TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2022

 | ISBN 978-0-xxxxxxx-x-x
http://?

Five Starter Pieces: Quantum Information Science via Semi-definite Programs

Vikesh Siddhu
JILA, University of Colorado/NIST, