with Schmidt vector $\boldsymbol{\lambda}$. These are valid entanglement monotones on pure states when $\alpha \in [0,1]$ [Vid00a]. The form of (5.16) for these monotones reduces to $E_{\alpha,\mathrm{wer}}(a) = 0$ when $a \in [0, \frac{1}{2}]$, and

$$E_{\alpha,\mathrm{wer}}(a) = \frac{1}{1-\alpha}\log\left(\left(\tfrac{1}{2} + \sqrt{a(1-a)}\right)^\alpha + \left(\tfrac{1}{2} - \sqrt{a(1-a)}\right)^\alpha\right) \tag{5.26}$$

when $a \in [\frac{1}{2}, 1]$. Numerical evidence suggests that (5.26) is strictly convex whenever $\alpha > 1$, and that (5.26) is strictly concave on the interval $a \in [\frac{1}{2}, 1]$ whenever $\alpha < \frac{1}{2}$. Thus $\widehat{E_\alpha}(\rho_{\mathrm{wer}}(a)) = E_{\alpha,\mathrm{wer}}(a)$ for $\alpha > 1$, and $\widehat{E_\alpha}(\rho_{\mathrm{wer}}(a)) = \max\{0, (2a-1)\log 2\}$ for $\alpha < \frac{1}{2}$.

5.3.2 Isotropic states

In this section, we present a general method for computing convex roofs of entanglement monotones evaluated on the isotropic states of a $d \times d$ bipartite system. In particular, we show explicit formulas for the convex roofs of the Vidal monotones as we did for the Werner states in Section 5.3.1. Using majorization, the result for the Vidal monotones is used to find a simple lower bound for any entanglement monotone on isotropic states. An outline for computing the convex roof of the Rényi entropies on isotropic states is also presented.

The isotropic states $\rho_{\mathrm{iso}}(b)$ defined in (5.10) are the states invariant under the action $U \cdot \rho = U \otimes \overline{U}\rho(U \otimes \overline{U})^\dagger$ from the d-dimensional unitary matrices U. Similar to our analysis of Werner states, for any $b \in [0,1]$ we consider the minimum entanglement of all pure states which twirl to $\rho_{\mathrm{iso}}(b)$ under this action as follows: Given an arbitrary entanglement monotone

E on pure states, define the function $E_{\text{iso}}\colon [0,1] \to \mathbb{R}$ as

$$E_{\text{iso}}(b) = \min\Big\{E(\psi) \,\Big|\, \langle\psi|\Phi_d|\psi\rangle = b\Big\}. \tag{5.27}$$

If we can determine a closed-form expression of (5.27) for a given entanglement monotone E, we can make use of Theorem 5.3 to compute the convex roof of E on isotropic states by computing $\widehat{E_{\text{iso}}}$. This result is greatly simplified if E_{iso} is already convex as a function of b. We use the result of the following lemma to simplify computations.

Lemma 5.7. *Let E be an entanglement monotone on pure states. For all $b \in [\frac{1}{d}, 1]$, it holds that*

$$E_{\text{iso}}(b) = \min\left\{E(\boldsymbol{\lambda}) \,\middle|\, \sum_{i=1}^{d}\sqrt{\lambda_i} = \sqrt{db}\right\}, \tag{5.28}$$

where the infimum is taken over all Schmidt vectors satisfying the condition. Furthermore, $E_{\text{iso}}(b) = 0$ whenever $b \in [0, \frac{1}{d}]$.

A closed-form expression for E_{iso} in the right-hand side of (5.28) can be computed for specific monotones E, which we show in the remainder of this section. In particular, we compute E_{iso} in the cases when E is a Vidal monotone or an entropy-type monotone. Lemma 5.7 is a generalization of the result in [TV00], and the proof is similar. To prove Lemma 5.7, we first provide the following lemma.

Lemma 5.8. *Let $b \in [\frac{1}{d}, 1]$ and let $|\psi\rangle$ be a pure state with Schmidt vector $\boldsymbol{\lambda}$ satisfying $\langle\psi|\Phi_d|\psi\rangle = b$. There exists a pure state $|\psi'\rangle = \sum_i \sqrt{\lambda_i'}|ii\rangle$ such that*

$$\begin{aligned}\langle\psi'|\Phi_d|\psi'\rangle &= \frac{1}{d}\left(\sum_{i=1}^{d}\sqrt{\lambda_i'}\right)^2 = b\\ &= \langle\psi|\Phi_d|\psi\rangle\end{aligned}$$

and $\boldsymbol{\lambda}' \succ \boldsymbol{\lambda}$, where $\boldsymbol{\lambda}'$ is the Schmidt vector for $|\psi'\rangle$.

Proof. We may suppose without loss of generality that $|\psi\rangle$ is of the same form as (5.20).

Thus

$$\begin{aligned} b = \langle\psi|\Phi_d|\psi\rangle &= \frac{1}{d}\left|\sum_{i=1}^{d}\sqrt{\lambda_i}U_{ii}\right|^2 \\ &\leq \frac{1}{d}\Big(\sum_{i=1}^{d}\sqrt{\lambda_i}\Big)^2, \end{aligned} \tag{5.29}$$

where we note that $|U_{ii}| \leq 1$ for all i since U is unitary. If $\sum_i \sqrt{\lambda_i} = \sqrt{db}$ then we may set $\boldsymbol{\lambda}' = \boldsymbol{\lambda}$ and we are done. Suppose instead that the inequality in (5.29) is strict. Define a continuous function $s\colon \mathbb{R}^d \to \mathbb{R}$,

$$s(\boldsymbol{\lambda}) = \frac{1}{d}\Big(\sum_{i=1}^{d}\sqrt{\lambda_i}\Big)^2. \tag{5.30}$$

We may suppose that the entries of $\boldsymbol{\lambda}$ are decreasing. For all $p \in [0,1]$ define the Schmidt vectors

$$\boldsymbol{\lambda}'(p) = (1-p)\boldsymbol{\lambda} + p(1,0,\ldots,0).$$

Note that $s(\boldsymbol{\lambda}'(p))$ is continuous and strictly decreasing as a function of p and that

$$\frac{1}{d} = s(\boldsymbol{\lambda}'(1)) < b < s(\boldsymbol{\lambda}'(0)) = s(\boldsymbol{\lambda}).$$

By continuity of s, there exists a $p \in (0,1)$ such that $s(\boldsymbol{\lambda}'(p)) = b$. Finally we note that $\boldsymbol{\lambda}'(p) \succ \boldsymbol{\lambda}$ for all p, which concludes the proof. □

We now supply the proof of Lemma 5.7. For an entanglement monotone E, recall that E_{iso} is defined as

$$E_{\text{iso}}(b) = \min\Big\{E(\boldsymbol{\psi})\,\Big|\,\langle\psi|\Phi_d|\psi\rangle = b\Big\}.$$

Proof (*of Lemma 5.7*). First consider the case $b \in [\frac{1}{d}, 1]$. For all pure states $|\psi\rangle$ satisfying $\langle\psi|\Phi_d|\psi\rangle = b$, from Lemma 5.8 we can find a pure state $|\psi'\rangle$ with Schmidt coefficients $\boldsymbol{\lambda}'$

satisfying $\langle\psi'|\Phi_d|\psi'\rangle = s(\boldsymbol{\lambda}') = b$ with $\boldsymbol{\lambda} \prec \boldsymbol{\lambda}'$. It follows that $E(\boldsymbol{\lambda}') = E(\psi') \leq E(\psi)$. Hence we may restrict the minimization in (5.27) to states of the form $|\psi\rangle = \sum_i \sqrt{\lambda_i}|ii\rangle$. This implies that the computation of $E_{\text{iso}}(b)$ may be simplified to

$$\begin{aligned} E_{\text{iso}}(b) &= \min\left\{E(\boldsymbol{\psi}) \,\middle|\, |\psi\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i}|ii\rangle \text{ and } s(\boldsymbol{\lambda}) = a\right\} \\ &= \min\left\{E(\boldsymbol{\lambda}) \,\middle|\, \sum_{i=1}^{d} \sqrt{\lambda_i} = \sqrt{db}\right\} \end{aligned}$$

as desired.

Last we consider the case when $b \in [0, \frac{1}{d}]$. Consider the pure state

$$|\psi\rangle = \sqrt{db}\,|11\rangle + \sqrt{1-db}\,|12\rangle.$$

Then $\langle\psi|\Phi_d|\psi\rangle = b$, but $E(\psi) = 0$ since $|\psi\rangle$ is separable. It follows that $E_{\text{iso}}(b) = 0$. This concludes the proof. □

Vidal monotones

Here, we present the results for evaluating the convex roofs of the Vidal monotones (2.11) on isotropic states. The Schmidt vector that minimizes $E_{k,\text{iso}}$ in (5.28) will be of the form.

$$\boldsymbol{\lambda} = \Big(\underbrace{t, \ldots, t}_{k}, \underbrace{\tfrac{1-kt}{d-k}, \ldots, \tfrac{1-kt}{d-k}}_{d-k}\Big), \tag{5.31}$$

with $t \geq \frac{1-kt}{d-k}$. This allows us to compute the convex roofs of the Vidal monotones on isotropic states.

Theorem 5.9. *Consider the convex roof of the Vidal monotones E_k on the isotropic states* *of $\mathbb{C}^d \otimes \mathbb{C}^d$. For $k \in \{1, \ldots, d-1\}$ and $b \in [0,1]$, it holds that*

$$\widehat{E_k}(\rho_{\text{iso}}(b)) = \begin{cases} 0, & b \in [0, \frac{k}{d}] \\ \frac{1}{d}\left(\sqrt{(1-b)k} - \sqrt{b(d-k)}\right)^2, & b \in [\frac{k}{d}, 1]. \end{cases} \tag{5.32}$$

Before proving Theorem 5.9, we must first prove a few lemmas. The following lemma shows that E_k vanishes on the isotropic states with $b \in [0, \frac{k}{d}]$.

Lemma 5.10. *For any integer $1 \leq k \leq d$, it holds that $E_{k,\text{iso}}(b) = 0$ for all $b \in [0, \frac{k}{d}]$.*

Proof. Since E_k is an entanglement monotone on pure states, the result of Lemma 5.7 shows that $E_{k,\text{iso}}(b) = 0$ whenever $b \in [0, \frac{1}{d}]$. So we may suppose that $k \geq 2$ and $b \in [\frac{1}{d}, \frac{k}{d}]$. Consider the function s defined in (5.30) restricted to the subset of Schmidt vectors $\boldsymbol{\lambda}$ that have at most k nonzero entries. The function s achieves the values $\frac{1}{d}$ and $\frac{k}{d}$ on this restriction, since

$$s\Big((1,0,\ldots,0)\Big) = \tfrac{1}{d} \quad \text{and} \quad s\left(\left(\tfrac{1}{k},\ldots,\tfrac{1}{k},0,\ldots,0\right)\right) = \tfrac{k}{d}.$$

The subset of Schmidt vectors in $\mathbb{R}^d$ containing at most k nonzero elements is also connected. By continuity of s, for any intermediate value $b \in [\frac{1}{d}, \frac{k}{d})$ there exists a Schmidt vector $\boldsymbol{\lambda}$ with at most k nonzero entries satisfying $s(\boldsymbol{\lambda}) = b$. Since $E_k(\boldsymbol{\lambda}) = 0$ for all such $\boldsymbol{\lambda}$, it follows that $E_{k,\text{iso}}(b) = 0$ whenever $\frac{1}{d} \leq b \leq \frac{k}{d}$. □

We now compute $E_{k,\text{iso}}(b)$. Convexifying this function as a function of b yields the desired result of Theorem 5.9.

Lemma 5.11. *Let $k \geq 1$ be an integer. It holds that*

$$E_{k,\text{iso}}(b) = \begin{cases} 0, & b \in [0, \frac{k}{d}] \\ \frac{1}{d}\left(\sqrt{(1-b)k} - \sqrt{b(d-k)}\right)^2, & b \in [\frac{k}{d}, 1]. \end{cases} \tag{5.33}$$

Proof. It was shown in Lemma 5.10 that $E_{k,\text{iso}}(b) = 0$ whenever $b \in [0, \frac{k}{d}]$, so it remains to compute $E_{k,\text{iso}}(b)$ when $b \in [\frac{k}{d}, 1]$. Computing $E_{k,\text{iso}}(b)$ may be restated as the following optimization problem:

$$\begin{aligned}\text{maximize: } & \lambda_1 + \cdots + \lambda_k \\ \text{subject to: } & \sum_{i=1}^{d} \lambda_i = 1 \text{ and } \sum_{i=1}^{d} \sqrt{\lambda_i} = \sqrt{db}.\end{aligned}$$

It is not difficult to see (by using Lagrange multipliers) that the optimal $\boldsymbol{\lambda}$ must be of the form

$$\boldsymbol{\lambda} = \Big(\underbrace{t, \ldots, t}_{k}, \underbrace{\tfrac{1-kt}{d-k}, \ldots, \tfrac{1-kt}{d-k}}_{d-k}\Big) \tag{5.34}$$

for some $t \in [\frac{1}{d}, \frac{1}{k}]$. For $\boldsymbol{\lambda}$ of this form, we see that

$$\begin{aligned}\sum_{i=1}^{d} \sqrt{\lambda_i} &= k\sqrt{t} + (d-k)\sqrt{\tfrac{1-kt}{d-k}} \\ &= k\sqrt{(1-t)} + \sqrt{(d-k)(1-kt)}.\end{aligned}$$

For $b \in [\frac{k}{d}, 1]$, the largest positive value of t that satisfies $k\sqrt{(1-t)} + \sqrt{(d-k)(1-kt)} = \sqrt{db}$ is given by

$$t = \frac{1}{k} - \frac{1}{kd}\Big(\sqrt{(1-b)k} - \sqrt{b(d-k)}\Big)^2. \tag{5.35}$$

For $\boldsymbol{\lambda}$ as given in (5.34) with t as in (5.35), it follows that

$$\begin{aligned}E_{k,\mathrm{iso}}(\boldsymbol{\lambda}) &= 1 - (\lambda_1 + \cdots + \lambda_k) \\ &= 1 - kt \\ &= \tfrac{1}{d}\Big(\sqrt{(1-b)k} - \sqrt{b(d-k)}\Big)^2,\end{aligned}$$

as desired. □

The proof of Theorem 5.9 now follows.

Proof (*of Theorem 5.9*). Note that $\widehat{E_k}(\rho_{\mathrm{iso}}(b)) = \widehat{E_{k,\mathrm{iso}}}(b)$ by Theorem 5.3, where $E_{k,\mathrm{iso}}$ is

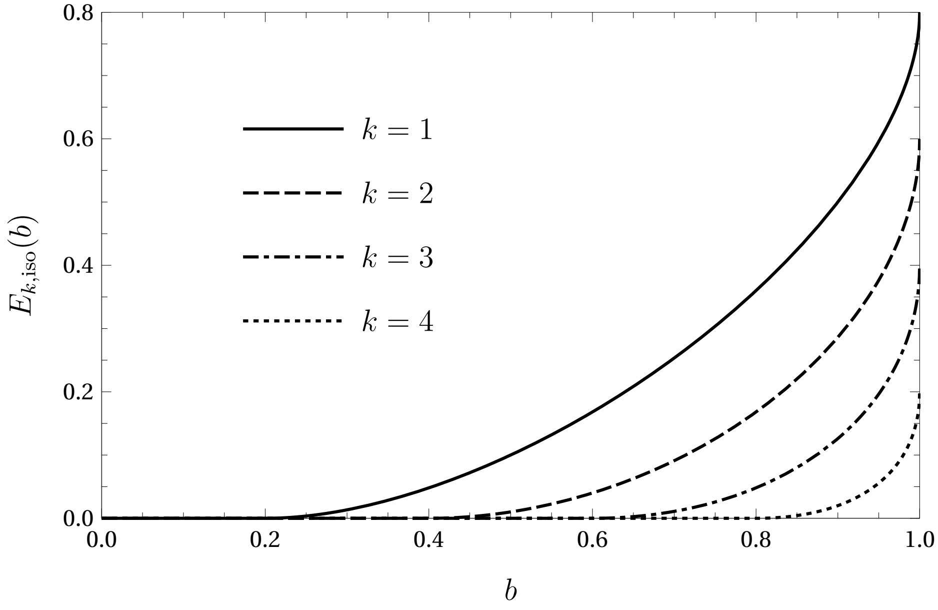


Figure 5.4: The convex roof of the Vidal monotones E_1, E_2, E_3, and E_4 evaluated on isotropic states with dimension $d = 5$.

the function defined in (5.27), and the entanglement monotone used is $E = E_k$. An explicit form of (5.27) for the Vidal monotones is computed in (5.33) of Lemma 5.11. It is clear that $E_{k,\mathrm{iso}}(b)$ in (5.33) is convex as a function of b, which may be confirmed by examining its second derivative. Thus $E_{k,\mathrm{iso}} = \widehat{E_{k,\mathrm{iso}}}$, which concludes the proof. □

The convex roofs of the Vidal monotones can be trivially computed for $k \geq d$, in which case $E_{k,\mathrm{iso}}(b) = 0$ for all $k \geq d$ and any b. A plot of the Vidal monotones (5.32) evaluated on isotropic states $\rho_{\mathrm{iso}}(b)$ with $d = 5$ is shown in Fig 5.4.

It is perhaps interesting to note that the equation

$$y = \left(\sqrt{(1-x)\tfrac{k}{d}} - \sqrt{(1-\tfrac{k}{d})x}\right)^2$$

is part of the unique ellipse which is tangent to the x-axis at the point $(\frac{k}{d}, 0)$, is tangent to the y-axis at the point $(0, \frac{k}{d})$, and goes through the point $(1, 1 - \frac{k}{d})$.

The resulting computations of the Vidal monotones on isotropic states can be used to construct a lower bound for any arbitrary entanglement monotone evaluated on isotropic states. For any $d \geq 2$ and any $b \in [0,1]$, define the following Schmidt vector:

$$\boldsymbol{\lambda}^b = \begin{pmatrix} 1 - E_{1,\mathrm{iso}}(b) \\ E_{1,\mathrm{iso}}(b) - E_{2,\mathrm{iso}}(b) \\ \vdots \\ E_{d-2,\mathrm{iso}}(b) - E_{d-1,\mathrm{iso}}(b) \\ E_{d-1,\mathrm{iso}}(b) \end{pmatrix}. \tag{5.36}$$

For each k, it holds that $E_k(\boldsymbol{\lambda}^b) = E_{k,\mathrm{iso}}(b)$. By construction, it holds that $E_k(\psi) \geq E_k(\boldsymbol{\lambda}^b)$ for any pure state $|\psi\rangle$ that twirls to $\rho_{\mathrm{iso}}(b)$ (i.e. satisfying $\langle\psi|\Phi_d|\psi\rangle = b$). Thus $\boldsymbol{\lambda} \prec \boldsymbol{\lambda}^b$ where $\boldsymbol{\lambda}$ is a Schmidt vector of any pure state that twirls to $\rho_{\mathrm{iso}}(b)$. This implies that we can use $\boldsymbol{\lambda}^b$ to construct a lower bound for any entanglement monotone E evaluated on isotropic states. In particular, it holds that

$$\widehat{E}(\rho_{\mathrm{iso}}(b)) \geq E(\boldsymbol{\lambda}^b) \tag{5.37}$$

for the convex roof of any possible entanglement monotone E evaluated on isotropic states.

Generalized entropy measures

It is also possible to study the convex roof of generalized measures of entropy, as studied in [BS03], rather than entanglement measures. Generalized entropy measures are functions of the form $H_f(\boldsymbol{\lambda}) = \sum_i f(\lambda_i)$ for functions f that satisfy the following conditions:

(i) $f(0) = 0$;

(ii) f is either strictly concave or strictly convex on the interval $[0,1]$; and

(iii) the first derivative f' exists and is continuous on the interval $(0,1)$.

This includes the entropy of entanglement when $f(x) = -x \log x$, as well as quantities that are related to the Rényi entropies when $f(x) = x^\alpha$. In [BS03], it is shown how to compute the minimum and maximum values of one generalized entropy $H_f(\boldsymbol{\lambda})$ for all Schmidt vectors $\boldsymbol{\lambda}$

with some other generalized entropy $H_g(\boldsymbol{\lambda}) = c$ held constant. It turns out that the Schmidt vectors minimizing or maximizing these quantities are either of the form

$$\boldsymbol{\lambda} = \Big(t, \tfrac{1-t}{d-1}, \dots, \tfrac{1-t}{d-1}\Big), \tag{5.38}$$

where $t \geq \frac{1-t}{d-1}$, or

$$\boldsymbol{\lambda} = \Big(t, \dots, t, 1-kt, 0, \dots, 0\Big) \tag{5.39}$$

where $t \geq 1 - kt$, and there are $k = \lfloor \frac{1}{t} \rfloor$ probabilities equal to t. We can then make use of the following theorem from [BS03].

Theorem 5.12 (Theorem 1 in [BS03])**.** *Let $f\colon [0,1] \to \mathbb{R}$ and $g\colon [0,1] \to \mathbb{R}$ both satisfy conditions* (i)-(iii) *for generalized entropy measures listed above. The following statements hold.*

1. *If $f' \circ (g')^{-1}$ is strictly convex (concave), then the maximum (minimum) H_f that can be achieved for fixed H_g is obtained by a probability distribution of the form in* (5.38)*.*
2. *If $f' \circ (g')^{-1}$ is strictly convex (concave), then the minimum (maximum) H_f that can be achieved for fixed H_g is obtained by a probability distribution of the form in* (5.39)*.*

Note that g in Theorem 5.12 is either strictly concave or convex, so it must hold that g' is invertible on the interval $(0, 1)$.

Given a function f that satisfies the conditions above, we can define an entropy measure on pure states by $S_f(\psi) = H_f(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ here is the vector of Schmidt coefficients of $|\psi\rangle$. This can be extended to mixed states via the convex roof construction. Evaluating the convex roof of such an entropy measure on isotropic states $\rho_{\text{iso}}(b)$ amounts to minimizing $H_f(\boldsymbol{\lambda})$, subject to the constraint $\sum_i \sqrt{\lambda_i} = \sqrt{db}$. In particular, we can evaluate functions of the form

$$H_{f,\text{iso}}(b) = \inf\left\{ H_f(\boldsymbol{\lambda}) \,\middle|\, \sum_{i=1}^{d} \sqrt{\lambda_i} = \sqrt{db} \right\} \tag{5.40}$$

for $b \in [\frac{1}{d}, 1]$. The constraint in (5.40) can be rewritten as $\sqrt{db} = H_g(\boldsymbol{\lambda})$, where we choose $g(x) = \sqrt{x}$. If f satisfies the conditions in Theorem 5.12, then we may use Theorem 5.12 to compute the value in (5.40). Note that $(g')^{-1}(x) = \frac{1}{4x^2}$, so it suffices to check if $f'(\frac{1}{4x^2})$ is either strictly concave or convex as a function of x.

Using $\boldsymbol{\lambda}$ of the form in (5.38), solving for t with respect to the constraint $\sum_{i=1}^{d} \sqrt{\lambda_i} = \sqrt{db}$ such that $H_f(\boldsymbol{\lambda})$ is minimized yields

$$t = 1 - \frac{1}{d}\left(\sqrt{1-b} - \sqrt{b(d-1)}\right)^2 . \tag{5.41}$$

Therefore, if $f'(\frac{1}{4x^2})$ is strictly concave, it follows that $H_{f,\text{iso}}(b) = f(t) + (d-1)f(\frac{1-t}{d-1})$, where the value of t is taken from (5.41).

Using $\boldsymbol{\lambda}$ of the form in (5.39), solving for t with respect to the constraint $\sum_{i=1}^{d} \sqrt{\lambda_i} = \sqrt{db}$ such that $H_f(\boldsymbol{\lambda})$ is minimized yields

$$t = \frac{\left(\sqrt{dbk} + \sqrt{k+1-db}\right)^2}{k(k+1)^2}, \tag{5.42}$$

where $k = \lfloor db \rfloor$. It follows that if $f'(\frac{1}{4x^2})$ is strictly convex then $H_{f,\text{iso}}(b) = \lfloor db \rfloor f(t) + f(1 - \lfloor db \rfloor t)$, where the value of t is taken from (5.42). Example values of t in Eqs. (5.41) and (5.42) as functions of b for $d = 5$ are plotted in Fig 5.5.

Generalized concurrences

Using the methods above, it is also possible to compute convex roofs of some of the *generalized concurrence monotones* [Gou05a]. These are defined as follows. For $k = 1, 2, \ldots, d$, let S_k

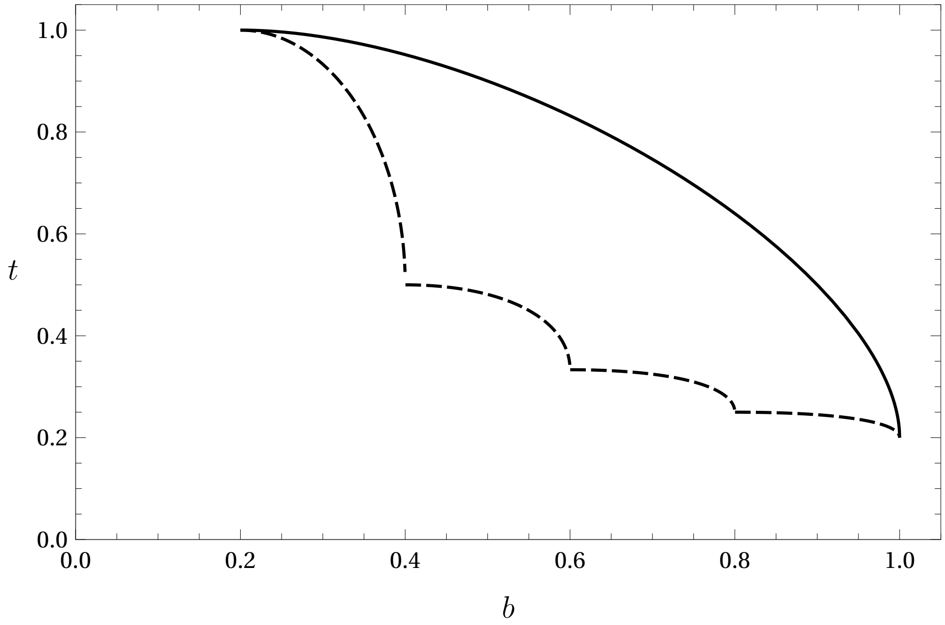


Figure 5.5: Example values of t from Eqs. (5.41) (solid line) and (5.42) (dashed line) as functions of b for $d = 5$.

be the k^{th} elementary symmetric polynomial of d variables. That is,

$$S_1(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \lambda_i,$$

$$S_2(\boldsymbol{\lambda}) = \sum_{i<j} \lambda_i \lambda_j,$$

$$\vdots$$

$$S_d(\boldsymbol{\lambda}) = \prod_{i=1}^{d} \lambda_i.$$

Note that $S_k(\frac{1}{d}, \ldots, \frac{1}{d}) = \frac{1}{d^k}\binom{d}{k}$. The generalized concurrence monotones are defined by

$$C_k(\boldsymbol{\lambda}) = \frac{d}{\binom{d}{k}^{1/k}} S_k(\boldsymbol{\lambda})^{1/k}.$$

These symmetric functions are also concave [Gou05a], and thus are valid entanglement monotones on pure states. Each C_k achieves a maximum value of 1 on the maximally entangled pure state of two qudits. Note that C_d is sometimes called the G-concurrence [Gou05a].

Here, we compute the convex roofs of C_2 and C_d on isotropic states. For $b \in [\frac{1}{d}, 1]$, we minimize C_2 and C_d over all Schmidt vectors that satisfy $\sum_{i=1}^d \lambda_i = 1$ and $\sum_{i=1}^d \sqrt{\lambda_i} = \sqrt{db}$.

We first compute $\widehat{C_2}$ for isotropic states. Note that

$$S_2(\boldsymbol{\lambda}) = \frac{1}{2}\Big(1 - \sum_{i=1}^{d} \lambda_i^2\Big).$$

Hence, minimizing $S_2(\boldsymbol{\lambda})$ is equivalent to maximizing $\sum_{i=1}^d \lambda_i^2$. By Theorem 5.12, the optimal value of this is achieved by the Schmidt vector of the form in (5.38) with the value t from (5.41). Thus,

$$C_{2,\mathrm{iso}}(b) = \frac{\sqrt{d}}{d-1}\sqrt{(1-t)(d(1+t)-2)}, \tag{5.43}$$

with t from (5.41) and $b \in [\frac{1}{d}, 1]$. The function in (5.43) is strictly concave as a function of b, thus its convex roof is the linear function

$$\widehat{C_{2,\mathrm{iso}}}(b) = \begin{cases} 0, & 0 \leq b \leq \frac{1}{d} \\ \frac{db-1}{d-1}, & \frac{1}{d} \leq b \leq 1. \end{cases} \tag{5.44}$$

The convex roof of the 2-concurrence on isotropic states reduces to $\widehat{C_2}(\rho_{\mathrm{iso}}(b)) = \widehat{C_{2,\mathrm{iso}}}(b)$. This agrees with the result from [ETS15].

To compute the convex roof of the G-concurrence $\widehat{C_d}$ for isotropic states, note that

$$\log S_d(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \log \lambda_i.$$

Thus minimizing $S_d(\boldsymbol{\lambda})$ is equivalent to maximizing $\sum_{i=1}^d \log \lambda_i$. By Theorem 5.12, the optimal value will be achieved by the Schmidt vector of the form in (5.39) with the value t from (5.42). Thus $C_{d,\mathrm{iso}}(b) = 0$ for $b \leq 1 - \frac{1}{d}$, and

$$C_{d,\mathrm{iso}}(b) = d\left(t^{d-1} - (d-1)t^d\right)^{1/d} \tag{5.45}$$

for $b > 1 - \frac{1}{d}$, where

$$t = \frac{1}{d(d-1)} \left(\sqrt{(d-1)b} + \sqrt{1-b} \right)^2 .$$

The expression in (5.45) is strictly concave as a function of b; thus its convex roof is the linear function

$$\widehat{C_{d,\mathrm{iso}}}(b) = \begin{cases} 0, & 0 \leq b \leq 1 - \frac{d}{d} \\ db - d + 1, & 1 - \frac{1}{d} \leq b \leq 1. \end{cases} \tag{5.46}$$

Hence, the convex roof of the G-concurrence on isotropic states reduces to $\widehat{C_d}(\rho_{\mathrm{iso}}(b)) = \widehat{C_{d,\mathrm{iso}}}(b)$. This agrees with the result from [SEG$^+$16].

5.3.3 Extension to some non-symmetric states

Here we show how to use the results from the previous sections to compute convex roof entanglement monotones for some states that are not necessarily symmetric.

For a subgroup $\mathcal{G} \subset \mathrm{LU}$ of local unitaries and an entanglement monotone E on pure states, recall that we can define the function

$$E_{\mathcal{G}}(\rho) := \min\{E(\psi) \,|\, \mathcal{T}_{\mathcal{G}}(|\psi\rangle\langle\psi|) = \rho\}$$

on $\mathcal{G}$-invariant states ρ, where the minimization is taken over all pure states that twirl to ρ. A pure state $|\psi\rangle$ is said to minimize the entanglement of ρ (with respect to $\mathcal{G}$ and E) if $\mathcal{T}_{\mathcal{G}}(|\psi\rangle\langle\psi|) = \rho$ and $E(\psi) = E_{\mathcal{G}}(\rho)$. We also consider the orbit of $|\psi\rangle$ under the group $\mathcal{G}$, which we denote as

$$\mathrm{orb}_{\mathcal{G}}(\psi) = \left\{ g|\psi\rangle\langle\psi|g^\dagger \,|\, g \in \mathcal{G} \right\}.$$

Theorem 5.13. *Let $\mathcal{G} \subset \mathrm{LU}$ be a subgroup of local unitaries, let ρ be a $\mathcal{G}$-invariant state, and let $|\psi\rangle$ be a pure state that minimizes the entanglement of ρ with respect to E and $\mathcal{G}$. If*

$\widehat{E_{\mathcal{G}}}(\rho) = E_{\mathcal{G}}(\rho)$, *then*

$$\widehat{E}(\sigma) = E(\psi) \quad \text{for all } \sigma \in \operatorname{conv}\left(\operatorname{orb}_{\mathcal{G}}(\psi)\right), \tag{5.47}$$

where conv *denotes the convex hull.*

Proof. Suppose the conditions of the theorem are satisfied and let $\sigma \in \operatorname{conv}\left(\operatorname{orb}_{\mathcal{G}}(\psi)\right)$. Since $\mathcal{G}$ is a subgroup of local unitaries, it holds that $E(g|\psi\rangle) = E(\psi)$ for all $g \in \mathcal{G}$. It follows that $\widehat{E}(\sigma) \leq E(\psi)$ from the definition of the convex roof. Furthermore, since $\mathcal{T}_{\mathcal{G}}$ is an LOCC channel, it holds that $\widehat{E}(\mathcal{T}_{\mathcal{G}}(\sigma)) \leq \widehat{E}(\sigma)$. Note that $\rho = \mathcal{T}_{\mathcal{G}}(\sigma)$ and $\widehat{E}(\rho) = \widehat{E_{\mathcal{G}}}(\rho)$. The result follows. □

Theorem 5.13 allows us to compute the convex hull on a larger class of non-symmetric states if we can find $\mathcal{G}$-invariant states such that $\widehat{E_{\mathcal{G}}}(\rho) = E_{\mathcal{G}}(\rho)$. On the other hand, if $\widehat{E_{\mathcal{G}}}(\rho) < E_{\mathcal{G}}(\rho)$, we can still compute $\widehat{E}$ on a larger class of non-symmetric states under certain conditions.

5.3.4 Convex roofs on other symmetries

In [VW01], it was shown how to extend the convex roof formula for the entanglement of formation E_F from the Werner and isotropic states to a larger family of OO-invariant states. Here, we show that this can be done for any entanglement monotone. Furthermore, we extend the convex roof formulas to the phase-permutation invariant states as well.

Let $\mathcal{G}$ and $\mathcal{G}'$ be subgroups of the local unitaries with $\mathcal{G}' \subset \mathcal{G}$. The commutants of $\mathcal{G}$ and $\mathcal{G}'$ satisfy $\operatorname{comm}(\mathcal{G}) \subset \operatorname{comm}(\mathcal{G}')$, so the family of $\mathcal{G}$-invariant states forms a subset of the $\mathcal{G}'$-invariant states. If it is known how to compute the convex roofs of entanglement monotones on $\mathcal{G}$-invariant states, then we can apply the result of Theorem 5.13 to compute convex roofs on some $\mathcal{G}'$-invariant states that are also in the convex hull of the orbit of some minimizing pure state. That is, if $|\psi\rangle$ is a minimizing pure state for some $\mathcal{G}$-invariant state

ρ, the convex roofs of entanglement monotones can be evaluated on all states that are in the intersection

$$\mathcal{T}_{\mathcal{G}'}(\mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)) \cap \operatorname{conv}\left(\operatorname{orb}_{\mathcal{G}}(\psi)\right)$$

The minimizing pure states for Werner states are always the states $|\psi_a\rangle$ as defined in (5.19). We first show which of the phase-permutation Werner states $\rho^G_{\mathrm{wer}}(a,b)$ and OO-invariant states $\rho_{\mathrm{O}}(a,b)$ are in the orbits of these minimizing pure states. These are exactly the states depicted in region A of both Fig. 5.1 and Fig. 5.2. This allows us to extend the formulas for convex roof entanglement monotones from the Werner states to this larger family of states.

Lemma 5.14. *Let $a \in [\frac{1}{2}, 1]$. Then*

1. $\rho^G_{\mathrm{wer}}(a,b) \in \operatorname{conv}(\operatorname{orb}_{\mathrm{wer}}(\psi_a))$ *for all* $b \in [0, 1-a]$*; and*

2. $\rho_{\mathrm{O}}(a,b) \in \operatorname{conv}(\operatorname{orb}_{\mathrm{wer}}(\psi_a))$ *for all* $b \in [0, \frac{2}{d}(1-a)]$.

That is, all states in region A *of Fig 5.1 and region* A *of Fig 5.2 are in the convex hulls of the orbits of the corresponding minimizing pure states for* $\rho_{\mathrm{wer}}(a)$.

Proof (*of Lemma 5.14 part 1*). By convexity, it suffices to check only the states on the boundary. That is, we check $\rho^G_{\mathrm{wer}}(a,b)$ with $b = 0$ and $b = 1-a$. In both cases, we find a pure state $|\psi\rangle \in \operatorname{orb}_{\mathrm{wer}}(\psi_a)$ such that $\mathcal{T}^G_{\mathrm{wer}}(\psi) = \rho^G_{\mathrm{wer}}(a,b)$.

Note that $\langle\psi_a|Q|\psi_a\rangle = 0$. Hence $\mathcal{T}^G_{\mathrm{wer}}(\psi_a) = \rho^G_{\mathrm{wer}}(a,0)$ and it follows that

$$\rho^G_{\mathrm{wer}}(a,0) \in \operatorname{conv}(\operatorname{orb}_{\mathrm{wer}}(\psi_a)).$$

For $\rho^G_{\mathrm{wer}}(a, 1-a)$, consider the unitary block matrix

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \\ \frac{i}{\sqrt{2}} & \frac{-i}{\sqrt{2}} & \\ & & \mathbb{1} \end{pmatrix} \tag{5.48}$$

that acts non-trivially only on the span of $\{|1\rangle, |2\rangle\}$. Then

$$U \otimes U|\psi_a\rangle = \sqrt{\frac{1-a}{2}}(|11\rangle + |22\rangle) - i\sqrt{\frac{a}{2}}(|12\rangle - |21\rangle).$$

It holds that $\langle\psi_a|U^\dagger \otimes U^\dagger Q U \otimes U|\psi_a\rangle = 1-a$ and thus $\mathcal{T}^G_{\text{wer}}(U \otimes U\psi_a) = \rho^G_{\text{wer}}(a, 1-a)$, which completes the proof. □

Proof (*of Lemma 5.14 part 2*). By convexity, it suffices to check only the states on the boundary, i.e. $\rho^G_{\text{wer}}(a,b)$ with $b = 0$ and $b = \frac{2(1-a)}{d}$. In both cases, we will find a pure state $|\psi\rangle \in \text{orb}_{\text{wer}}(\psi_a)$ such that $\mathcal{T}_{\text{O}}(\psi) = \rho^G_{\text{O}}(a,b)$. Note that $\langle\psi_a|\Phi_d|\psi_a\rangle = 0$. Thus $|\psi_a\rangle$ twirls to $\mathcal{T}^G_{\text{O}}(\psi_a) = \rho_{\text{O}}(a,0)$. With the same U as in (5.48), it holds that

$$\langle\psi_a|U^\dagger \otimes U^\dagger \Phi_d U \otimes U|\psi_a\rangle = \frac{2(1-a)}{d}.$$

This implies that $\mathcal{T}_{\text{O}}(U \otimes U\psi_a) = \rho^G_{\text{wer}}(a, \frac{2(1-a)}{d})$ which completes the proof. □

A similar statement can be made for isotropic states. Here, however, the form of the Schmidt coefficients of the minimizing pure state $|\phi_b\rangle = \sum_{i=1}^d \sqrt{\lambda_i}|ii\rangle$ for the isotropic state $\rho_{\text{iso}}(b)$ will depend on which entanglement monotone E is being considered. As above, the convex roof of E can be evaluated on any state in the convex hull of the orbit of $|\phi_b\rangle$. In the following lemma, we show which phase-permutation isotropic states and which OO-invariant states are in the convex hulls of these orbits. For any E, all phase-permutation isotropic states $\rho_{\text{iso}}(a,b)$ in region B of Fig. 5.3 are in the convex hull of the orbit of the minimizing pure state $|\phi_b\rangle$. In most cases, all OO-invariant states $\rho_{\text{O}}(a,b)$ in region B of Fig. 5.1 are also in the convex hull of the orbit of $|\phi_b\rangle$.

Lemma 5.15. *Let E be an entanglement monotone on pure states and let $b \in [\frac{1}{d}, 1]$. Let $|\phi_b\rangle = \sum_{i=1}^d \sqrt{\lambda_i}|ii\rangle$ be the pure state that minimizes E for $\rho_{\text{iso}}(b)$. Then*

1. $\rho^G_{\text{iso}}(a,b) \in \text{conv}(\text{orb}_{\text{iso}}(\phi_b))$ *for all* $a \in [0, 1-b]$, *and*

2. If $\boldsymbol{\lambda}$ *is of the form in either* (5.31) *or* (5.39)*, then* $\rho_{\mathrm{O}}(a,b) \in \mathrm{conv}(\mathrm{orb}_{\mathrm{iso}}(\phi_b))$ *for all* $a \in [0, \frac{d(1-b)}{2(d-1)}]$.

That is, all states in region B *of Fig 5.1 and region* B *of Fig 5.3 are in the convex hulls of the orbits of the corresponding minimizing pure states for* $\rho_{\mathrm{iso}}(b)$*.*

Recall that, for any entanglement monotone E and any $b \in [\frac{1}{d}, 1]$, the pure state that minimizes (5.28) will be of the form

$$|\phi_b\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i}|ii\rangle \tag{5.49}$$

where the Schmidt coefficients satisfy $\sum_{i=1}^{d} \sqrt{\lambda} = \sqrt{db}$.

Proof (*of Lemma 5.15 part 1*). As above, it suffices to check only the states on the boundary. That is, we check $\rho_{\mathrm{iso}}^{G}(a,b)$ with $a = 0$ and $a = 1-b$. In both cases, we will find a pure state $|\psi\rangle \in \mathrm{orb}_{\mathrm{iso}}(\phi_b)$ such that $\mathcal{T}_{\mathrm{iso}}^{G}(\psi) = \rho_{\mathrm{iso}}^{G}(a,b)$. Note that $\langle\phi_b|Q|\phi_b\rangle = 1$ and thus

$$\langle\phi_b|(Q - \Phi_d)|\phi_b\rangle = 1 - b.$$

Hence $\mathcal{T}_{\mathrm{iso}}^{G}(\phi_b) = \rho_{\mathrm{iso}}^{G}(1-b,b)$ and thus $\rho_{\mathrm{iso}}^{G}(1-b,b) \in \mathrm{conv}(\mathrm{orb}_{\mathrm{iso}}(\phi_b))$. For $\rho_{\mathrm{iso}}^{G}(0,b)$, we use the discrete Fourier transform unitary matrix

$$U = \frac{1}{\sqrt{d}} \sum_{j,k=1}^{d} \omega^{jk}|j\rangle\langle k|,$$

where $\omega = e^{\frac{2i\pi}{d}}$ is the d^{th} root of unity. It holds that

$$\begin{aligned}
\langle\phi_b|U^{\dagger} \otimes \overline{U}^{\dagger} Q U \otimes \overline{U}|\phi_b\rangle &= \frac{1}{d^2} \sum_{k=1}^{d} \Big(\sum_{j=1}^{d} \sqrt{\lambda_j} |\omega^{jk}|^2 \Big)^2 \\
&= \frac{1}{d} \Big(\sum_{j=1}^{d} \sqrt{\lambda_j} \Big)^2 \\
&= b.
\end{aligned}$$

Thus $\langle\phi_b|U^\dagger\otimes\overline{U}^\dagger(Q-\Phi_d)U\otimes\overline{U}|\phi_b\rangle = 0$. This implies that $\mathcal{T}^G_{\text{iso}}(U\otimes\overline{U}|\phi_b\rangle) = \rho^G_{\text{iso}}(0,b)$, which completes the proof. □

Proof (*of Lemma 5.15 part 2*). It suffices to check only the states on the boundary. That is, we check $\rho^G_{\text{O}}(a,b)$ with $a=0$ and $a=\frac{d(1-b)}{2(d-1)}$. Note that $\langle\phi_b|W_-|\phi_b\rangle = 0$ and thus $\mathcal{T}_{\text{O}}(\phi_b) = \rho_{\text{O}}(0,b)$. Hence $\rho_{\text{O}}(0,b)\in\text{conv}(\text{orb}_{\text{iso}}(\phi_b))$. For $\rho_{\text{O}}(\frac{d(1-b)}{2(d-1)},b)$, it suffices to find a unitary U such that

$$\langle\phi_b|(U\otimes\overline{U})^\dagger W_-(U\otimes\overline{U})|\phi_b\rangle \geq \frac{d(1-b)}{2(d-1)}.$$

We split the proof into two parts. First suppose that $\boldsymbol{\lambda}$ is of the form

$$\boldsymbol{\lambda} = \left(t,\ldots,t,\tfrac{1-kt}{d-k},\ldots,\tfrac{1-kt}{d-k}\right)$$

with $|\phi_b\rangle = \sum_{i=1}^d\sqrt{\lambda_i}|ii\rangle$ and

$$\left(\sum_{i=1}^d\sqrt{\lambda_i}\right)^2 = \left(k\sqrt{t}+\sqrt{(d-k)(1-kt)}\right)^2 = db.$$

For distinct indices $j,k\in\{1,2,\ldots,d\}$ with $j<k$, define the unitary matrices

$$U_{j,k} = \sum_{l\neq j,k}|l\rangle\langle l| + \frac{1}{\sqrt{2}}\Big(|j\rangle\langle j| + |j\rangle\langle k| + i|k\rangle\langle j| - i|k\rangle\langle k|\Big)$$

that act non-trivially only on the subspace spanned by $\{|j\rangle,|k\rangle\}$ and trivially elsewhere. Note that U in (5.48) is $U_{1,2}$ in this notation. Furthermore note that

$$\langle\phi_b|(U_{j,k}\otimes\overline{U_{j,k}})^\dagger W_-(U_{j,k}\otimes\overline{U_{j,k}})|\phi_b\rangle = \frac{\left(\sqrt{\lambda_j}-\sqrt{\lambda_k}\right)^2}{2}.$$

Let $U = (U_{1,d})(U_{2,d-1}) \cdots (U_{m,d+1-m})$, where $m = \min\{k, d-k\}$. Then

$$\begin{aligned}
\langle \phi_b | (U \otimes \overline{U})^\dagger W_- (U \otimes \overline{U}) | \phi_b \rangle &= \frac{(\sqrt{\lambda_1} - \sqrt{\lambda_d})^2}{2} + \cdots + \frac{(\sqrt{\lambda_m} - \sqrt{\lambda_{d-m+1}})^2}{2} \\
&= \frac{m}{2(d-k)} \left(\sqrt{(d-k)t} - \sqrt{1-kt} \right)^2 \\
&= \frac{m}{2(d-k)} \left(\frac{d(1-b)}{k} \right) \\
&= \frac{d(1-b)}{2} \frac{\min\{k, d-k\}}{k(d-k)} \\
&\geq \frac{d(1-b)}{2} \frac{1}{d-1}
\end{aligned}$$

with equality if and only if $k = d-1$ or $k = 1$ (or $b = 1$). The result follows.

The proof of the other case is analogous. In this case, suppose that $\boldsymbol{\lambda}$ is of the form

$$\boldsymbol{\lambda} = \Big(t, \ldots, t, 1-kt, 0, \ldots, 0 \Big)$$

with $(\sum_{i=1}^d \sqrt{\lambda_i})^2 = (k\sqrt{t} + \sqrt{1-kt})^2 = db$. Using the unitary

$$U = (U_{1,d})(U_{2,d-1}) \cdots (U_{\lfloor \frac{d}{2} \rfloor, d - \lfloor \frac{d}{2} \rfloor + 1}),$$

it is not difficult to show that

$$\langle \phi_b | (U \otimes \overline{U})^\dagger W_- (U \otimes \overline{U}) | \phi_b \rangle \geq \frac{d(1-b)}{2(d-1)}$$

with equality if and only if $k = d-1$ (or $b = 1$). □

For every entanglement monotone considered in this chapter, the Schmidt coefficients of the minimizing pure states have this desired form. This allows us to extend the convex roofs of these entanglement monotones from the isotropic states to this larger family of states.

If $E_{\text{wer}}(a)$ and $E_{\text{iso}}(b)$ are already convex as functions of a and b, then Lemmas 5.14 and 5.15, together with Theorem 5.13, allow us to extend these convex roof formulas to any state

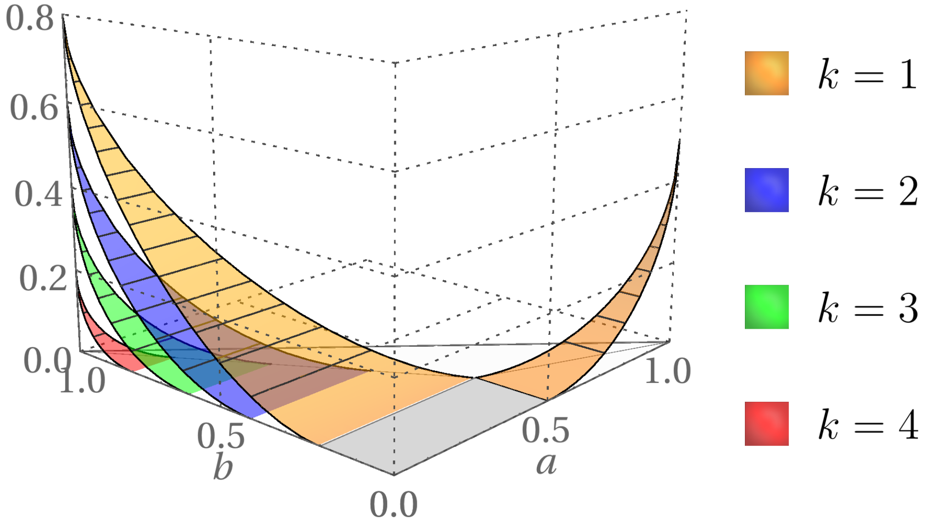


Figure 5.6: Convex roofs of the Vidal monotones E_k (for $d = 5$ and $k = 1, 2, 3, 4$) evaluated on regions A and B of the OO-invariant states. For the surfaces on the left-hand side, k varies from 1 to 4 from the upper to the lower level. Only E_1 is non-vanishing on the right-hand side.

in regions A of Figs. 5.1 and 5.2 and regions B of Figs. 5.1 and 5.3. It is noteworthy that the value of the convex roof for any entanglement monotone for these states depends only on one of the expectations $\langle \rho, W_{-} \rangle$ or $\langle \rho, \Phi_d \rangle$. As an example, the convex roofs of the Vidal monotones on the OO-invariant states with dimension $d = 5$ are shown in Fig 5.6.

If $E_{\text{wer}}(a)$ and $E_{\text{iso}}(b)$ are not convex, e.g. if there is some value b so that $\widehat{E_{\text{iso}}}(b) < E_{\text{iso}}(b)$, then we may still extend the formula to all of these states as long as E is continuous.

5.4 Entanglement conversion witnesses for pure-to-mixed states

It was shown in [GG08] that a pure state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ can be converted into an arbitrary mixed state $\rho \in \mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ if and only if there exists an pure state decomposition $\{p_i, |\varphi_i\rangle\}$ of ρ that satisfies

$$E_k(\psi) \geq \sum_i p_i E_k(\varphi_i)$$

for all positive integers k, where $\rho = \sum_i p_i|\varphi_i\rangle\langle\varphi_i|$. This necessary and sufficient condition for LOCC transformation can be encoded into the following complete conversion witness:

$$W(\psi,\rho) = \max_{\{p_i,|\varphi_i\rangle\}} \min_k \left(E_k(\psi) - \sum_i p_i E_k(\varphi_i) \right), \tag{5.50}$$

where the maximum is taken over all pure state decompositions $\rho = \sum_i p_i|\varphi_i\rangle\langle\varphi_i|$. The function W is a complete witness in the sense that $W(\psi,\rho) \geq 0$ if and only if $|\psi\rangle$ can be converted into ρ via LOCC. Although this function cannot be computed for arbitrary mixed states, it can be simplified for certain classes of mixed states ρ. In particular, we compute $W(\psi,\rho)$ explicitly in the case when ρ is a state on $\mathbb{C}^2 \otimes \mathbb{C}^d$ for any d (i.e. in the case when at least one subsystem is a qubit). We can also make extensive use of symmetry to compute $W(\psi,\rho)$ in the case when ρ is highly symmetric (i.e. Werner or isotropic states).

The conversion witness in (5.50) simplifies to a known necessary and sufficient condition for converting a pure state $|\psi\rangle$ to a mixed state ρ in the case when ρ is a state of a system in which one subsystem is a qubit [Vid00b]. Indeed, for pure states $|\varphi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^d$ with any $d \geq 2$, it holds that $E_k(\varphi) = 0$ whenever $k \geq 2$ since $|\varphi\rangle$ can have at most two nonzero Schmidt coefficients. If ρ is any mixed state on $\mathbb{C}^2 \otimes \mathbb{C}^d$, then the minimization over k in (5.50) can be eliminated, since only E_1 can be nonzero. In this case, the conversion witness in (5.50) simplifies to $W(\psi,\rho) = E_1(\psi) - \widehat{E_1}(\rho)$. This implies the following theorem.

Theorem 5.16. *For any bipartite mixed state ρ on $\mathbb{C}^2 \otimes \mathbb{C}^d$ and for any bipartite pure state $|\psi\rangle$ of systems of any size, it holds that $|\psi\rangle \xrightarrow{\text{LOCC}} \rho$ if and only if $E_1(\psi) \geq \widehat{E_1}(\rho)$.*

Furthermore, it was shown in [Vid00b] that $\widehat{E_1}$ for an arbitrary mixed state of two qubits simplifies to

$$\widehat{E_1}(\rho) = \frac{1 - \sqrt{1 - C(\rho)^2}}{2},$$

where $C(\rho)$ is the concurrence [Woo98] of ρ. Hence, a pure state $|\psi\rangle$ can be converted into a mixed state ρ on $\mathbb{C}^2 \otimes \mathbb{C}^2$ if and only if $C(\psi) \geq C(\rho)$.

As the following theorem shows, the value of $\widehat{E_1}$ gives a necessary and sufficient condition for converting any pure states into Werner states of arbitrary dimension as well.

Theorem 5.17. *For any bipartite pure state $|\psi\rangle$ and any $a \in [\frac{1}{2}, 1]$, it holds that $|\psi\rangle \xrightarrow{\text{LOCC}} \rho_{\text{wer}}(a)$ if and only if $\lambda_1 \le \frac{1}{2} + \sqrt{a(1-a)}$, where λ_1 is the largest Schmidt coefficient of $|\psi\rangle$.*

Note that if $a \in [0, \frac{1}{2}]$ then $\rho_{\text{wer}}(a)$ is separable and thus $|\psi\rangle \xrightarrow{\text{LOCC}} \rho_{\text{wer}}(a)$ holds trivially. The theorem states the conditions for conversion in the case when $\rho_{\text{wer}}(a)$ is entangled.

Proof. Let $a \in [\frac{1}{2}, 1]$ and suppose that $|\psi\rangle \xrightarrow{\text{LOCC}} \rho_{\text{wer}}(a)$. Then it must be the case that $E_1(\psi) \ge \widehat{E_1}(\rho_{\text{wer}}(a))$ since E_1 is an entanglement monotone. The result follows, since $E_1(\psi) = 1 - \lambda_1$ and $\widehat{E_1}(\rho_{\text{wer}}(a)) = \frac{1}{2} - \sqrt{a(1-a)}$. On the other hand, if $\lambda_1 \le \frac{1}{2} + \sqrt{a(1-a)}$ then $\boldsymbol{\lambda} \prec \boldsymbol{\lambda}^a$, where $\boldsymbol{\lambda}$ is the vector of Schmidt coefficients of $|\psi\rangle$ and $\boldsymbol{\lambda}^a$ is the vector of Schmidt coefficients of $|\psi_a\rangle$ given in (5.19). It follows that $|\psi\rangle$ can be converted into $|\psi_a\rangle$ by LOCC, but $|\psi_a\rangle$ can be converted into $\rho_{\text{wer}}(a)$ via LOCC, since $\mathcal{T}_{\text{wer}}(|\psi_a\rangle\langle\psi_a|) = \rho_{\text{wer}}(a)$ and the twirling operation $\mathcal{T}_{\text{wer}}$ is LOCC. This concludes the proof. □

We have shown that the conversion witness in (5.50) can be computed explicitly in the cases when ρ is a Werner state or any state on a $\mathbb{C}^2 \otimes \mathbb{C}^d$ system, but it remains unknown if it can be computed explicitly for any other classes of states. However, it may still be useful to consider upper and lower bounds of this quantity, since these would give either necessary or sufficient conditions for LOCC conversion from $|\psi\rangle$ into ρ. In particular, in the case when $\rho = \rho_{\text{iso}}(b)$ is an isotropic state, a lower bound for (5.50) can be found. The following theorem gives a no-go conversion witness for detecting when pure states cannot be converted into isotropic states.

Theorem 5.18. *Let $|\psi\rangle$ be a pure state and $b \in [\frac{1}{d}, 1]$. If $|\psi\rangle \xrightarrow{\text{LOCC}} \rho_{\text{iso}}(b)$ then $W_{\text{iso}}(\boldsymbol{\lambda}, b) \ge 0$, where*

$$W_{\text{iso}}(\boldsymbol{\lambda}, b) = \max_{\boldsymbol{\mu}} \min_{k} \Big(E_k(\boldsymbol{\lambda}) - E_k(\boldsymbol{\mu}) \Big) \tag{5.51}$$

and the the maximum is taken over all Schmidt vectors $\boldsymbol{\mu}$ that satisfy $\sum_i \sqrt{\mu_i} = \sqrt{db}$.

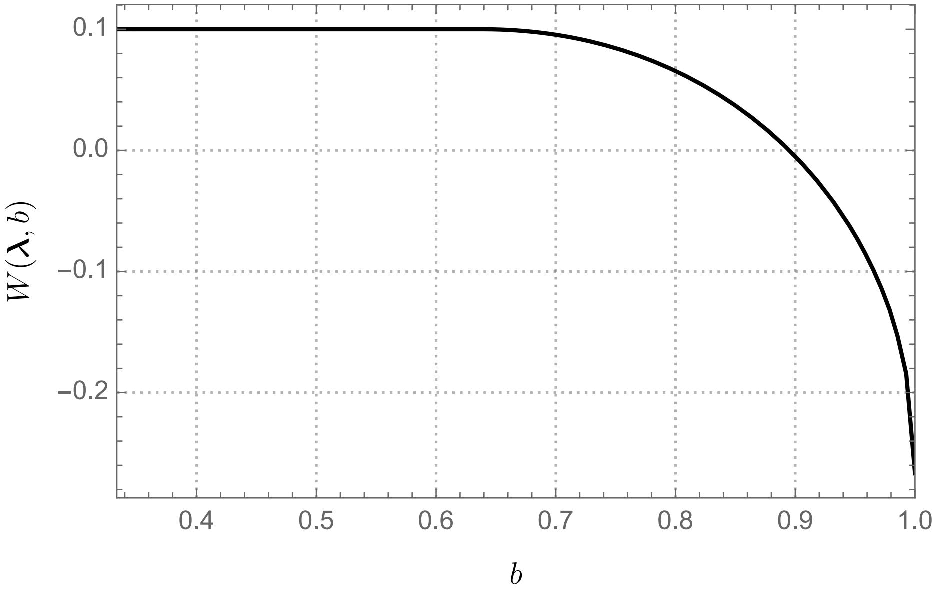


Figure 5.7: An example of the witness in (5.51) computed for $d = 3$ and $\boldsymbol{\lambda} = (\frac{6}{10}, \frac{3}{10}, \frac{1}{10})$. It appears that $W(\boldsymbol{\lambda}, b) < 0$ whenever $b > 0.895$.

Proof. In the case when $\rho = \rho_{\text{iso}}(b)$, it is clear that a lower bound for the witness W in (5.50) can be given by

$$W(\psi, \rho_{\text{iso}}(b)) \geq \max_{|\varphi\rangle} \min_{k} \Big(E_k(\psi) - E_k(\varphi) \Big), \tag{5.52}$$

where the maximum is taken over all $|\varphi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ such that $\langle\varphi|\Phi_d|\varphi\rangle = b$. The left-hand side of the inequality in (5.52) can be further simplified to the desired expression in (5.51). □

In particular, if $W_{\text{iso}}(\psi, b) < 0$ then $|\psi\rangle \overset{\text{LOCC}}{\not\longrightarrow} \rho_{\text{iso}}(b)$. Although the formula for this witness is now much simpler than the general one in (5.50), it still cannot be computed analytically for arbitrary Schmidt vectors $\boldsymbol{\lambda}$. However, we present a way to numerically compute these witnesses efficiently in Appendix A. An example of the witness in (5.51) with $d = 3$ and $\boldsymbol{\lambda} = (\frac{6}{10}, \frac{3}{10}, \frac{1}{10})$ is shown in Fig 5.7. In this case, it appears that $W(\boldsymbol{\lambda}, b) < 0$ whenever $b > 0.895$. Hence the conversion $|\psi\rangle \xrightarrow{\text{LOCC}} \rho_{\text{iso}}(b)$ is not possible when $b > 0.895$, where $|\psi\rangle$ is the pure state with Schmidt coefficients $\boldsymbol{\lambda}$.

5.5 Summary

We computed the convex roof of entanglement monotones on certain classes of symmetric states. This generalized the work of Refs. [VW01] and [TV00], where the entanglement of formation was computed for Werner and isotropic states. In particular, we computed the convex roof for any entanglement monotone on Werner states. The convex roof of certain types of monotones was also computed on isotropic states. We were able to extend these formula for the convex roofs to many non-symmetric states as well. In particular, for many states with other types of symmetries (i.e., for OO-invariant states as well as phase-permutation Werner and isotropic type states), we were also able to compute the convex roofs of these monotones.

We also constructed a necessary and sufficient condition in the form of a conversion witness that determines when a bipartite pure state can be converted to any Werner state by LOCC. A similar conversion witness was constructed for detecting when a pure state can be converted into an isotropic state, but the condition was only necessary and not sufficient.

This work sheds light on the structure of bipartite entanglement of symmetric states, an area of research that is still quite active. Recently, work has been done on computing convex roofs of certain entanglement monotones on larger classes of symmetric states [SEG$^+$16]. Investigations into further types of symmetries and other entanglement monotones will prove fruitful in the complete characterization of the LOCC convertibility of bipartite quantum entanglement.

Chapter 6

Convex quantum resource theories

Recall that a *resource theory* in quantum information $(\mathscr{F}, \mathscr{O})$ is defined in terms of its *free states* $\mathscr{F}$ and the *free operations* $\mathscr{O}$. In any resource theory, the main problem in the single-shot setting is to determine necessary and sufficient conditions for the convertibility of resource states. That is, given resources ρ and ρ', we would like to have convenient conditions for when there exists a free operation that converts ρ into ρ'. Finding a complete set of *resource monotones* is one method of giving such conditions, but such monotones are only useful if they can be computed in practice. This chapter introduces the study of *convex* resource theories in quantum mechanics, in which the sets of free states and free operations are both closed and convex. Using tools from convex analysis (in particular those of conic programming), necessary and sufficient conditions for conversion of resources within these resource theories.

We first recall some of the definitions of resource theories in quantum information that were outlined in Section 2.5. Let $(\mathscr{F}, \mathscr{O})$ be a resource theory where $\mathscr{F}$ and $\mathscr{O}$ are families of density matrices and quantum channels respectively such that $\mathscr{F}(\mathcal{H}) \subset \mathrm{D}(\mathcal{H})$ denote subsets of *free* density matrices on system $\mathcal{H}$ and $\mathscr{O}(\mathcal{H}, \mathcal{H}')$ denote the subset of *free* operations from states on system $\mathcal{H}$ to states on system $\mathcal{H}'$. The *resources* in such a resource theory are all density matrices. We write $\rho \xrightarrow{\mathscr{O}} \rho'$ if there exists a free channel $\mathcal{E} \in \mathscr{O}$ such that $\mathcal{E}(\rho) = \rho'$.

The resource theory is said to be *closed* if $\mathscr{F}_{\mathcal{H}}$ and $\mathscr{O}_{\mathcal{H},\mathcal{H}'}$ are topologically closed as subsets of states and channels for each $\mathcal{H}$ and $\mathcal{H}'$, and is said to be *convex* if its sets of free states and channels are convex.

The resource theory of bipartite entanglement under LOCC is an example of a resource theory that is convex, but not closed [CLM+14]. Meanwhile, the resource theory of non-Gaussianity is an example of a resource theory that is not convex [BESP03].

This chapter first develops the tools for studying convex resource theories in quantum information and developing necessary and sufficient conditions for resource transformation for general convex resource theories. These results will be used to study the *resource theory of PPT states and channels*, in which the free states are the PPT density matrices of bipartite quantum systems and the free operations are the channels whose Choi representation is PPT (defined in Section 2.4.2). In particular, a complete set of resource monotones that can be computed using semidefinite programming techniques are constructed for this resource theory. The *resource theory of separable states and channels* will also be analyzed, but these results are less practical since determining whether an operator is separable is difficult in practice. Most notably, however, we will show that any resource theory where the set of free operations consists of any finite polytope approximation to the set of separable channels will not lead to a resource theory with bound entanglement. In particular, any finite approximation to the set of separable channels must contain an entangling operation, and some two-qubit entangled state must be free in such a resource theory.

The remainder of this chapter will be dedicated to studying convertibility criteria in closed, convex quantum resource theories. Section 6.1 introduces some necessary mathematical tools that will be needed to study convex resource theories. For such resource theories, we can find a complete set of resource monotones that can be computed as the solution to a conic program, which will be presented in Section 6.2. Applications of this result to some resource theories will be presented in section 6.4.

6.1 Tools for studying convex resource theories

The primary purpose of this section is to lay the mathematical groundwork for complete characterization of convex quantum resource theories. Primarily, we will lay the groundwork for determining when there exists a channel $\mathcal{E}$ in some restricted convex set of channels such that $\mathcal{E}(\rho) = \rho'$ for some given states ρ and ρ'. Rather than work with the channels directly, it will be more convenient to work with the Choi representations of the channels $J(\mathcal{E})$. When $\mathcal{E} : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H}')$ is a linear map of linear operators, its Choi representation is an operator $J(\mathcal{E}) \in \mathrm{L}(\mathcal{H}' \otimes \mathcal{H})$. For much of this chapter, sets of quantum channels from $\mathcal{H}$ to $\mathcal{H}'$ will be identified with the set of Choi matrices of these channels.

Consider a resource theory $(\mathscr{F}, \mathscr{O})$ of quantum states and operations. Given systems $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{A}'}$ in the resource theory, the set of free operations from states on system A to system A' in this resource theory will be denoted $\mathscr{O}_{\mathsf{A},\mathsf{A}'}$. If $\mathscr{O}_{\mathsf{A},\mathsf{A}'}$ is closed and convex, a channel $\mathcal{E}$ from A to A' is free if and only if it is in $\mathrm{cone}(\mathscr{O}_{\mathsf{A},\mathsf{A}'})$ (i.e. in the conic hull of the set of free operations). It is typically easier to instead consider the Choi matrices of such channels, and consider the cone in the Choi representation defined by $\mathcal{K} = \mathrm{cone}(J(\mathscr{O}_{\mathsf{A},\mathsf{A}'}))$ as a subset $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}'} \otimes \mathcal{H}_{\mathsf{A}})$, where $J(\mathscr{O}_{\mathsf{A},\mathsf{A}'})$ denotes the set of Choi matrices of channels in $\mathscr{O}_{\mathsf{A},\mathsf{A}'}$. The main idea of convex analysis that is used in this chapter is that the *dual cone* $\mathcal{K}^* = (J(\mathscr{O}_{\mathsf{A},\mathsf{A}'}))^*$ acts as the set of *witnesses* to the set of free channels. That is, if $\mathcal{E}$ is a free channel, then $\langle W, J(\mathcal{E})\rangle \geq 0$ for all $W \in \mathcal{K}^*$. Furthermore, since $\mathscr{O}_{\mathsf{A},\mathsf{A}'}$ is convex and closed, it holds that $\mathcal{K}^{**} = \mathcal{K}$. This is an important fact that will be utilized in this section.

The main result of this section is Theorem 6.1 which states the following. Given states $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1})$ and $\rho_2 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$, and a closed convex cone $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})$, there exists a quantum channel $\mathcal{E}$ converting ρ_1 to ρ_2 such that $J(\mathcal{E}) \in \mathcal{K}$ if and only if, for all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$, all $W \in \mathcal{K}^*$, and all $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$, either

$$\mathbb{1}_{\mathsf{A}_2} \otimes Y \not\geq \sigma \otimes \rho_1^T + W \tag{6.1}$$

or $\mathrm{Tr}(Y) \geq \langle \sigma, \rho_2 \rangle$. In particular, if we can find operators $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$, $W \in \mathcal{K}^*$, and $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$ satisfying both $\mathbb{1}_{\mathsf{A}_2} \otimes Y \geq \sigma \otimes \rho_1^T + W$ and $\mathrm{Tr}(Y) < \langle \sigma, \rho_2 \rangle$, then it must be the case that ρ_1 cannot be converted to ρ_2 with any quantum channel in the cone $J(\mathcal{E}) \in \mathcal{K}$. This condition is a *no-go* conversion witness for convertibility.

The proof of this main result makes use of conic programming techniques introduced in Section 3.4.2. In particular, the conic version of Farkas' Lemma (Theorem 3.12), which gives conditions for one or another systems of equations to have a solution (but not both), is used to give the necessary and sufficient condition stated here.

As we shall see, this necessary and sufficient condition for convertibility can also be stated in terms of computing the value of certain quantities, q_{c} and $M_{\mathcal{K},\sigma}$, which will be defined in Section 6.1.1. The main result stating necessary and sufficient conditions for conversion are stated in Section 6.1.2. Properties of the function $M_{\mathcal{K},\sigma}$ will be discussed in Section 6.1.3, including conic program formulations.

6.1.1 A quantity related to $H_{\min}$

Let $\mathcal{H}_{\mathsf{A}_1}$ and $\mathcal{H}_{\mathsf{A}_2}$ be finite dimensional Hilbert spaces. A measure of "quantum correlations" of density operators $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})$ is given by

$$\mathsf{q}_{\mathsf{c}}(\mathsf{A}_2|\mathsf{A}_1)_\rho := \inf\{\mathrm{Tr}(Y) \,|\, Y \in \mathrm{H}(\mathcal{H}_{\mathsf{B}}) \text{ s.t. } \mathbb{1}_{\mathsf{A}} \otimes Y \geq \rho\}. \tag{6.2}$$

It is a measure of the amount of information that can be gained about system A_2 by measuring system A_1. This quantity is related to the (*conditional*) *quantum min-entropy* [KRS09] of ρ by

$$\mathsf{q}_{\mathsf{c}}(\mathsf{A}|\mathsf{B})_\rho = 2^{-H_{\min}(\mathsf{A}_2|\mathsf{A}_1)_\rho}, \tag{6.3}$$

which is a single-shot analog of the conditional quantum entropy [DMHB13]. The conditional entropy measures the average uncertainty one has about system A_2 after measuring system A_1, and it may be interpreted as the distance of ρ from a maximally entangled state

[KRS09]. This concept is useful in quantum cryptography, for example in the context of privacy amplification [VV14] and numerous other applications in single-shot quantum information processing [Ren05, KRS09, TCR09].

If the systems A_1 and A_2 are clear from context, we simply write $\mathsf{q_c}(\rho)$ for $\mathsf{q_c}(\mathsf{A}_2|\mathsf{A}_1)_\rho$. This function can also be extended to arbitrary operators $Z \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})$ that are not density operators,

$$\mathsf{q_c}(Z) := \inf\{\mathrm{Tr}(Y) \,|\, Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1}) \text{ s.t. } \mathbb{1}_{\mathsf{A}_2} \otimes Y \geq Z\}, \tag{6.4}$$

although it is unclear what the physical interpretation of $\mathsf{q_c}(Z)$ is if $Z \ngeq 0$. Nonetheless, it will play an important role in the analysis of convex resource theories in this chapter.

The function $\mathsf{q_c} : \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1}) \to \mathbb{R}$ has the following useful properties. For a given $Z \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})$, the quantity $\mathsf{q_c}(Z)$ can be computed from the following pair of dual semidefinite programs:

$$\begin{array}{rlcrl} \text{maximize:} & \langle Z, X\rangle & \qquad & \text{minimize:} & \mathrm{Tr}(Y) \\ \text{subject to:} & \mathrm{Tr}_{\mathsf{A}_2}(X) = \mathbb{1}_{\mathsf{A}_1} & & \text{subject to:} & \mathbb{1}_{\mathsf{A}_2} \otimes Y \geq Z \\ & X \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1}),\ X \geq 0 & & & Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1}) \end{array} \tag{6.5}$$

Strict feasibility holds in both the primal and dual problems in (6.5), so strong duality follows from Slater's theorem (Theorem 3.11). Indeed, we may choose $X = \frac{1}{\mathrm{Tr}(\mathbb{1}_{\mathsf{A}_2})}\mathbb{1}_{\mathsf{A}_2} \otimes \mathbb{1}_{\mathsf{A}_1}$ so that $X > 0$ and $\mathrm{Tr}_{\mathsf{A}_2}(X) = \mathbb{1}_{\mathsf{A}_1}$ to see that the primal problem is strictly feasible, while we may choose an operator $Y > \lambda_{\max}(Z)\mathbb{1}_{\mathsf{A}_1}$ such that $\mathbb{1}_{\mathsf{A}_2} \otimes Y > Z$ to see that the dual problem is strictly feasible.

6.1.2 Condition for the existence of a channel in a cone

Let $\rho \in \mathrm{D}(\mathcal{H})$ and $\rho' \in \mathrm{D}(\mathcal{H}')$ be density operators and let $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}' \otimes \mathcal{H})_+$ be a closed, convex cone of positive operators. In this section, we determine the necessary and sufficient conditions for the existence of a quantum channel $\mathcal{E}$ whose Choi matrix is in the cone $J(\mathcal{E}) \in \mathcal{K}$ that converts ρ to ρ'.

The main result of this section is given in Theorem 6.1, which states useful necessary and sufficient conditions for the existence of a channel in a cone $\mathcal{K}$ that converts one state to another. Before stating the theorem, we must first define the following function. Given a cone $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})$ and a density operator, define the function $M_{\mathcal{K},\sigma} : \mathrm{D}(\mathcal{H}_{\mathsf{A}_1}) \to \mathbb{R}$ by

$$M_{\mathcal{K},\sigma}(\rho) := \inf_{W \in \mathcal{K}^*} \mathsf{q_c}(\sigma \otimes \rho^T + W). \tag{6.6}$$

Further properties of this function, including a conic program formulation for its computation, will be explored in Section 6.1.3. Later, in Section 6.2, it will be shown that the function $M_{\mathcal{K},\sigma}$ form a complete set of resource monotones in convex resource theories.

Theorem 6.1. *Let $\mathcal{H}_{\mathsf{A}_1}$ and $\mathcal{H}_{\mathsf{A}_2}$ be Hilbert spaces, let $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})_+$ be a closed convex cone, and let $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1})$ and $\rho_2 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$. The following are equivalent.*

(i) There exists a quantum channel $\mathcal{E}$ such that $J(\mathcal{E}) \in \mathcal{K}$ and $\mathcal{E}(\rho_1) = \rho_2$.

(ii) For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$, all $W \in \mathcal{K}^*$, and all $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$, if

$$\mathbb{1}_{\mathsf{A}_2} \otimes Y \geq \sigma \otimes \rho_1^T + W$$

then $\mathrm{Tr}(Y) \geq \mathrm{Tr}(\sigma\rho_2)$.

(iii) For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$ and all $W \in \mathcal{K}^*$, it holds that

$$\mathsf{q_c}(\sigma \otimes \rho_1^T + W) \geq \mathrm{Tr}(\sigma\rho_2). \tag{6.7}$$

(iv) For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$, it holds that

$$M_{\mathcal{K},\sigma}(\rho_1) \geq \mathrm{Tr}(\sigma\rho_2). \tag{6.8}$$

Here, $\mathsf{q_c}$ *and* $M_{\mathcal{K},\sigma}$ *are the functions defined in* (6.4) *and* (6.6)*, respectively.*

Proof. The proof of the equivalence of statements (i) and (ii) of Theorem 6.1 relies on conic programming techniques, and follows directly from Lemma 6.2 (which is stated and proved below).

The equivalence of statements (ii), (iii), and (iv) is straightforward from the definitions of $\mathsf{q_c}(\cdot)$ and $M_{\mathcal{K},\sigma}(\cdot)$. Indeed, consider an arbitrary density operator $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$. To prove the equivalence of (ii) and (iii), let $W \in \mathcal{K}^*$ be arbitrary. Then, by definition of $\mathsf{q_c}(\cdot)$ it holds that $\mathsf{q_c}(\sigma \otimes \rho_1^T + W) \geq \mathrm{Tr}(\sigma\rho_2)$ for all Y if and only if $\mathrm{Tr}(Y) \geq \mathrm{Tr}(\sigma\rho_2)$ holds for all Y satisfying $\mathbb{1}_{\mathsf{A}_2} \otimes Y \geq \sigma \otimes \rho_1^T + W$. Analogously, by definition of $M_{\mathcal{K},\sigma}(\cdot)$, it holds that $M_{\mathcal{K},\sigma}(\rho_1) \geq \mathrm{Tr}(\sigma\rho_2)$ if and only if $\mathsf{q_c}(\sigma \otimes \rho_1^T + W) \geq \mathrm{Tr}(\sigma\rho_2)$ holds for all $W \in \mathcal{K}^*$. □

Lemma 6.2 provides the bulk of the proof of Theorem 6.1. Indeed, the equivalence of statements (i) and (ii) of Theorem 6.1 follows directly from Lemma 6.2. A linear map $\mathcal{E} : \mathrm{H}(\mathcal{H}_{\mathsf{A}_1}) \to \mathrm{H}(\mathcal{H}_{\mathsf{A}_2})$ is a quantum channel if and only if its Choi representation $X = J(\mathcal{E})$ satisfies $X \geq 0$ and $\mathrm{Tr}_{\mathsf{A}_2}(X) = \mathbb{1}_{\mathsf{A}_1}$. Furthermore, it holds that $\mathcal{E}(\rho_1) = \rho_2$ if and only if $\mathrm{Tr}_{\mathsf{A}_1}(X(\mathbb{1}_{\mathsf{A}_2} \otimes \rho_1^T)) = \rho_2$. Since we assume that $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})_+$ (i..e, the cone is a subset of the positive operators), any $X \in \mathcal{K}$ automatically satisfies $X \geq 0$. The proof Lemma 6.2 of makes use of the conic version of Farkas' Lemma (discussed in Section 3.4.3).

Lemma 6.2. *Let* $\mathcal{H}_{\mathsf{A}_1}$ *and* $\mathcal{H}_{\mathsf{A}_2}$ *be Hilbert spaces, let* $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})_+$ *be a closed convex cone, and let* $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1})$ *and* $\rho_2 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$*. The following are equivalent.*

(i) There exists $X \in \mathcal{K}$ *so that* $\mathrm{Tr}_{\mathsf{A}_2}(X) = \mathbb{1}_{\mathsf{A}_1}$ *and* $\mathrm{Tr}_{\mathsf{A}_1}(X(\mathbb{1}_{\mathsf{A}_2} \otimes \rho_1^T)) = \rho_2$*.*

(ii) For all $Y_1 \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$ *and* $Y_2 \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2})$*, and all* $W \in \mathcal{K}^*$*, if*

$$\mathbb{1}_{\mathsf{A}_2} \otimes Y_1 + Y_2 \otimes \rho_1^T \geq W$$

then $\mathrm{Tr}(Y_1) + \mathrm{Tr}(Y_2\rho_2) \geq 0$*.*

(iii) For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$, all $W \in \mathcal{K}^*$, and all $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$, if

$$\mathbb{1}_{\mathsf{A}_2} \otimes Y \geq \sigma \otimes \rho_1^T + W$$

then $\mathrm{Tr}(Y) \geq \mathrm{Tr}(\sigma\rho_2)$.

Proof. The proof the equivalence of (i) and (ii) is an application of the conic version of Farkas' Lemma. Indeed, consider the linear map $\Phi : \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1}) \to \mathrm{H}(\mathcal{H}_{\mathsf{A}_1}) \oplus \mathrm{H}(\mathcal{H}_{\mathsf{A}_2})$ defined by

$$\Phi(X) = \begin{pmatrix} \mathrm{Tr}_{\mathsf{A}_2}(X) & 0 \\ 0 & \mathrm{Tr}_{\mathsf{A}_1}(X(\mathbb{1}_{\mathsf{A}_2} \otimes \rho_1^T)) \end{pmatrix}.$$

Then an operator $X \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})$ satisfies $\mathrm{Tr}_{\mathsf{A}_2}(X) = \mathbb{1}_{\mathsf{A}_1}$ and $\mathrm{Tr}_{\mathsf{A}_1}(X(\mathbb{1}_{\mathsf{A}_2} \otimes \rho_1^T)) = \rho_2$ if and only if it satisfies

$$\Phi(X) = \begin{pmatrix} \mathbb{1}_{\mathsf{A}_1} & 0 \\ 0 & \rho_2 \end{pmatrix}. \tag{6.9}$$

The dual of this map is given by

$$\Phi^*\left(\begin{pmatrix} Y_1 & \cdot \\ \cdot & Y_2 \end{pmatrix}\right) = \mathbb{1}_{\mathsf{A}_2} \otimes Y_1 + Y_2 \otimes \rho_1^T.$$

To make use of the conic version of Farkas' Lemma, we first prove that $\Phi(\mathcal{K})$ is closed. Note that

$$\Phi^*\left(\begin{pmatrix} \mathbb{1}_{\mathsf{A}_1} & 0 \\ 0 & 0 \end{pmatrix}\right) = \mathbb{1}_{\mathsf{A}_2} \otimes \mathbb{1}_{\mathsf{A}_1}$$

and $\mathbb{1}_{\mathsf{A}_2} \otimes \mathbb{1}_{\mathsf{A}_1} \in \mathrm{int}(\mathrm{H}(\mathcal{H}_2 \otimes \mathcal{H}_1)) \subseteq \mathrm{int}(\mathcal{K})$. Hence $\Phi(\mathcal{K})$ is closed by Lemma 3.13. By the conic version of Farkas' Lemma (Theorem 3.12), there exists an $X \in \mathcal{K}$ satisfying (6.9) if and only if

$$\mathrm{Tr}(Y_1) + \langle Y_2, \rho_2 \rangle = \left\langle \begin{pmatrix} Y_1 & \cdot \\ \cdot & Y_2 \end{pmatrix}, \begin{pmatrix} \mathbb{1}_{\mathsf{A}_1} & 0 \\ 0 & \rho_2 \end{pmatrix} \right\rangle \geq 0$$

holds for all $Y_1 \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$ and $Y_2 \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2})$ such that

$$\Phi^*\left(\begin{pmatrix} Y_1 & 0 \\ 0 & Y_2 \end{pmatrix}\right) = \mathbb{1}_{\mathsf{A}_2} \otimes Y_1 + Y_2 \otimes \rho_1^T \in \mathcal{K}^*. \tag{6.10}$$

Since $\mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})_+ \subseteq \mathcal{K}^*$, the condition in (6.10) holds if and only if there exists a $W \in \mathcal{K}^*$ such that

$$\mathbb{1}_{\mathsf{A}_2} \otimes Y_1 + Y_2 \otimes \rho_1^T \geq W.$$

Indeed, it holds that $\mathbb{1}_{\mathsf{A}_2} \otimes Y_1 + Y_2 \otimes \rho_1^T \geq W$ if and only if there is a positive operator $P \geq 0$ such that $\mathbb{1}_{\mathsf{A}_2} \otimes Y_1 + Y_2 \otimes \rho_1^T = P + W$. Since $P \in \mathcal{K}^*$ as well, it follows that $P + W \in \mathcal{K}^*$ since $\mathcal{K}^*$ is a cone. This completes the proof of the equivalence of (i) and (ii).

Note that the implication $\neg$(iii)$\Rightarrow$ $\neg$(ii) is clear, since any counter example to (iii) is also a counter example to (ii). Indeed, if the operators W, σ, and Y provide a counterexample to (iii), we may choose the operators $Y_1 = Y$ and $Y_2 = -\sigma$ to provide a counterexample to (ii). So it remains to show that $\neg$(ii)$\Rightarrow$ $\neg$(iii). Suppose there exist operators $Y_1 \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$, $Y_2 \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2})$, and $W \in \mathcal{K}^*$ such that $\mathbb{1}_{\mathsf{A}_2} \otimes Y_1 + Y_2 \otimes \rho_1^T \geq W$ but $\mathrm{Tr}(Y_1) + \mathrm{Tr}(Y_2\rho_2) < 0$. We may choose a real number $t > 0$ such that

$$t\left(Y_2 - \frac{\mathrm{Tr}(Y_2)}{\mathrm{Tr}(\mathbb{1}_{\mathsf{A}_2})}\mathbb{1}_{\mathsf{A}_2}\right) \leq \frac{1}{\mathrm{Tr}(\mathbb{1}_{\mathsf{A}_2})}\mathbb{1}_{\mathsf{A}_2}.$$

Indeed, the value $1/t = d\lambda_{\max}(Y_2) - \mathrm{Tr}(Y_2)$ suffices when $Y \not\propto \mathbb{1}_{\mathsf{A}_2}$. Then we may set

$$\begin{aligned} \sigma &= \frac{1}{\mathrm{Tr}(\mathbb{1}_{\mathsf{A}_2})}\mathbb{1}_{\mathsf{A}_2} - t\left(Y_2 - \frac{\mathrm{Tr}(Y_2)}{\mathrm{Tr}(\mathbb{1}_{\mathsf{A}_2})}\mathbb{1}_{\mathsf{A}_2}\right) \\ \text{and} \quad Y &= tY_1 + \frac{1 + t\,\mathrm{Tr}(Y_2)}{\mathrm{Tr}(\mathbb{1}_{\mathsf{A}_2})}\rho_1^T \end{aligned}$$

such that $\operatorname{Tr}(\sigma\rho_2) = \frac{1+t\operatorname{Tr}(Y_2)}{\operatorname{Tr}(\mathbb{1}_{\mathsf{A}_2})} - t\operatorname{Tr}(Y_2\rho_2)$ and

$$\begin{aligned}\operatorname{Tr}(Y) &= t\operatorname{Tr}(Y_1) + \frac{1+t\operatorname{Tr}(Y_2)}{\operatorname{Tr}(\mathbb{1}_{\mathsf{A}_2})} \\ &< \operatorname{Tr}(\sigma\rho_2).\end{aligned}$$

Note that $t\mathbb{1}_{\mathsf{A}_2} \otimes Y_1 \geq t(-Y_2 \otimes \rho_1^T + W)$ since $t > 0$, and thus

$$\begin{aligned}\mathbb{1}_{\mathsf{A}_2} \otimes Y &= \frac{1+t\operatorname{Tr}(Y_2)}{\operatorname{Tr}(\mathbb{1}_{\mathsf{A}_2})}\mathbb{1}_{\mathsf{A}_2} \otimes \rho_1^T + t\mathbb{1}_{\mathsf{A}_2} \otimes Y_1 \\ &\geq \frac{1+t\operatorname{Tr}(Y_2)}{\operatorname{Tr}(\mathbb{1}_{\mathsf{A}_2})}\mathbb{1}_{\mathsf{A}_2} \otimes \rho_1^T - tY_2 \otimes \rho_1^T + tW \\ &= \sigma \otimes \rho_1^T + tW,\end{aligned}$$

where $tW \in \mathcal{K}^*$, which yields a counterexample to (iii). This completes the proof. □

Statements (ii) and (iii) of Theorem 6.1 are a sort of *conversion witness* since they give simple conditions for when it is *not* possible to find a channel $\mathcal{E}$ in the cone $\mathcal{K}$ such that $\mathcal{E}(\rho_1) = \rho_2$. That is, if we can find a density operator $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}'})$, along with a witness $W \in \mathcal{K}^*$ and an operator $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$, such that $\operatorname{Tr} Y < \langle \rho_2, \sigma \rangle$ but

$$\mathbb{1}_{\mathsf{A}_2} \otimes X \geq \sigma \otimes \rho^T + W,$$

then it is *not* possible to find a channel in the cone that converts ρ_1 to ρ_2. This can be a useful way to characterize when conversions between states in resource theories are not possible without resorting to monotones. However, as we shall see in the next section, Theorem 6.1 can be used to construct a complete family of resource monotones for quantum resource theories, which can be useful is certain cases.

6.1.3 Properties of $M_{\mathcal{K},\sigma}$

For a given cone $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})_+$, from the definition of $M_{\mathcal{K},\sigma}$ we see that the value $M_{\mathcal{K},\sigma}(\rho)$ can be computed as the solution to the following optimization problem:

$$\begin{aligned} \text{minimize:} \quad & \mathrm{Tr}(Y) \\ \text{subject to:} \quad & \mathbb{1}_{\mathsf{A}_2} \otimes Y - \sigma \otimes \rho_1^T \in \mathcal{K}^*. \end{aligned} \tag{6.11}$$

It is straightforward to see that (6.11) is the dual problem of a conic program whose primal problem is:

$$\begin{aligned} \text{maximize:} \quad & \mathrm{Tr}(X(\sigma \otimes \rho^T)) \\ \text{subject to:} \quad & X \in \mathcal{K} \\ & \mathrm{Tr}_{\mathsf{A}_2}(X) = \mathbb{1}_{\mathsf{A}_1}. \end{aligned} \tag{6.12}$$

We note that strong duality holds for the dual and primal problems given in (6.11) and (6.12), respectively (as long as the cone $\mathcal{K}$ contains at least one quantum channel). Indeed, the problem in (6.11) is always strongly feasible, since we may choose any $Y > \mathbb{1}_{\mathsf{A}_1}$ such that $\mathbb{1}_{\mathsf{A}_2} \otimes Y - \sigma \otimes \rho_1 > 0$. Then as long as the problem in (6.12) is feasible, the optimal values of these two problems coincide. Thus the quantity $M_\sigma(\rho)$ may be computed as the optimal value of either problem.

There is another representation of the problem in (6.12) that is more natural from a quantum informational perspective. The requirements that $X \in \mathcal{K}$ and $\mathrm{Tr}_{\mathsf{A}_2}(X) = \mathbb{1}_{\mathsf{A}_1}$ imply that X must be the Choi representation $X = J(\mathcal{E})$ of some quantum channel $\mathcal{E} : \mathrm{H}(\mathcal{H}_{\mathsf{A}_1}) \to \mathrm{H}(\mathcal{H}_{\mathsf{A}_2})$ such that $J(\mathcal{E}) \in \mathcal{K}$. Furthermore, note that

$$\langle J(\mathcal{E}), \sigma \otimes \rho^T \rangle = \langle \sigma, \mathcal{E}(\rho) \rangle$$

for any quantum channel $\mathcal{E}$. Hence the optimization problem in (6.12) may be written as

$$\begin{aligned} \text{maximize:} \quad & \langle \sigma, \mathcal{E}(\rho) \rangle \\ \text{subject to:} \quad & J(\mathcal{E}) \in \mathcal{K} \\ & \mathcal{E} \text{ a quantum channel.} \end{aligned}$$

Hence, by strong duality, the quantities $M_{\mathcal{K},\sigma}(\rho)$ may be expressed as

$$M_{\mathcal{K},\sigma}(\rho) = \inf\{\langle \sigma, \mathcal{E}(\rho) \rangle \,|\, J(\mathcal{E}) \in \mathcal{K}\}, \tag{6.13}$$

where the infimum is taken over all quantum channels $\mathcal{E}$.

For any σ, the fact that $M_{\mathcal{K},\sigma}(\rho) \geq \langle \sigma, \mathcal{E}(\rho) \rangle$ holds for any channel $\mathcal{E}$ with $J(\mathcal{E}) \in \mathcal{K}$ is straightforward from the representation of $M_{\mathcal{K},\sigma}$ given in (6.13). Hence the condition that

$$M_{\mathcal{K},\sigma}(\rho_1) \geq \operatorname{Tr}(\sigma \rho_2) \quad \text{for all } \sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2}) \tag{6.14}$$

is easily seen to be a necessary condition for the existence of a channel in the cone that converts ρ_1 to ρ_2 (i.e., if there exists a $\mathcal{E}$ such that $\mathcal{E}(\rho_1) = \rho_2$). However, the fact that the condition in (6.14) is also a *sufficient* condition for the existence of a channel in the cone is not straightforward, and application of the conic version of Farkas' Lemma is required to prove sufficiency. Finally, we state a few useful facts about $M_{\mathcal{K},\sigma}$ that will be used for characterizing these functions as resource monotones in convex resource theories.

Proposition 6.3. *Let $\mathcal{H}_{\mathsf{A}_1}$ and $\mathcal{H}_{\mathsf{A}_2}$ be finite-dimensional Hilbert spaces, let $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1})$ and $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$ be density operators, and let $\mathcal{K} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{A}_1})_+$ be a closed convex cone.*

1. *It holds that*

$$0 \leq \lambda_{\min}(\sigma) \leq M_{\mathcal{K},\sigma}(\rho) \leq \lambda_{\max}(\sigma) \leq 1. \tag{6.15}$$

2. Suppose $\mathsf{A}_1 = \mathsf{A}_2 = \mathsf{A}$*. If* $J(\mathrm{id}_{\mathsf{A}}) \in \mathcal{K}$ *then*

$$M_{\mathcal{K},\sigma}(\rho) \geq \langle \sigma, \rho \rangle. \tag{6.16}$$

Proof. To prove 1., let $P \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_2})$ be the projection operator onto the eigenvector of σ with maximal eigenvalue such that $\mathrm{Tr}(\sigma P) = \lambda_{\max}(\sigma)$. Note that $0 \leq \lambda_{\max}(\sigma) \leq 1$ since σ is a density operator. Furthermore, computing the maximal eigenvalue of σ can be computed from the semidefinite program

$$\begin{aligned} \text{maximize:}\ & \mathrm{Tr}(\sigma X) \\ \text{subject to:}\ & X \geq 0 \text{ and } \mathrm{Tr}(X) = 1, \end{aligned}$$

and $\mathcal{E}(\rho)$ satisfies both $\mathcal{E}(\rho) \geq 0$ and $\mathrm{Tr}\,\mathcal{E}(\rho) = 1$ since $\mathcal{E}$ is a quantum channel and ρ is a density operator. It follows that $M_{\mathcal{K},\sigma}(\rho) \leq \langle \sigma, P \rangle = \lambda_{\max}(\sigma)$.

Part 2. follows directly from the formulation of $M_{\mathcal{K},\sigma}$ given in (6.13), since

$$M_{\mathcal{K},\sigma}(\rho) \geq \langle \sigma, \mathrm{id}(\rho) \rangle = \langle \sigma, \rho \rangle$$

if the Choi representation of the identity channel resides in $\mathcal{K}$. □

6.2 Complete set of monotones for convex resource theories

Theorem 6.1 of the previous section gives the main necessary and sufficient conditions for conversion between resources in a resource theory that will be used in this thesis, but those conditions are given only as *conversion witnesses.* This section outlines how *resource monotones* can be constructed from these conditions using conic programs. The main result of this section is Theorem 6.5, which states that, for a closed, convex quantum resource theory

$(\mathscr{F},\mathscr{O})$, the family of functions $\{M_{\mathscr{O},\sigma}\}_\sigma$ defined in (6.17) forms a complete set of resource monotones. These monotones are indexed by all possible resource states σ in the resource theory, and they can be computed as solutions to a conic programs.

Given a set of channels $\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}$ from $\mathcal{H}_{\mathsf{A}_1}$ to $\mathcal{H}_{\mathsf{A}_2}$, the set of Choi matrices of these channels is $J(\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}) = \{J(\mathcal{E}) \,|\, \mathcal{E} \in \mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}\}$. If $\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}$ is closed and convex, then determining whether a channel $\mathcal{E}$ is in $\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}$ is equivalent to checking whether $\langle W, J(\mathcal{E})\rangle \geq 0$ for all witnesses in the dual cone $W \in (J(\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}))^*$. We now define the functions that will be used as resource monotones in convex quantum resource theories.

Definition 6.4. Let $(\mathscr{F},\mathscr{O})$ be a quantum resource theory and let $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$ be any density operator in the resource theory. The *$\mathscr{O}$-monotone with respect to σ* is defined as

$$M_{\mathscr{O},\sigma}(\rho) := \inf_{W \in J(\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2})^*} \mathsf{q_c}(\sigma \otimes \rho^T + W) \tag{6.17}$$

for all states $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1})$.

For convenience, it will often be useful to write the dual cone of the set of channels simply as $\mathcal{W} = J(\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2})^*$. The remainder of this section is devoted to proving that the functions $\{M_{\mathscr{O},\sigma}\}_\sigma$ defined in (6.17) do in fact form a complete set of resource monotones. That is, we will show that $M_{\mathscr{O},\sigma}(\rho) \geq M_{\mathscr{O},\sigma}(\mathcal{E}(\rho))$ for any free operator $\mathcal{E} \in \mathscr{O}$ and any resource states ρ and σ, and that $M_{\mathscr{O},\sigma}(\rho) \geq M_{\mathscr{O},\sigma}(\rho')$ holds for all resource states σ if and only if $\rho \xrightarrow{\mathscr{O}} \rho'$.

We first make a few useful remarks. In any resource theory $(\mathscr{F},\mathscr{O})$, note that $\rho \xrightarrow{\mathscr{O}} \rho$ always holds for any resource state ρ since the identity map is always free. We now give the main theorem of this section, which states that the functions $M_{\mathscr{O},\sigma}$ do in fact form a complete set of monotones.

Theorem 6.5. *Let $(\mathscr{F},\mathscr{O})$ be a closed and convex quantum resource theory. Let $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1})$ and $\rho_2 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$ be density operators in the resource theory. The following are equivalent.*

(i) *It holds that $\rho_1 \xrightarrow{\mathscr{O}} \rho_2$.*

(ii) *For all* $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$*, all* $W \in (J(\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}))^*$*, and all* $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1})$*, if*

$$\mathbb{1}_{\mathsf{A}} \otimes Y \geq \sigma \otimes \rho^T + W \tag{6.18}$$

then $\mathrm{Tr}(Y) \geq \langle \sigma, \rho_2 \rangle$.

(iii) *For all Hilbert spaces* $\mathcal{H}_{\mathsf{A}}$ *in the resource theory and all* $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}})$*, it holds that*

$$M_{\mathscr{O},\sigma}(\rho_1) \geq M_{\mathscr{O},\sigma}(\rho_2). \tag{6.19}$$

Proof. The equivalence of (i) and (ii) has already been established in Theorem 6.1. To prove that (i) implies (iii), suppose that $\rho_1 \xrightarrow{\mathscr{O}} \rho_2$. Then there exists a free channel $\mathcal{E} \in \mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}$ such that $\mathcal{E}(\rho_1) = \rho_2$. Note that $\mathcal{E}' \circ \mathcal{E} \in \mathscr{O}_{\mathsf{A}_1,\mathsf{A}}$ for any free channel $\mathcal{E}' \in \mathscr{O}_{\mathsf{A}_2,\mathsf{A}}$, hence

$$\begin{aligned} M_{\mathscr{O},\sigma}(\rho_2) &= M_{\mathscr{O},\sigma}(\mathcal{E}(\rho_1)) \\ &= \max\{\langle \sigma, \mathcal{E}' \circ \mathcal{E}(\rho_1)\rangle \,|\, \mathcal{E}' \in \mathscr{O}_{\mathsf{A}',\mathsf{A}''}\} \\ &\leq \max\{\langle \sigma, \mathcal{E}''(\rho_1)\rangle \,|\, \mathcal{E}'' \in \mathscr{O}_{\mathsf{A}',\mathsf{A}''}\} \\ &= M_{\mathscr{O},\sigma}(\rho_1). \end{aligned}$$

To prove the implication (iii)⇒(i), suppose that (6.19) holds for all σ. Note that $M_{\mathscr{O},\sigma}(\rho_2) \geq \langle \sigma, \rho_2 \rangle$ holds for all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2})$ since $\rho_2 \xrightarrow{\mathscr{O}} \rho_2$ (since the identity channel $\mathrm{id}_{\mathsf{A}_2}$ is in $\mathscr{O}_{\mathsf{A}_2,\mathsf{A}_2}$). Hence $M_{\mathscr{O},\sigma}(\rho_1) \geq \langle \rho_2, \sigma \rangle$ for all σ by assumption, which implies that $\rho_1 \xrightarrow{\mathscr{O}} \rho_2$. This concludes the proof. □

As we have shown in Section 6.1.3, the monotones $M_{\mathscr{O},\sigma}$ can be computed as solutions to conic programs. If the cone $\mathcal{K} = \mathrm{cone}(J(\mathscr{O}_{\mathsf{A}_1,\mathsf{A}_2}))$ can be defined in terms of semidefinite constraints (that is, if the constraint that dictates when a channel $\mathcal{E}$ is in $\mathscr{O}$ can be written as a semidefinite constraint), then these monotones can furthermore be computed as solutions to semidefinite programs. This will be the case in many of the examples that are studied in the next section.

Note that any monotone $M_{\mathscr{O},\sigma}$ defined this way is constant on all free states in the resource theory. We can compare the resource monotones for different states σ. If σ is a pure free state, i.e., $\sigma = |\psi\rangle\langle\psi|$ and $\sigma \in \mathscr{F}(\mathcal{H})$, then the monotone is constant $M_\sigma(\rho)$ is constant for all states ρ. Indeed, if σ is free then the completely positive trace-preserving map defined by $X \mapsto \mathrm{Tr}(X)\sigma$ must be a free operation. Hence

$$M_{\mathscr{O},\sigma}(\rho) \geq \langle \sigma, \sigma \rangle = 1$$

for all resource states ρ. However, any such monotone must satisfy $M_{\mathscr{O},\sigma}(\rho) \leq 1$ for all ρ, hence $M_{\mathscr{O},\sigma}(\rho) = 1$ for all ρ if σ is pure and free. It remains unclear how to interpret the monotonicity of $M_{\mathscr{O},\sigma}$ for different states σ that are not free.

6.3 Maximal quantum resource theories

We now consider resource theories in which the set of free operations is the maximal set of resource non-generating quantum channels for a given family of convex sets of free states. The structure of the dual cone of witnesses for detecting when a channel is free in these resources has s simple structure. It is based on the cone of witnesses for the set of free states in these resource theories. We then discuss some abstract mathematical examples of these resource theories, such as *affine* resource theories (first introduced in [Gou16]) and *polytopic* resource theories, in which the resource theories have a simple structure and semidefinite programming can be exploited.

It must first be defined what we mean when the free operations of a resource theory are *maximal*. Such a resource theory has the largest possible family of free operations given a family of free states. In particular, the resource theory is completely determined by the structure of its free states, and such a resource theory is constructed as follows. Consider a family $\mathscr{F} = \{\mathscr{F}(\mathcal{H})\}_{\mathcal{H}}$ of subsets of density matrices on finite-dimensional hilbert spaces, such that $\mathscr{F}(\mathcal{H}) \subseteq \mathrm{D}(\mathcal{H})$ for each $\mathcal{H}$ in the family. Define the resource theory that has this family of states as its free resources, and whose family of free operations consists exactly of all of the quantum channels $\mathcal{E} \in \mathrm{C}(\mathcal{H}, \mathcal{H}')$ that are *resource non-generating* with respect to this family of free states. Here, a quantum channel is said to be resource non-generating with respect to a family of free states $\mathscr{F}$ if $\mathcal{E}(\rho) \in \mathscr{F}(\mathcal{H}')$ for all $\rho \in \mathscr{F}(\mathcal{H})$. If $\mathscr{O}(\mathcal{H}, \mathcal{H}')$ is the set of all resource non-generating channels for each set of free states $\mathscr{F}(\mathcal{H})$ and $\mathscr{F}(\mathcal{H}')$, the resulting resource theory $(\mathscr{F}, \mathscr{O})$ is maximal.

In the case when the free states are the bipartite separable states, the corresponding maximal resource consists of the so-called *non-entangling* operations (or *separability preserving* operations) [BP08, BP09, BCY11]. This family of bipartite quantum channels consists of all the operations that do not generate entanglement given a separable input. The family of non-entangling operations is much larger than LOCC, but LOCC is much more difficult to describe mathematically. Studying the larger (but unphysical) resource theory of non-entangling operations still allows us to make important statements about entanglement theory.

We now study resource theories that are both convex whose family of free operations is the maximal set of resource non-generating ones. We first examine the structure of the dual cone of witnesses to the sets of free channels in such resource theories. Let $\mathscr{F}(\mathcal{H})$ and $\mathscr{F}(\mathcal{H}')$ be closed convex sets of density operators on hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, and consider $\mathscr{O}(\mathcal{H}, \mathcal{H})$ to be the maximal set of resource non-generating channels with respect to these sets of free states. Then a channel $\mathcal{E} : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H})$ in this resource theory is free if and only if $\mathcal{E}(\rho) \in \mathscr{F}(\mathcal{H}')$ for all $\rho \in \mathcal{H}$. The dual witness cone for the set of free states on $\mathcal{H}'$ is

$$\mathscr{F}(\mathcal{H}')^* = \{W \in \mathrm{L}(\mathcal{H}) \mid \langle W, \sigma \rangle \geq 0 \ \forall \sigma \in \mathscr{F}(\mathcal{H}')\}.$$

Since $\mathscr{F}(\mathcal{H}')$ is convex and closed, for a given $\sigma \in \mathrm{D}(\mathcal{H}')$ it holds that $\sigma \in \mathscr{F}(\mathcal{H}')$ if and only if $\langle W, \sigma\rangle \geq 0$ for all $W \in \mathscr{F}(\mathcal{H}')^*$. Then $\mathcal{E}$ is a free channel if and only if

$$\begin{aligned} 0 &\leq \langle W, \mathcal{E}(\rho)\rangle \\ &= \langle W \otimes \rho^T, J(\mathcal{E})\rangle \end{aligned}$$

for all $W \in \mathscr{F}(\mathcal{H}')^*$ and all $\rho \in \mathscr{F}(\mathcal{H})$. Hence the dual witness cone for the maximal set of resource non-generating channels is given by

$$\mathcal{W} = J(\mathscr{O}(\mathcal{H}, \mathcal{H}'))^* = \operatorname{cone}\{W \otimes \rho^T \,|\, W \in \mathscr{F}(\mathcal{H}')^*,\, \rho \in \mathscr{F}(\mathcal{H})\} \tag{6.20}$$

such that a channel $\mathcal{E}$ is free if and only if $\langle W, J(\mathcal{E})\rangle \geq 0$ for all $W \in \mathcal{W}$.

6.3.1 Examples of quantum resource theories with maximal sets of free operations

Affine resource theories

A quantum resource theory $(\mathscr{F}, \mathscr{O})$ is said to be *affine* if any state $\rho \in \mathrm{D}(\mathcal{H})$ that is affine combination of free states in $\mathscr{F}(\mathcal{H})$ is also free. That is, it holds that

$$\operatorname{aff}(\mathscr{F}(\mathcal{H})) \cap \mathrm{D}(\mathcal{H}) = \mathscr{F}(\mathcal{H})$$

for all systems $\mathcal{H}$ in the resource theory. For example, the resource theories of coherence and thermodynamics are affine. Any affine subset is closed and convex, and the set of density operators on $\mathcal{H}$ is closed and convex, so $\mathscr{F}(\mathcal{H})$ must also be closed and convex since it is the intersection of two closed, convex sets. Hence the free states of any affine resource theory are also convex. If we consider the set of free operations to be maximal, we may therefore apply the observations above. Affine resource theories have been studied in [Gou16].

Consider an affine resource theory $(\mathscr{F}, \mathscr{O})$. Define the subspaces of hermitian operators

$\mathcal{V} = \operatorname{span}_{\mathbb{R}}(\mathscr{F}(\mathcal{H}_{\mathsf{A}}))$ and $\mathcal{V}' = \operatorname{span}_{\mathbb{R}}(\mathscr{F}(\mathcal{H}_{\mathsf{A}'}))$. Then a channel $\mathcal{E} : \mathrm{L}(\mathcal{H}_{\mathsf{A}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A}'})$ is "free" in this resource theory if and only if $\mathcal{E}(\mathcal{V}) \subset \mathcal{V}'$. The Choi matrix of a free map must satisfy

$$\langle J(\mathcal{E}), Y \otimes X^T \rangle = 0$$

for all $X \in \mathcal{V}$ and all $Y \in \mathcal{V}'^{\perp}$. Define the set

$$\mathcal{W} = \operatorname{cone}(\{Y \otimes X^T \mid X \in \mathcal{V},\, Y \in \mathcal{V}'^{\perp}\}) \tag{6.21}$$

Note that $\mathcal{W} \subseteq \mathrm{H}(\mathcal{H}_{\mathsf{A}'} \otimes \mathcal{H}_{\mathsf{A}})$ is actually a subspace, and is defined such that $J(\mathcal{E})$ is the Choi matrix of a free map if and only if $\langle J(\mathcal{E}), W \rangle = 0$ for all $W \in \mathcal{W}$. Applying Theorem 6.5 to these resource theories, we have the following result.

Theorem 6.6. *Consider an affine resource theory $(\mathscr{F}, \mathscr{O})$ where the sets of free states are affine and the family of free operations is maximal. Let $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{A}})$ and $\rho' \in \mathrm{D}(\mathcal{H}_{\mathsf{A}'})$ be density operators in this resource theory. The following are equivalent.*

(i) There exists a affine resource non-generating channel $\mathcal{E}$ such that $\mathcal{E}(\rho) = \rho'$.

(ii) For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}'})$, all $W \in \mathcal{W}$ defined in (6.21)*, and all $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}})$, if*

$$\mathbb{1}_{\mathsf{A}'} \otimes Y \geq \sigma \otimes \rho^T + W$$

then $\mathrm{Tr}(Y) \geq \langle \sigma, \rho' \rangle$.

(iii) For all density operators $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}'})$, it holds that

$$M_{\mathscr{O},\sigma}(\rho) \geq M_{\mathscr{O},\sigma}(\rho')$$

where $M_{\mathscr{O},\sigma}$ is defined by

$$M_{\mathscr{O},\sigma}(\rho) = \inf_{\substack{W \in \mathcal{W} \\ Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}})}} \left\{ \mathrm{Tr}(Y) \,\middle|\, \mathbb{1}_{\mathsf{A}'} \otimes Y - W \geq \sigma \otimes \rho^T \right\}. \tag{6.22}$$

Since $\mathcal{W}$ is a subspace of $\mathrm{H}(\mathcal{H}_{\mathsf{A}'} \otimes \mathcal{H}_{\mathsf{A}})$, evaluation of the optimization in (6.22) can be obtained as the solution to a semidefinite program. Indeed, given bases $\{V_1, \ldots, V_n\}$ and $\{V_1', \ldots, V_m'\}$ of $\mathcal{V}$ and $\mathcal{V}'^{\perp}$ respectively, any $W \in \mathcal{W}$ can be written as $W = \sum_{j,k} w_{j,k} V_j' \otimes V_k^T$ for some real numbers $w_{j,k} \in \mathbb{R}$.

Polytopic resource theories

Here we consider quantum resource theories in which the set of free states $\mathscr{F}(\mathcal{H}) \subseteq \mathrm{D}(\mathcal{H})$ is always a polytope. A polytope is the convex hull of finitely many points

$$\mathscr{F}(\mathcal{H}) = \mathrm{conv}\{P_1, \ldots, P_n\}.$$

with $P_1, \ldots, P_n \in \mathrm{D}(\mathcal{H})$. Alternatively, this polytope can be described as the intersection of a finite collection of half spaces

$$\mathscr{F}(\mathcal{H}) = \{\rho \,|\, \rho \geq 0,\ \mathrm{Tr}(\rho) = 1,\ \langle \rho, Q_j \rangle \geq 0 \text{ for all } j = 1, \ldots, N\}$$

for some finite collection of operators $\{Q_j\} \subseteq \mathrm{H}(\mathcal{H})$ which defines the faces of the polytope. In such a resource theory, we can consider the set of free operations to be the maximal set of quantum channels that bring free states to free states. That is, a channel $\mathcal{E}$ is free if $\mathcal{E}(\mathscr{F}(\mathcal{H}')) \subseteq \mathscr{F}(\mathcal{H})$. Since the set of free states is a polytope, a operation is free if and only if it maps all of the extreme points of the polytope into the polytope. These "polytope-preserving" maps are exactly the resource non-generating operations in this resource theory.

That is, a channel $\mathcal{E}$ is free if and only if

$$\langle \mathcal{E}(P_k), Q_j \rangle \geq 0$$

for all j and k, where $\{P_k\}$ are the vertices of the polytope and $\{Q_j\}$ are operators that define the faces of the polytope. This is equivalent to the fact that the Choi matrix $J(\mathcal{E})$ must satisfy

$$\langle J(\mathcal{E}), Q_j \otimes P_k^T \rangle \geq 0$$

for all j and k. Consider the closed convex cone

$$\begin{aligned} \mathcal{W} &= \operatorname{cone}(\{Q_j \otimes P_k^T \mid j, k\}) \\ &= \Big\{ \sum_{j,k} p_{jk} Q_j \otimes P_k^T \,\Big|\, p_{jk} \geq 0 \Big\}. \end{aligned} \tag{6.23}$$

Then $\mathcal{W}$ is the dual cone of witnesses to the set of free channels that preserve the polytope. That is, $\mathcal{E}$ is free if and only if $\langle W, J(\mathcal{E}) \rangle \geq 0$ holds for all $W \in \mathcal{W}$ (i.e., if $J(\mathcal{E}) \in \mathcal{W}^*$). We can now apply Theorem 6.5 to the polytopic resource theories to see that the functions $M_{\mathscr{O},\sigma}$ defined by the semidefinite program

$$\begin{aligned} M_{\mathscr{O},\sigma}(\rho) = \text{minimize: } & \operatorname{Tr}(Y) \\ \text{subject to: } & p_{jk} \geq 0 \\ & \mathbb{1}_{\mathsf{A}_2} \otimes Y - \sum_{j,k} p_{jk} Q_j \otimes P_k^T \geq \sigma \otimes \rho^T \end{aligned}$$

form a complete set of resource monotones.

Theorem 6.7. *Consider a polytopic resource theory as defined above, where the set of free states is a polytope and the set of free operations consists all polytope-preserving channel. Let $\rho, \rho' \in \mathrm{D}(\mathcal{H})$ be density operators, and let $\mathcal{K}$ be the cone defined in* (6.23)*. The following are equivalent.*

(i) There exists a polytope-preserving channel $\mathcal{E}$ such that $\mathcal{E}(\rho) = \rho'$.

(ii) For all $\sigma \in \mathrm{D}(\mathcal{H})$, all $p_{jk} \geq 0$, and all $Y \in \mathrm{H}(\mathcal{H})$, if

$$\mathbb{1} \otimes Y \geq \sigma \otimes \rho^T + \sum_{j,k} p_{jk} Q_j \otimes P_k^T$$

then $\mathrm{Tr}(Y) \geq \langle \sigma, \rho' \rangle$.

(iii) For all density operators $\sigma \in \mathrm{D}(\mathcal{H})$, it holds that $M_{\mathscr{O},\sigma}(\rho) \geq M_{\mathscr{O},\sigma}(\rho')$.

The resource monotones $M_{\mathscr{O},\sigma}(\rho)$ in this resource theory can be computed as solutions to semidefinite programs. As an example of a polytopic resource theory that appears in quantum information theory is the resource theory of *magic states* (or *non-stabilizer states*) [DAGS17], which has useful applications in quantum computation. The free resources in this resource theory are the stabilizer states.

6.4 Applications to entanglement theory

We now consider resource theories of states and operations on bipartite systems, which will give interesting results relating to entanglement theory. In particular, the results from the study of general convex resource theories discussed earlier in this chapter will be used to study the resource theories of separable operations and of PPT operations.

Given two bipartite systems $\mathcal{H}_{\mathsf{A}_1} \otimes \mathcal{H}_{\mathsf{B}_1}$ and $\mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{B}_2}$, we use the shorthand $\mathsf{A} = (\mathsf{A}_1\mathsf{A}_2)$ and $\mathsf{B} = (\mathsf{B}_1\mathsf{B}_2)$ to denote the systems corresponding to Alice and Bob separately, and we write

$$\begin{aligned} \mathcal{H}_{\mathsf{A}_2} \otimes \mathcal{H}_{\mathsf{B}_2} \otimes \mathcal{H}_{\mathsf{A}_1} \otimes \mathcal{H}_{\mathsf{B}_1} &= \mathcal{H}_{\mathsf{A}_2\mathsf{B}_2} \otimes \mathcal{H}_{\mathsf{A}_1\mathsf{B}_1} \\ &= \mathcal{H}_{\mathsf{AB}} \end{aligned}$$

for simplicity. The set of bipartite channels $\mathrm{L}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$ is denoted $\mathrm{C}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1}, \mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$.

Given a quantum channel $\mathcal{E}$ on these systems, its Choi representation is a operator on the space $J(\mathcal{E}) \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})$.

The framework of convex resource theories applied to separable operations is presented in Section 6.4.1. It will be shown that any finite approximation to the set of separable operations necessarily yields a resource theory with no bound entanglement. In particular, there will be two-qubit entangled states that are free in such a resource theory. The study of PPT states and operations as a resource theory is presented in Section 6.4.1. Convertibility of resources in this resource theory, and the computation of the complete set of resource monotones, can be cast as semidefinite programs in this case.

6.4.1 Resource theory of separable operations

Recall that a bipartite quantum channel $\mathcal{E} : \mathrm{L}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$ is said to be *separable* if its Choi matrix $J(\mathcal{H}) \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})$ is separable with respect to the $\mathsf{A} : \mathsf{B}$ bipartition, i.e., if $J(\mathcal{E}) \in \mathrm{Sep}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$. The *resource theory of separable operations* is the resource theory of bipartite hilbert spaces in which the free states and free channels are the separable ones. Separable operations were first studied in [Rai97, VP98], and the set of separable channels was shown to be strictly larger than the set of LOCC channels in [BDF$^+$99]. Transformations between pure states using separable operations has been studied in [GG07, GG08, Ghe10], while asymptotic distillation of entanglement using separable operations was studied in [Rai97]. The relationships between separable operations and LOCC have been studied in [SRBL17].

We now use the framework for studying convex resource theories to study the resource theory of separable operations. Let $\mathcal{W}_{\mathsf{A:B}} = (\mathrm{Sep}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}}))^*$ be the dual cone of entanglement witnesses for the cone of separable operators on this space, such that $\langle W, X \rangle \geq 0$ for all witnesses $W \in \mathcal{W}_{\mathsf{A:B}}$ and all separable positive operators $X \in \mathrm{Sep}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$. Then a channel $\mathcal{E} \in \mathrm{C}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1}, \mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$ is separable if and only if $\langle W, J(\mathcal{E}) \rangle \geq 0$ holds for all elements of the cone $W \in \mathcal{W}_{\mathsf{A:B}}$. We may now apply Theorem 6.5 to the resource theory of separable operations.

Theorem 6.8. *Let $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1})$ and $\rho_2 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$ be bipartite density operators. The following are equivalent.*

1. *There exists a separable channel $\mathcal{E}$ such that $\mathcal{E}(\rho_1) = \rho_2$.*
2. *For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$, all entanglement witnesses $W \in \mathcal{W}_{\mathsf{A:B}}$, and all $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1})$, if*

   $$\mathbb{1}_{\mathsf{A}_2\mathsf{B}_2} \otimes Y \geq \sigma \otimes \rho_1^T + W$$

   then $\mathrm{Tr}(Y) \geq \langle \sigma, \rho_2 \rangle$.
3. *For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$, it holds that $M_{\mathrm{Sep},\sigma}(\rho_1) \geq M_{\mathrm{Sep},\sigma}(\rho_2)$,*

where $M_{\mathrm{Sep},\sigma}(\rho)$ is the separable resource monotone defined by

$$M_{\mathrm{Sep},\sigma}(\rho) = \inf_{\substack{W \in \mathcal{W}_{\mathsf{A:B}} \\ Y \in \mathrm{H}(\mathcal{H}_{A_1B_1})}} \{\mathrm{Tr}\, Y \mid \mathbb{1}_{\mathsf{A}_2\mathsf{B}_2} \otimes Y - W \geq \sigma \otimes \rho_1^T\}.$$

The functions $M_{\mathrm{Sep},\sigma}(\rho)$ form a complete set of resource monotones for the resource theory of separable operations, and they can be computed as the solution to a conic program. These monotones are not necessarily useful in practice, however, since the it cannot be cast as a semidefinite program. Indeed, the cone of entanglement witnesses $\mathcal{W}_{\mathsf{A:B}}$ is difficult to characterize computationally.

Finite approximations to separable operators

We can instead try to simplify this resource theory mathematically by considering a polytope approximation to the set of separable operations. To do this, consider only a finite collection of entanglement witnesses $\{W_i\}_{i=1}^n$ where each $W_i \in \mathcal{W}_{\mathsf{A:B}}$ is an entanglement witnesses, and define

$$\widetilde{\mathcal{W}}_{\mathsf{A:B}} = \mathrm{cone}(\{W_i\}_{i=1}^n)$$

as the conic closure of these witnesses. We see that $\widetilde{\mathcal{W}}_{\mathsf{A:B}} \subset \mathcal{W}_{\mathsf{A:B}}$ and this containment is strict. Indeed, a finite collection of witnesses cannot be used to detect all entangled operators, since that would contradict the hardness of the separability problem [Gur03]. Hence the dual cone to $\widetilde{\mathcal{W}}_{\mathsf{A:B}}$ is strictly larger than the cone of separable operators $\mathrm{Sep}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$. Define $\mathcal{K} = (\widetilde{\mathcal{W}}_{\mathsf{A:B}})^*$ as this dual cone, then $\mathcal{K} \subset \mathrm{Sep}(\mathcal{H}_{\mathsf{A}} : \mathcal{H}_{\mathsf{B}})$ and this cone is also a polytope. We can now consider a resource theory whose free operations are those that are contained in $\mathcal{K}$. It will be shown that any such resource theory must have some entangled states as free resources. In particular, there cannot exist bound entanglement in such a resource theory.

Using Theorem 6.1, we can give necessary and sufficient conditions for conversion of resources within this resource theory. Let $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1})$ and $\rho_2 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$ be bipartite density operators. Then there exists a channel $\mathcal{E} \in \mathcal{K} = (\widetilde{\mathcal{W}}_{\mathsf{A:B}})^*$ such that $\mathcal{E}(\rho_1) = \rho_2$ if and only if, for all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_3})$, all $W \in \widetilde{\mathcal{W}}_{\mathsf{A:B}}$, and all $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_2})$,

$$\mathbb{1}_{\mathsf{A}_2\mathsf{B}_2} \otimes X \geq \sigma \otimes \rho_1^T + W \tag{6.24}$$

implies that $\mathrm{Tr}(Y) \geq \langle \sigma, \rho_2 \rangle$.

Next note the following fact about entanglement witnesses on $\mathcal{H}_{\mathsf{AB}} = \mathcal{H}_{\mathsf{A}_2\mathsf{B}_2\mathsf{A}_1\mathsf{B}_1}$. Given any $W \in \mathcal{W}_{\mathsf{A:B}}$, the partial trace of W over $\mathsf{A}_1\mathsf{B}_1$, which we write as

$$W^{(\mathsf{A}_2\mathsf{B}_2)} = \mathrm{Tr}_{\mathsf{A}_1\mathsf{B}_1}(W),$$

is an entanglement witness for operators on $\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2}$. Indeed, let $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$ be a separable density matrix. Then

$$\begin{aligned} \langle W^{(\mathsf{A}_2\mathsf{B}_2)}, \rho \rangle &= \langle W, \rho \otimes \mathbb{1}_{\mathsf{A}_1\mathsf{B}_1} \rangle \\ &\geq 0 \end{aligned}$$

since $\rho \otimes \mathbb{1}_{\mathsf{A}_1\mathsf{B}_1} \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})$ is separable with respect to $\mathsf{A} : \mathsf{B}$. In fact, the operator $\rho \otimes \mathbb{1}_{\mathsf{A}_1\mathsf{B}_1}$ is exactly the Choi representation of the constant channel that discards the input system and produces ρ.

Consider now the entanglement witnesses $\{W_i^{(\mathsf{A_2B_2})}\}_{i=1}^n$, where the W_i are the witnesses given earlier. By the above observation, each $W_i^{(\mathsf{A_2B_2})}$ is an entanglement witness for states on $\mathcal{H}_{\mathsf{A_2B_2}}$. Since $\{W_i^{(\mathsf{A_2B_2})}\}_{i=1}^n$ is only a finite collection of entanglement witnesses, there must exist a two-qubit entangled state ρ_2 that is *not* detected by these witnesses. That is, there exists a ρ_2 that is entangled but that $\langle W_i^{(\mathsf{A_2B_2})}, \rho_2\rangle \geq 0$ for each of these witnesses.

Let $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A_1B_1}})$ be *any* density matrix. We will show that there exists a free channel in this resource theory such that $\mathcal{E}(\rho_1) = \rho_2$. Indeed, let $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A_2B_2}})$, $W \in \widetilde{\mathcal{W}}_{\mathsf{A:B}}$, and $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A_1B_1}})$ be arbitrary such that $\operatorname{Tr} Y < \langle \sigma, \rho_2\rangle$. We will show that (6.24) does not hold. Note that $\rho_2 \otimes \mathbb{1}_{\mathsf{A_1B_1}} \geq 0$, but

$$\langle \mathbb{1}_{\mathsf{A_2B_2}} \otimes Y - \sigma \otimes \rho_1^T - W, \rho_2 \otimes \mathbb{1}_{\mathsf{A_1B_1}}\rangle = \underbrace{\mathrm{Tr}(Y) - \langle \sigma, \rho_2\rangle}_{<0} - \underbrace{\langle W^{(\mathsf{A_2B_2})}, \rho_2\rangle}_{\geq 0}$$
$$< 0.$$

So it must be the case that $\mathbb{1}_{\mathsf{A_2B_2}} \otimes X - \sigma \otimes \rho_1^T - W \ngeq 0$. This completes the proof that there exists a free channel in this resource theory such that $\mathcal{E}(\rho_1) = \rho_2$. Since ρ_1 was arbitrary, yet $\rho_1 \xrightarrow{\tilde{\mathcal{W}}^*} \rho_2$, it follows that ρ_2 must be free in this resource theory (since ρ_2 can be obtained by every other resource state in the resource theory).

In particular, this implies that any polytope approximation to the set of separable channels must contain an entangling channel that maps a separable state to a two-qubit entangled state. While only the single-shot case is considered here, this also has important implications to the asymptotic case for this resource theory, in which the free operations are defined as some polytope approximation to the separable channels. Every two-qubit entangled state is distillable using separable operations, so any such state must also be distillable using the operations in the cone $\mathcal{K}$ (the polytope approximation to the set of separable operations used here). Hence, in this resource theory, arbitrary amounts of entanglement could be distilled out of any state (including any separable state). This is remarkable, since any finite approximation to the set of separable channels leads to a resource theory in which arbitrary amounts of entanglement are free.

Separable states are always obtainable via separable operations

We now show that transformations from any state ρ_1 to any separable state ρ_2 is always possible in SEP. Indeed, we will show that if $\rho_1 \to \rho_2$ is not possible via separable operations, then it must be that ρ_2 is entangled. This fact is already well known, and it is something that we already implicit assumed in order to consider the separable operations as a resource theory. However, it is important to prove this here using the formalism for convex resource theories to show the effectiveness of this formalism.

Let $\rho_1 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_1 B_1})$ and $\rho_2 \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$ be any arbitrary bipartite density operators and suppose $\rho_1 \to \rho_2$ is *not* possible by separable operations. Then there must exist a density operator $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$, an entanglement witness $W \in \mathcal{W}_{\mathsf{A:B}} \subset \mathrm{H}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2\mathsf{A}_1\mathsf{B}_1})$, and some $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1})$ such that

$$\mathbb{1}_{\mathsf{A}_2\mathsf{B}_2} \otimes Y - \sigma \otimes \rho_1^T - W \geq 0 \tag{6.25}$$

but $\mathrm{Tr}(Y) < \langle \sigma, \rho_2 \rangle$. Since the operator in (6.25) is positive definite and the operator $\rho_2 \otimes \mathbb{1}_{\mathsf{A}_1\mathsf{B}_1} \geq 0$ is nonzero, it follows that

$$\begin{aligned} 0 &\leq \langle \mathbb{1}_{\mathsf{A}_2\mathsf{B}_2} \otimes Y - \sigma \otimes \rho_1^T - W,\ \rho_2 \otimes \mathbb{1}_{\mathsf{A}_1\mathsf{B}_1} \rangle \\ &= \mathrm{Tr}(Y) - \langle \sigma, \rho_2 \rangle - \langle W^{(\mathsf{A}_2\mathsf{B}_2)}, \rho_2 \rangle. \end{aligned}$$

Then $\langle W^{(\mathsf{A}_2\mathsf{B}_2)}, \rho_2 \rangle < 0$ since $\mathrm{Tr}(Y) - \langle \sigma, \rho_2 \rangle < 0$. Hence $W^{(\mathsf{A}_2\mathsf{B}_2)} = \mathrm{Tr}_{\mathsf{A}_1\mathsf{B}_1}(W)$ is an entanglement witness for ρ_2 and thus ρ_2 must be entangled.

6.4.2 Resource theory of PPT operations

Recall that a channel $\mathcal{E} \in \mathrm{C}(\mathcal{H}_{\mathsf{A_1B_1}}, \mathcal{H}_{\mathsf{A_2B_2}})$ is PPT if its Choi representation is PPT with respect to the bipartite splitting $\mathsf{A} : \mathsf{B}$. That is, if its Choi matrix $J(\mathcal{E}) \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})$ satisfies $J(\mathcal{E})^{T_{\mathsf{B}}} \geq 0$, where $T_{\mathsf{B}} = T_{\mathsf{B_1B_2}}$ represents the operation that transposes Bob's systems. Hence a channel $\mathcal{E}$ is PPT if and only if

$$\langle J(\mathcal{E}), X^{T_{\mathsf{B}}} \rangle \geq 0 \qquad \text{and} \qquad \langle J(\mathcal{E}), X \rangle \geq 0$$

hold for all $X \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})$ with $X \geq 0$. So we may take the $\mathcal{W}$ to be the set

$$\mathcal{W} = \{X + Y^{T_{\mathsf{B}}} \,|\, X, Y \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}}),\ X \geq 0 \text{ and } Y \geq 0\}$$

so that $\mathcal{W}$ is the dual cone of witnesses to the PPT channels. There is another useful characterization of PPT channels that is presented in the following lemma.

Lemma 6.9. *Let $\mathcal{E} \in \mathrm{C}(\mathcal{H}_{\mathsf{A_1B_1}}, \mathcal{H}_{\mathsf{A_2B_2}})$ be a bipartite quantum channel. Then $\mathcal{E}$ is a PPT channel if and only if the map $\mathcal{E}^{T_{\mathsf{B}}} : \mathrm{L}(\mathcal{H}_{\mathsf{A_1B_1}}) \to \mathrm{L}(\mathcal{H}_{\mathsf{A_2B_2}})$ is completely positive, where $\mathcal{E}^{T_{\mathsf{B}}}$ is defined by*

$$\mathcal{E}^{T_{\mathsf{B}}}(X) = \mathcal{E}(X^{T_{\mathsf{B_1}}})^{T_{\mathsf{B_2}}}$$

for all $X \in \mathrm{L}(\mathcal{H}_{\mathsf{A_1B_1}})$.

Proof. It is straightforward to show that the Choi matrix of the map $\mathcal{E}^{T_{\mathsf{B}}}$ is given by $J(\mathcal{E}^{T_{\mathsf{B}}}) = J(\mathcal{E})^{T_{\mathsf{B}}}$, and thus $J(\mathcal{E})^{T_{\mathsf{B}}} \geq 0$ if and only if $J(\mathcal{E}^{T_{\mathsf{B}}}) \geq 0$. □

The composition of PPT channels is a PPT channel. Indeed, if $\mathcal{E}$ and $\mathcal{E}'$ are PPT channels, then

$$\begin{aligned}(\mathcal{E}' \circ \mathcal{E})^{T_{\mathsf{B}}} &= T_{\mathsf{B}} \circ \mathcal{E}' \circ \mathcal{E} \circ T_{\mathsf{B}} \\ &= \mathcal{E}'^{T_{\mathsf{B}}} \circ \mathcal{E}'^{T_{\mathsf{B}}},\end{aligned}$$

which is completely positive since the composition of completely positive maps is completely positive. Furthermore, the identity channel is PPT and a PPT channel applied to a PPT state always yields another PPT state. Hence we may consider the resource theory of PPT states and channels, where the free states are PPT states and the free operations are the PPT channels. Furthermore, the sets of PPT states and channels are closed and convex, so we may apply the formalism of convex quantum resource theories to the PPT resource theory.

Theorem 6.10. *Let ρ_1 and ρ_2 be bipartite states. The following are equivalent.*

(i) There exists a PPT channel $\mathcal{E}$ such that $\mathcal{E}(\rho_1) = \rho_2$.

(ii) For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$, all $Z \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})$ with $Z \geq 0$, and all $Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1})$, if

$$\mathbb{1}_{\mathsf{A}_2\mathsf{B}_2} \otimes X \geq \sigma \otimes \rho^T + Z^{T_\mathsf{B}}$$

then $\mathrm{Tr}(Y) \geq \langle \sigma, \rho_2 \rangle$.

(iii) For all $\sigma \in \mathrm{D}(\mathcal{H}_{\mathsf{A}_2\mathsf{B}_2})$, it holds that

$$M_{\mathrm{PPT},\sigma}(\rho_1) \geq M_{\mathrm{PPT},\sigma}(\rho_2)$$

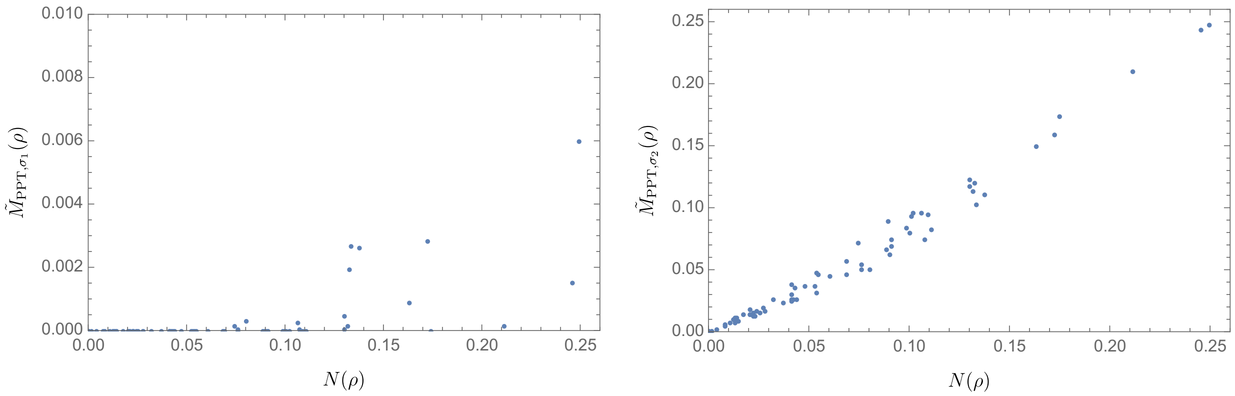
where $M_{\mathrm{PPT},\sigma}$ can be computed from the following semidefinite program:

$$\begin{aligned} \text{minimize: } & \mathrm{Tr}(Y) \\ \text{subject to: } & Y \in \mathrm{H}(\mathcal{H}_{\mathsf{A}_1\mathsf{B}_1}),\ Z \in \mathrm{H}(\mathcal{H}_{\mathsf{AB}})_+ \\ & \mathbb{1}_{\mathsf{A}_2\mathsf{B}_2} \otimes Y - Z^{T_\mathsf{B}} \geq \sigma \otimes \rho^T \end{aligned}$$

In particular, we see that we can characterize conversion of bipartite states in the resource theory of PPT operations by means of a complete set of SDP-computable monotones. The following MATLAB code can be used to compute the value of the PPT monotone $M_{\text{PPT},\sigma}(\rho)$ using cvx for semidefinite programming. In the code, we use dA1, dB1, dA2, and dB2 to denote the dimensions of $\mathcal{H}_{\mathsf{A}_1}$, $\mathcal{H}_{\mathsf{B}_1}$, $\mathcal{H}_{\mathsf{A}_2}$, and $\mathcal{H}_{\mathsf{B}_2}$.

```
d1 = dA1*dB1;
d2 = dA2*dB2;
cvx_begin sdp
   variable Z(d1*d2,d1*d2) hermitian
   variable Y(d1,d1) hermitian
   minimize trace(Y)
   subject to
      Z >=0;
      kron(eye(d2),Y) - PartialTranspose(Z,[2,4],[dA2,dB2,dA1,dB1])
                >= kron(sigma,transpose(rho));
cvx_end
```

Note that the monotones $M_{\text{PPT},\sigma}$ must necessarily be constant on all PPT states. To make the monotones also *faithful* (i.e., so that $M_{\text{PPT},\sigma}(\rho) = 0$ for all "free" PPT states ρ), we can define $\tilde{M}_{\text{PPT},\sigma}(\rho) := M_{\text{PPT}\sigma}(\rho) - M_{\text{PPT}\sigma}(\mathbb{1}_{\mathsf{A}_1\mathsf{B}_1}/d_1)$ such that $\tilde{M}_{\text{PPT},\sigma}(\rho) = 0$ for all PPT states. Examples of the results from computing monotones of this form are displayed in Figure 6.1. For two different two-qubit states (one randomly generated entangled state σ_1 and the maximally entangled state $\sigma_2 = |\phi^+\rangle\langle\phi^+|$), the monotones $\tilde{M}_{\text{PPT},\sigma_1}$ and $\tilde{M}_{\text{PPT},\sigma_2}$ have been computed for 100 randomly generated two-qubit states ρ. The negativity, another monotone under PPT operations that is well known, is compared to both monotones. The behaviors and physical interpretations of these monotones are not yet understood, but it is clear that they are distinctly different from the negativity and from each other. All computations were performed using MATLAB, and evaluations of the monotones $\tilde{M}_{\text{PPT},\sigma}$ were performed using the code displayed above.



(a) Comparison of the monotone $\tilde{M}_{\mathrm{PPT},\sigma_1}(\rho)$ to the negativity for randomly generated two-qubit states, where σ_1 is a fixed random two-qubit state.

(b) Comparison of the monotone $\tilde{M}_{\mathrm{PPT},\sigma_2}(\rho)$ to the negativity for randomly generated two-qubit states, where $\sigma_2 = |\phi^+\rangle\langle\phi^+|$ is the maximally entangled state of two qubits.

Figure 6.1: Comparison of different PPT monotones to the negativity for 100 randomly generated states of two qubits.

6.5 Conclusion

Resource theories of quantum states are frameworks within the formalism of quantum information theory in which the allowed transformations that may be performed on the states of systems interest are restricted. Such a restriction induces a hierarchy among states where more "resourceful" states can be mapped to less resourceful ones via the allowed transformations. For example, in the well-studied theory of entanglement, the set of allowable transformations of bipartite states is restricted to only the ones that can be implemented with LOCC. Significant progress has been made toward the construction and classification of quantum resource theories in general [HO13, CFS16], but most work in this area concerns quantum resource theories in the asymptotic regime, where rates of conversion are considered in the limit of a large number of identical systems [GS08, BH15]. In practical applications where the resources available are only finite, different tools must be used.

This chapter presented a framework for conversion among resources in convex resource theories in the single-shot regime in quantum information. The question of convertibility of states can, in many resource theories, be cast in terms of a feasibility semidefinite program

[Gou16], or more generally as a conic program. This representation of the question of convertibility makes use of the Choi matrix representation of a quantum operation. While this formulation is useful in its own right, making use of duality relations for conic programs allowed us to find tractable necessary and sufficient conditions for convertibility of resources within certain quantum resource theories. These conditions can be given as a complete family of resource monotones, and these monotones can be computed in practice by means of a semidefinite program in cases where the resource theory has a simple mathematical structure. In particular, we investigated a complete family of monotones for affine and polytopic resource theories, as well as for the resource theory of PPT operations, where the semidefinite program falls out of the mathematical structure of the theory.

So far, this method of using duality properties of conic programs has been used to characterize when exact transformations are possible within a resource theory. In physical implementations of quantum information, however, exact copies of desired resources are not necessary, and only an approximation of the target state is necessary. It may be possible to explore convertibility in resource theories for the case when only approximate transformations are desired. That is, using this formalism of conic programs one may be able to find necessary and sufficient conditions that determine when a transformation exists that yields the desired outcome state up to some approximation. This question can also be cast in terms a semidefinite program in some cases. This might also be able to produce new monotones that can be used to test for convertibility up to some approximation.

Some of the results in this chapter yield interesting implications regarding the resource theory of entanglement. In particular, we showed that any finite approximation to the set of separable operations leads to a resource theory in which arbitrary amounts of entanglement can be extracted asymptotically from any state. This highlights the fine structure of entanglement, since the resource theory of separable operations alone leads to bound entanglement. It may also be possible to apply this formalism of convex resource theories to other approximations to the set of channels that can be implemented by LOCC. We may consider, for example, the set k-extendible maps [PBHS11], defined as the set of channels whose Choi representation is k-extendible. The class of k-extendible maps are directly related to the criteria of entanglement based on k-extendability of states, an important method in the detection of entanglement [DPS05, BCY11, NOP09]. This class of maps can be described easily mathematically, and the question of convertibility can be cast as a semidefinite program.

Bibliography

[ADVW02] Koenraad Audenaert, Bart De Moor, Karl Gerd H. Vollbrecht, and Reinhard F. Werner. Asymptotic relative entropy of entanglement for orthogonally invariant states. *Phys. Rev. A*, 66(3):032310, sep 2002.

[BB84] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proc. IEEE Int. Conf. Comput. Syst. Signal Process.*, pages 175–179, Bangalore, India, 1984.

[BBC+93] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70(13):1895–1899, mar 1993.

[BCP14] T. Baumgratz, M. Cramer, and M. B. Plenio. Quantifying coherence. *Phys. Rev. Lett.*, 113(14):1–5, 2014.

[BCY11] Fernando G S L Brandão, Matthias Christandl, and Jon Yard. Faithful Squashed Entanglement. *Commun. Math. Phys.*, 306(3):805–830, sep 2011.

[BDF+99] Charles H. Bennett, David P. DiVincenzo, Christopher A Fuchs, Tal Mor, Eric Rains, Peter W. Shor, John A. Smolin, and William K Wootters. Quantum nonlocality without entanglement. *Phys. Rev. A*, 59(2):1070–1091, feb 1999.

[BDM$^+$99] Charles H. Bennett, David DiVincenzo, Tal Mor, Peter W. Shor, John Smolin, and Barbara M. Terhal. Unextendible Product Bases and Bound Entanglement. *Phys. Rev. Lett.*, 82(26):5385–5388, jun 1999.

[BDSW96] Charles H. Bennett, David DiVincenzo, John Smolin, and William Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54(5):3824–3851, nov 1996.

[Bei13] Salman Beigi. Sandwiched ReÌĄnyi divergence satisfies data processing inequality. *J. Math. Phys.*, 54(12):122202, jun 2013.

[BESP03] Daniel E. Browne, Jens Eisert, Stefan Scheel, and Martin B. Plenio. Driving non-Gaussian to Gaussian states with linear optics. *Phys. Rev. A*, 67(6):062320, 2003.

[BG15] Fernando G. S. L. Brandão and Gilad Gour. Reversible Framework for Quantum Resource Theories. *Phys. Rev. Lett.*, 115(7):070503, aug 2015.

[BH15] Fernando G. S. L. Brandão and Aram W Harrow. Estimating operator norms using covering nets. sep 2015.

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, 1997.

[BHO$^+$13] Fernando G. S. L. Brandão, Michał Horodecki, Jonathan Oppenheim, Joseph M. Renes, and Robert W. Spekkens. Resource Theory of Quantum States Out of Thermal Equilibrium. *Phys. Rev. Lett.*, 111(25):250404, dec 2013.

[BL06] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples.* Springer, 2006.

[BL13] Orest Bucicovschi and Jiri Lebl. On the continuity and regularity of convex extensions. *J. Convex Anal.*, 20(4):1113–1126, dec 2013.

[BP08] Fernando G S L Brandão and Martin B Plenio. Entanglement theory and the second law of thermodynamics. *Nat. Phys.*, 4(11):873–877, nov 2008.

[BP09] Fernando G S L Brandão and Martin B Plenio. Entanglement manipulation under non-entangling operations. *J. Phys. Conf. Ser.*, 143:012009, dec 2009.

[BP10] Fernando G. S. L. Brandão and Martin B. Plenio. A Generalization of Quantum Stein's Lemma. *Commun. Math. Phys.*, 295(3):791–828, feb 2010.

[BS03] Dominic W Berry and Barry C Sanders. Bounds on general entropy measures. *J. Phys. A. Math. Gen.*, 36(49):12255–12265, dec 2003.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[BW92] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69(20):2881–2884, nov 1992.

[BŻ06] I Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge University Press, 2006.

[Car10] Eric A. Carlen. Trace inequalities and quantum entropy: an introductory course. In Roert Sims and Daniel Ueltschi, editors, *Entropy quantum Arizona Sch. Anal. with Appl.*, pages 73–140. American Mathematical Society, 2010.

[CFS16] Bob Coecke, Tobias Fritz, and Robert W. Spekkens. A mathematical theory of resources. *Inf. Comput.*, 250:59–86, oct 2016.

[CG16] Eric Chitambar and Gilad Gour. Are Incoherent Operations Physically Consistent? – A Critical Examination of Incoherent Operations. pages 1–25, 2016.

[Cho75] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra Appl.*, 10(3):285–290, jun 1975.

[Cla06] Lieven Clarisse. The distillability problem revisited. *Quantum Inf. Comput.*, 6(6):539–560, oct 2006.

[CLM$^+$14] Eric Chitambar, Debbie Leung, Laura Mančinska, Maris Ozols, and Andreas Winter. Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask). *Commun. Math. Phys.*, 328(1):303–326, may 2014.

[DAGS17] Hoan Bui Dang, Mehdi Ahmadi, Gilad Gour, and Barry C. Sanders. Quantum resource theory of non-stabilizer states in the single-shot regime. 2017.

[DJ14] N. Dinh and V. Jeyakumar. Farkas' lemma: three decades of generalizations for mathematical optimization. *TOP*, 22(1):1–22, apr 2014.

[DL14] Nilanjana Datta and Felix Leditzky. A limit of the quantum Rényi divergence. *J. Phys. A Math. Theor.*, 47(4):045304, jan 2014.

[DMHB13] Nilanjana Datta, Milan Mosonyi, Min Hsiu Hsieh, and Fernando G S L Brandao. A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels. *IEEE Trans. Inf. Theory*, 59(12):8014–8026, 2013.

[DPS05] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Detecting multipartite entanglement. *Phys. Rev. A*, 71(3):032333, mar 2005.

[DSS$^+$00] David P. DiVincenzo, Peter W. Shor, John A. Smolin, Barbara M. Terhal, and Ashish V. Thapliyal. Evidence for bound entangled states with negative partial transpose. *Phys. Rev. A*, 61(6):062312, may 2000.

[ES13] Christopher Eltschka and Jens Siewert. Negativity as an Estimator of Entanglement Dimension. *Phys. Rev. Lett.*, 111(10):100503, sep 2013.

[ES14] Christopher Eltschka and Jens Siewert. Quantifying entanglement resources. *J. Phys. A Math. Theor.*, 47(42):424005, oct 2014.

[ETS15] Christopher Eltschka, Géza Tóth, and Jens Siewert. Partial transposition as a direct link between concurrence and negativity. *Phys. Rev. A*, 91(3):032327, mar 2015.

[Far02] Julius Farkas. Theorie der einfachen Ungleichungen. *J. für die reine und Angew. Math.*, (124):1–27, 1902.

[Fey82] Richard P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21(6-7):467–488, jun 1982.

[FGR11] Shmuel Friedland, Gilad Gour, and Aidan Roy. Local extrema of entropy functions under tensor products. *Quantum Inf. Comput.*, 11(11):1028–1044, may 2011.

[FL13] Rupert L. Frank and Elliott H. Lieb. Monotonicity of a relative ReÌĄnyi entropy. *J. Math. Phys.*, 54(12):122201, jun 2013.

[Fri13] Tobias Fritz. Resources, may 2013.

[GB08] Michael C. Grant and Stephen P. Boyd. Graph Implementations for Nonsmooth Convex Programs. In *Recent Adv. Learn. Control*, volume 371, pages 95–110. Springer London, London, 2008.

[GF13] Gilad Gour and Shmuel Friedland. The Minimum Entropy Output of a Quantum Channel Is Locally Additive. *IEEE Trans. Inf. Theory*, 59(1):603–614, jan 2013.

[GG07] Vlad Gheorghiu and Robert B. Griffiths. Entanglement transformations using separable operations. *Phys. Rev. A*, 76(3):032310, sep 2007.

[GG08] Vlad Gheorghiu and Robert Griffiths. Separable operations on pure states. *Phys. Rev. A*, 78(2):020304, aug 2008.

[GG15] Mark W. Girard and Gilad Gour. Computable entanglement conversion witness that is better than the negativity. *New J. Phys.*, 17(9):093013, sep 2015.

[GG17] Mark W. Girard and Gilad Gour. Entanglement monotones and transformations of symmetric bipartite states. *Phys. Rev. A*, 95(1):012308, jan 2017.

[GGF14] Mark W. Girard, Gilad Gour, and Shmuel Friedland. On convex optimization problems in quantum information theory. *J. Phys. A Math. Theor.*, 47(50):505302, dec 2014.

[Ghe10] Vlad Gheorghiu. *Separable Operations, Graph Codes and the Location of Quantum Information.* PhD thesis, 2010.

[GM12] Bernd Gärtner and Jiri Matousek. *Approximation Algorithms and Semidefinite Programming.* Springer-Verlag, 2012.

[GMN$^+$15a] Gilad Gour, Markus P. Müller, Varun Narasimhachar, Robert W. Spekkens, and Nicole Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Phys. Rep.*, 583:1–58, jul 2015.

[GMN$^+$15b] Gilad Gour, Markus P. Müller, Varun Narasimhachar, Robert W. Spekkens, and Nicole Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Phys. Rep.*, 583:1–58, 2015.

[Gou05a] Gilad Gour. Family of concurrence monotones and its applications. *Phys. Rev. A*, 71(1):012318, jan 2005.

[Gou05b] Gilad Gour. Infinite number of conditions for local mixed-state manipulations. *Phys. Rev. A*, 72(2):022323, aug 2005.

[Gou16] Gilad Gour. Quantum resource theories in the single-shot regime. oct 2016.

[GS08] Gilad Gour and Robert W. Spekkens. The resource theory of quantum reference frames: Manipulations and monotones. *New J. Phys.*, 10, 2008.

[Gur03] Leonid Gurvits. Classical deterministic complexity of Edmonds' Problem and quantum entanglement. In *Proc. thirty-fifth ACM Symp. Theory Comput. - STOC '03*, page 10, New York, New York, USA, mar 2003. ACM Press.

[GZFG15] Mark W. Girard, Yuriy Zinchenko, Shmuel Friedland, and Gilad Gour. Erratum: Numerical estimation of the relative entropy of entanglement [Phys. Rev. A 82, 052336 (2010)]. *Phys. Rev. A*, 91(2):029901, 2015.

[HGKM10] Matthew B. Hastings, Iván González, Ann B. Kallin, and Roger G. Melko. Measuring renyi entanglement entropy in quantum Monte Carlo simulations. *Phys. Rev. Lett.*, 104(15):2–5, 2010.

[HH99] Michał Horodecki and Paweł Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Phys. Rev. A*, 59(6):4206–4216, jun 1999.

[HHHH09] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81(2):865–942, jun 2009.

[HHO03] Michał Horodecki, Paweł Horodecki, and Jonathan Oppenheim. Reversible transformations from pure to mixed states and the unique measure of information. *Phys. Rev. A*, 67(6):062104, jun 2003.

[HJ94] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1994.

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, 2 edition, 2013.

[HM13] Aram W. Harrow and Ashley Montanaro. Testing Product States, Quantum Merlin-Arthur Games and Tensor Optimization. *J. ACM*, 60(1):1–43, feb 2013.

[HMPB11] Fumio Hiai, Milan Mosonyi, Dénes Petz, and Cédric Bény. Quantum \$f\$-divergences and error correction. *Rev. Math. Phys.*, 23(07):691–747, aug 2011.

[HO13] Michał Horodecki and Jonathan Oppenheim. (Quantumness in the Context of) Resource Theories. *Int. J. Mod. Phys. B*, 27(01n03):1345019, jan 2013.

[Hor97] Paweł Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Phys. Lett. A*, 232(5):333–339, aug 1997.

[Hor01] Michał Horodecki. Entanglement measures. *Quantum Inf. Comput.*, 1(1):3–26, 2001.

[Ish03] Satoshi Ishizaka. Analytical formula connecting entangled states and the closest disentangled state. *Phys. Rev. A*, 67(6):060301, jun 2003.

[KRS09] Robert König, Renato Renner, and Christian Schaffner. The Operational Meaning of Min- and Max-Entropy. *IEEE Trans. Inf. Theory*, 55(9):4337–4347, sep 2009.

[LHL17] Zi-Wen Liu, Xueyuan Hu, and Seth Lloyd. Resource Destroying Maps. *Phys. Rev. Lett.*, 118(6):060502, feb 2017.

[LMT15] Bruno F. Lourenço, Masakazu Muramatsu, and Takashi Tsuchiya. Solving SDP Completely with an Interior Point Oracle. *Optim. Online*, (C):1–20, 2015.

[LY08] David G. Luenberger and Yinyu Ye. *Linear and Nonlinear Programming*, volume 116 of *International Series in Operations Research & Management Science*. Springer, Boston, MA, 2008.

[MD09] Milán Mosonyi and Nilanjana Datta. Generalized relative entropies and the capacity of classical-quantum channels. *J. Math. Phys.*, 50(7):072104, 2009.

[MH11] Milán Mosonyi and Fumio Hiai. On the Quantum R{é}nyi Relative Entropies and Related Capacity Formulas. *IEEE Trans. Inf. Theory*, 57(4):2474–2487, apr 2011.

[MI08] Adam Miranowicz and Satoshi Ishizaka. Closed formula for the relative entropy of entanglement. *Phys. Rev. A*, 78(3):032310, may 2008.

[Mis11] Jarosław Adam Miszczak. Generating and using truly random quantum states in Mathematica. *Comput. Phys. Commun.*, 183(1):12, feb 2011.

[MLDS$^+$13] Martin Müller-Lennert, FreÌĄdeÌĄric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum ReÌĄnyi entropies: A new generalization and some properties. *J. Math. Phys.*, 54(12):122203, jun 2013.

[Mon13] Ashley Montanaro. Weak Multiplicativity for Random Quantum Channels. *Commun. Math. Phys.*, 319(2):535–555, jan 2013.

[MS14] Iman Marvian and Robert W Spekkens. Extending Noether's theorem by quantifying the asymmetry of quantum states. *Nat. Commun.*, 5(May):6, 2014.

[Nie99] Michael A. Nielsen. Conditions for a Class of Entanglement Transformations. *Phys. Rev. Lett.*, 83(2):436–439, jul 1999.

[NN94] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming.* Society for Industrial and Applied Mathematics, jan 1994.

[NOP09] Miguel Navascués, Masaki Owari, and Martin B. Plenio. Power of symmetric extensions for entanglement detection. *Phys. Rev. A - At. Mol. Opt. Phys.*, 80(5):1–16, 2009.

[NV01] Michael A. Nielsen and Guifré Vidal. Majorization and the interconversion of bipartite states. *Quantum Inf. Comput.*, 1(1):76–93, 2001.

[OP93] Masanori Ohya and Dénes Petz. *Quantum Entropy and its Use.* Springer, 1993.

[PBHS11] Lukasz Pankowski, Fernando G. S. L. Brandao, Michal Horodecki, and Graeme Smith. Entanglement distillation by extendible maps. *Quantum Inf. Comput.*, 13(9-10):0751–0770, sep 2011.

[Pet86] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports Math. Phys.*, 23(1):57–65, feb 1986.

[Pet10] Dénes Petz. From f-Divergence to Quantum Quasi-Entropies and Their Use. *Entropy*, 12(3):304–325, mar 2010.

[PPHH10] Lukasz Pankowski, Marco Piani, Michał Horodecki, and Paweł Horodecki. A Few Steps More Towards NPT Bound Entanglement. *IEEE Trans. Inf. Theory*, 56(8):4085–4100, aug 2010.

[PV07] Martin B. Plenio and Shashank Virmani. An introduction to entanglement measures. *Quantum Inf. Comput.*, 7(1):1–51, apr 2007.

[Rai97] Eric M. Rains. Entanglement purification via separable superoperators. jul 1997.

[Rai99a] E. M. Rains. Bound on distillable entanglement. *Phys. Rev. A*, 60(1):179–184, jul 1999.

[Rai99b] E. M. Rains. Rigorous treatment of distillable entanglement. *Phys. Rev. A*, 60(1):173–178, jul 1999.

[Rai00] E. M. Rains. Erratum: Bound on distillable entanglement [Phys. Rev. A 60, 179 (1999)]. *Phys. Rev. A*, 63(1):019902, dec 2000.

[Rai01] E. M. Rains. A semidefinite program for distillable entanglement. *IEEE Trans. Inf. Theory*, 47(7):2921–2933, aug 2001.

[RB01] Robert Raussendorf and Hans J. Briegel. A One-Way Quantum Computer. *Phys. Rev. Lett.*, 86(22):5188–5191, may 2001.

[Ren01] James Renegar. *A Mathematical View of Interior-point Methods in Convex Optimization.* MPS-SIAM Series on Optimization. SIAM, 2001.

[Ren05] Renato Renner. *Security of Quantum Key Distribution.* PhD thesis, dec 2005.

[Roc70] R. Tyrell Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[San10] Yuval Rishu Sanders. *Resource Theories in Quantum Information.* PhD thesis, University of Calgary, 2010.

[SB35] Erwin Schrödinger and M. Born. Discussion of Probability Relations between Separated Systems. *Math. Proc. Cambridge Philos. Soc.*, 31(04):555, oct 1935.

[SEG$^+$16] Gael Sentís, Christopher Eltschka, Otfried Gühne, Marcus Huber, and Jens Siewert. Quantifying Entanglement of Maximal Dimension in Bipartite Mixed States. *Phys. Rev. Lett.*, 117(19):190502, nov 2016.

[Sha10] Naresh Sharma. Equality conditions for the quantum f-relative entropy and generalized data processing inequalities. In *2010 IEEE Int. Symp. Inf. Theory*, pages 2698–2702. IEEE, jun 2010.

[Sha12] Naresh Sharma. Equality conditions for the quantum f-relative entropy and generalized data processing inequalities. *Quantum Inf. Process.*, 11(1):137–160, feb 2012.

[Sla50] Morton Slater. Lagrange Multipliers Revisited. In *Traces Emerg. Nonlinear Program.*, number 80, pages 293–306. Springer Basel, Basel, 1950.

[SRBL17] Alexander Streltsov, Swapan Rana, Manabendra Nath Bera, and Maciej Lewenstein. Towards Resource Theory of Coherence in Distributed Scenarios. *Phys. Rev. X*, 7(1):011024, mar 2017.

[TCR09] Marco Tomamichel, Roger Colbeck, and Renato Renner. A Fully Quantum Asymptotic Equipartition Property. *IEEE Trans. Inf. Theory*, 55(12):5840–5847, dec 2009.

[Tuy08] Hoang Tuy. Cutting Plane Methods for Global Optimization. In Christodoulos A. Floudas and Panos M. Pardalos, editors, *Encycl. Optim.*, pages 591–595. Springer, NeW York, 2008.

[TV00] Barbara M. Terhal and Karl Gerd H. Vollbrecht. Entanglement of Formation for Isotropic States. *Phys. Rev. Lett.*, 85(12):2625–2628, sep 2000.

[Uhl10] Armin Uhlmann. Roofs and Convexity. *Entropy*, 12(7):1799–1832, jul 2010.

[VB96] Lieven Vandenberghe and Stephen Boyd. Semidefinite Programming. *SIAM Rev.*, 38(1):49–95, mar 1996.

[VHGE14] Victor Veitch, S. A. Hamed Mousavian, Daniel Gottesman, and Joseph Emerson. The resource theory of stabilizer quantum computation. *New J. Phys.*, 16, 2014.

[Vid00a] Guifré Vidal. Entanglement monotones. *J. Mod. Opt.*, 47(2-3):355–376, feb 2000.

[Vid00b] Guifré Vidal. Optimal local preparation of an arbitrary mixed state of two qubits: Closed expression for the single-copy case. *Phys. Rev. A - At. Mol. Opt. Phys.*, 62(6):062315–062311, 2000.

[Vid03] Guifré Vidal. Efficient Classical Simulation of Slightly Entangled Quantum Computations. *Phys. Rev. Lett.*, 91(14):147902, oct 2003.

[VP98] Vlatko Vedral and Martin B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57(3):1619–1633, mar 1998.

[VPRK97] Vlatko Vedral, Martin B. Plenio, M. A. Rippin, and P. L. Knight. Quantifying Entanglement. *Phys. Rev. Lett.*, 78(12):2275–2279, mar 1997.

[VV14] Umesh Vazirani and Thomas Vidick. Fully Device-Independent Quantum Key Distribution. *Phys. Rev. Lett.*, 113(14):140501, sep 2014.

[VW01] Karl Gerd H. Vollbrecht and Reinhard F. Werner. Entanglement measures under symmetry. *Phys. Rev. A*, 64(6):062307, nov 2001.

[Wat16] John Watrous. *Theory of Quantum Information.* 2016.

[Wer89] Reinhard F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40(8):4277–4281, 1989.

[Woo98] W. K. Wootters. Quantum entanglement as a quantifiable resource. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, 356(1743):1717–1731, 1998.

[WY16] Andreas Winter and Dong Yang. Operational Resource Theory of Coherence. *Phys. Rev. Lett.*, 116(12):1–6, 2016.

[Zal08] Constantin Zalinescu. On Zero Duality Gap and the Farkas Lemma for Conic Programming. *Math. Oper. Res.*, 33(2):991–1001, 2008.

[ZFG10] Yuriy Zinchenko, Shmuel Friedland, and Gilad Gour. Numerical estimation of the relative entropy of entanglement. *Phys. Rev. A*, 82(5):052336, nov 2010.

[ZPNC11] Karol Zyczkowski, Karol a. Penson, Ion Nechita, and Benoit Collins. Generating random density matrices. *J. Math. Phys.*, 52(6):062201, 2011.

Appendix A

Pure to isotropic conversion witness

For a fixed Schmidt vector $\boldsymbol{\lambda}$ we define

$$f_k(\boldsymbol{\mu}) = E_k(\boldsymbol{\lambda}) - E_k(\boldsymbol{\mu}),$$

and write this as $f_k(\boldsymbol{\mu}) = \mu_1 + \cdots + \mu_k - (\lambda_1 + \cdots + \lambda_k)$. The goal is to compute

$$W_{\text{iso}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \max_{\boldsymbol{\mu}} \min_{k} f_k(\boldsymbol{\mu}).$$

This can be split into $d-1$ separate optimization problems as follows. For each $k \in \{1, \ldots, d-1\}$, we maximize $f_k(\boldsymbol{\mu})$ over all $\boldsymbol{\mu}$ for which k yields the minimum. That is, maximize over all $\boldsymbol{\mu}$ for which $f_k(\boldsymbol{\mu}) \leq f_\ell(\boldsymbol{\mu})$ for all $\ell \in \{1, \ldots, d-1\}$. Minimizing this over all k yields the desired result

$$W_{\text{iso}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{k} \left[\max_{\boldsymbol{\mu}} \left\{ f_k(\boldsymbol{\mu}) | f_k(\boldsymbol{\mu}) \leq f_\ell(\boldsymbol{\mu}) \text{ for all } \ell \right\} \right]$$

where the maximizations are taken over all Schmidt vectors satisfying $\sum_{i=1}^{d} \sqrt{\mu_i} = \sqrt{db}$.

For each k, these suboptimization problems can be rewritten as follows:

$$\begin{aligned} \text{maximize: } & \sum_{i=1}^{k} \mu_i \\ \text{subject to: } & \sum_{i=1}^{d} \mu_i = 1 \\ & \sum_{i=1}^{d} \sqrt{\mu_i} = \sqrt{db} \\ & \sum_{i=2}^{\ell} \mu_i \leq \sum_{i=2}^{\ell} \lambda_i \text{ for all } \ell \in \{1, \ldots, k-1\} \\ & \sum_{i=k+1}^{\ell+1} \lambda_i \leq \sum_{i=k+1}^{\ell+1} \mu_i \text{ for all } \ell \in \{k+1, \ldots, d-1\}. \end{aligned}$$

There are d constraints for these d-dimensional optimization problems, so we may use the method of Lagrange multipliers to find optimal solutions.

Appendix B

Proofs for the conic version of Farkas' Lemma

This appendix supplies proofs of Theorem 3.12 and Lemma 3.13 for the conic version of Farkas' Lemma. Proofs of these results can be found in many textbooks on convex analysis and convex optimization[1]. However, for completeness and because of our use of nonstandard notation, we include proofs of the main results from the theory of conic programs here.

Proof (*of Theorem 3.12*). It is easy to see that both statements cannot hold simultaneously. Indeed, suppose (i) holds and let $Y \in \mathrm{H}(\mathcal{H}')$ such that $\Phi^*(Y) \in \mathcal{K}^*$. Then

$$\langle Y, B \rangle = \langle Y, \Phi(X) \rangle = \langle \Phi^*(Y), X \rangle \geq 0$$

since $\Phi^*(Y) \in \mathcal{K}^*$ and $X \in \mathcal{K}$, and thus (ii) does not hold. Suppose instead that (i) does not hold. Since $\Phi(\mathcal{K})$ is a closed and convex cone, and $B \notin \Phi(\mathcal{K})$, by the hyperplane separation theorem (Theorem 3.7) there exists $Y \in \mathrm{H}(\mathcal{H}')$ with $Y \neq 0$ such that

$$\langle Y, \Phi(X) \rangle \geq 0 > \langle Y, B \rangle$$

[1] See for example [LY08, Section 6.3], [GM12, Section 4.5], and [Ren01, Section 3.2]. See also [Zal08].

for all $X \in \mathcal{K}$. Thus $\langle Y, B \rangle < 0$ and $\langle \Phi^*(Y), X \rangle \geq 0$, so $\Phi^*(Y) \in \mathcal{K}^*$. This completes the proof. □

Proof (*of Lemma 3.13*). Let $\{Y_i\}$ be a Cauchy sequence in $\Phi(\mathcal{K})$. Then this sequence converges to some $Y \in \mathrm{H}(\mathcal{H}')$. We will show that $Y \in \Phi(\mathcal{K})$. Since $\{Y_i\}$ is bounded, there is a real number $c > 0$ such that

$$c \geq |\langle Y_i, Z \rangle|$$

for all i. Since $Y_i \in \Phi(\mathcal{K})$, there exists $X_i \in \mathcal{K}$ such that $Y_i = \Phi(X_i)$ for all i. If $Y = 0$ then $Y \in \Phi(\mathcal{K})$, so we may suppose without loss of generality that $Y_i \neq 0$, and thus $X_i \neq 0$ for all i. Hence

$$\langle Y_i, Z \rangle = \langle X_i, \Phi^*(Z) \rangle > 0$$

for all i since $\Phi^*(Z) \in \mathrm{int}(\mathcal{K}^*)$.

Define the sequence $\{\tilde{X}_i\}$ by $\tilde{X}_i = X_i / \|X_i\|$. Clearly $\{\tilde{X}_i\}$ is bounded and $\tilde{X}_i \in \mathcal{K}$. Since $\mathcal{K}$ is closed, the sequence has at least one accumulation point in $\mathcal{K}$ which we denote $\tilde{X}$. Note $\tilde{X} \in \mathcal{K}$ and $\tilde{X} \neq 0$, hence $\langle \tilde{X}, Z \rangle > 0$. Furthermore, for each i it holds that

$$0 < \langle \tilde{X}_i, Z \rangle \leq \frac{1}{\|X_i\|} c.$$

Suppose now that $\{X_i\}$ were unbounded, then $\langle \tilde{X}_i, Z \rangle \to 0$ since $\|X_i\| \to \infty$. But this is a contradiction, since $\lim_{i \to \infty} \langle \tilde{X}_i, Z \rangle = \langle \tilde{X}, Z \rangle > 0$. Hence $\{X_i\}$ must be bounded and the sequence must have at least one accumulation point $X \in \mathcal{K}$ such that $\Phi(X) = Y$. This completes the proof. □

Appendix C

Code for convex optimization problems

This appendix shows the MATLAB code for implementing the algorithms presented in 4.6 for computing the relative entropy of entanglement. This code can be found online at [?], and is a modification of the code that first appear in [ZFG10] and modified in [GZFG15].

```
function [Xopt,uBound,lBound,outerCount] = ...
    relEntropy(m,n,A,eps,maxIter,lineSearchEps)
% relEntropy
%--------------------
% Approximates the relative entropy of entanglement (REE) of a density
% matrix of an mxn bipartite system (relative to PPT states).
%  -Uses CVX solver for semidefinite programming (http://cvxr.com/cvx/).
%  -Uses PartialTranspose from QETLAB (http://www.qetlab.com) to compute
%    partial transposes of matrices.
%https://github.com/nathanieljohnston/QETLAB/blob/master/PartialTranspose.m
%
% Standard usage:  [Xopt,relEntr]=relEntropy(m,n,A)
% Variables:
```

```
%    m,n       - dimensions of the subsystems
%    A         - density matrix whose REE we are trying to compute
%    Xopt      - optimal PPT matrix that minimizes the relative entropy
%    relEnt    - output upper bound of relative entropy of entanglement
%
% Optional inputs with defaults:
% relEntropy(m,n,A,eps,maxIter,lineSearchEps)
%    eps           - precision such that |uBound-lBound|<eps (default: ...
   eps = 1e-5)
%    maxIter       - max number of iterations (default: maxIter = 200)
%    lineSearchEps - precision of intermediate line search (default: ...
   lineSearchEps = 1e-10;)
%
% We define a function traceAlogmA to compute trace(A*logm(A))to accept
% rank-defficnent matrices.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%  Check the input arguments    %%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if nargin<3
    error('Not enough arguments; input m,n,A')
end
% check if A is indeed positive semi-definite
if min(eig(A))<-1e-12 || max(max(abs(A-A')))>1e-12 || ...
   max(size(A)≠[m*n,m*n])
    error('A must be positive semi-definite (mn x mn)-matrix');
end
% check if A is trace-1 (within some numerical tolerance level)
if abs(1-trace(A))>1e-12
    error('A must be trace-1; |1-trace(A)| exceeds the allowed ...
        1e-12-tolerance');
end
%%%%%%%
```

```
% Set the parameters if not specified
%
% If not specified, set the default precision.
if nargin<4
    eps=1e-5;
end
% if not specified, set the default maximum number of outer iterations.
if nargin<5
    maxIter=200;
end
% If not specified, set the default line-search precision.
% This appears to impact our ability to find an eps-approximate solution.
% Fixed lineSearchEps=1e-10 seems to work better than the adaptive
% choice of eps^(3/2)
if nargin<6
    lineSearchEps=1e-10;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%  Check if input state is PPT   %%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if (min(real(eig(PartialTranspose(A,1,[m,n])))))≥0
    lBound=0;
    uBound=0;
    Xopt=A;
    disp('A is PPT, thus Xopt=A and relEntropy=0');
    return;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%% Initialize search  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```
% set blank output
lBound=-Inf;
uBound=Inf;
Xopt=[];
% set outer iteration counter
outerCount=0;
% set initial numerical status
status=0;
% initialze N
N=0;
outerCount=0;

% Initialize list of X.
%   This is the list of points in the interior of
%   the PPT states. We construct the tangent planes at each of these
%   points to create a polytope approximation of the epigraph.
X{N+1}=eye(m*n)/(m*n);

% Generate list of E.
%   Used for the approximate epigraph of the objective. These generate the
%   'Gateaux derivatives' of the tr(A*logm(X)) function at each X.
%   (Here we provide a generic script that would work for any N,
%   although for now we always start with N=0)
for i=0:N
    % build E^{(i)} from X^{(i)}
    [U,L]=eig(X{i+1});
    for j=1:m*n
        for k=1:m*n
            if L(j,j)==L(k,k) D(j,k)=1/L(j,j);
            else D(j,k)=(log(L(j,j))-log(L(k,k)))/(L(j,j)-L(k,k));
            end
        end
    end
    E{i+1}=U*(D.*(U'*A*U))*U';
end

% symmetrize E
for i=1:N+1
    E{i}=(E{i}+E{i}')/2;
end

% Define list of b's.
%   These are the values b{i}=-trace(A*logm(X))+trace(E*X) for each X ...
    and E.
for i=1:N+1
    b{i}=-trace(A*logm(X{i}))+trace(E{i}*X{i});
end
% Make into a vector so it can be used in cvx
bvect=zeros([N+1,1]);
for i=1:N+1
    bvect(i)=b{i};
end

% Start iterating until we reach the prescribed precision eps
% or unitl we exceed maximum number of iterations, i.e., outerCount>maxIter

% Set bestN index to point to the best upper bound out of X{i}, i=0,...,N
bestN=N;
% Re-initialize the bounds (we know something already)
lBound=-real(traceAlogmA(A));
uBound=-trace(A*logm(eye(m*n)/(m*n)));
% Set probSolved_flag to indicate wether the problem is solved yet
probSolved_flag=0;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%% Start optimization program %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

while ¬probSolved_flag && outerCount<maxIter %&& ¬status
    % Formulate and solve the approximation SDP problem.
    % Variables are for SDP are:
    %     (Y,t)   - where Y is positve n*m by n*m definite PPT matrix ...
       with trace(Y)=1
    %               and (Y,t) is in approximation to epigraph of ...
       -trace(A*log(Y))
    %               so that t≥lBound.
    %     s       - s>0
    % Main constraint is
    %     -trace(A*logm(X{i}))+trace(E{i}*X{i})-trace(E{i}*Y) ≤ t

    % Use CVX to solve SDP problem
    cvx_begin sdp quiet
    %cvx_precision high;
    variable t
    variable s(N+1)
    variable Y(m*n,m*n) hermitian
    expression V(N+1)
    for i=1:N+1
        V(i)=trace(E{i}*Y);
    end
    minimize t
    subject to
        t≥lBound;
        s≥0;
        PartialTranspose(Y,1,[m,n])≥0;
        trace(Y) == 1;
        Y ≥ 0;
        s+bvect-V==t*ones([N+1,1]);
    cvx_end
```

```

% After SDP, use line search to find better optimal.
Ystart=X{bestN+1};
Yend=Y;
% set the search ray direction
dY=Yend-Ystart;
% set the mid point and the objective derivative
Ynext=(Ystart+Yend)/2;
[U,L]=eig(Ynext);
for j=1:m*n
    for k=1:m*n
      if L(j,j)==L(k,k) D(j,k)=1/L(j,j);
      else D(j,k)=(log(L(j,j))-log(L(k,k)))/(L(j,j)-L(k,k));
      end
    end
end
Enext=U*(D.*(U'*A*U))*U';
df=-trace(Enext*dY);
% iterate (with 'cheap' norm)
while (norm(Yend(:)-Ystart(:))>lineSearchEps)
   if df<0 Ystart=Ynext;
   else Yend=Ynext;
   end
   % recompute the mid point and the objective derivative
   Ynext=(Ystart+Yend)/2;
   [U,L]=eig(Ynext);
   for j=1:m*n
        for k=1:m*n
             if L(j,j)==L(k,k) D(j,k)=1/L(j,j);
             else D(j,k)=(log(L(j,j))-log(L(k,k)))/(L(j,j)-L(k,k));
             end
        end
   end
```

```
       Enext=U*(D.*(U'*A*U))*U';
       df=-trace(Enext*dY);
    end
    %Use Xnext for next point in list of P
   Y=Ynext;
    N=N+1;
    X{N+1}=Y;
    lBound=max(lBound,t);
    if uBound>-trace(A*logm(Y))
         uBound=-trace(A*logm(Y));
         bestN=N;
    end

    s=sprintf('  [%d] lower bound: %e, upper bound: %e, gap: %e, relGap ...
       %d%%', outerCount, lBound, uBound, uBound-lBound, ...
       round(100*(uBound-lBound)/uBound));
    disp(s);

    [U,L]=eig(X{N+1});
    for j=1:m*n
        for k=1:m*n
           if L(j,j)==L(k,k) D(j,k)=1/L(j,j);
           else D(j,k)=(log(L(j,j))-log(L(k,k)))/(L(j,j)-L(k,k));
           end
        end
    end
    E{N+1}=U*(D.*(U'*A*U))*U';

    % Symmetrize E
    E{N+1}=(E{N+1}+E{N+1}')/2;

    % Define vector of b's
    b{N+1}=real(-trace(A*logm(X{N+1}))+trace(E{N+1}*X{N+1}));
    bvect=zeros([N+1,1]);
    for i=1:N+1
       bvect(i)=b{i};
    end

    % Verify if we found a solution
    if (uBound-lBound)<eps
        probSolved_flag=1;
    end

    % Increment outer iteration counter
    outerCount=outerCount+1;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%% Output %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Set the output
lBound=lBound+traceAlogmA(A);%trace(A*logm(A));
uBound=real(uBound+traceAlogmA(A));%trace(A*logm(A));
Xopt=X{bestN+1};

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [val] = traceAlogmA(A)
% overloads MATLAB's trace(A*logm(A)) to accept
% rank-defficient positive semi-definite A
% by computing the limiting value
D=eig(A);
idx=find(D);
val=sum(D(idx).*log(D(idx)));
```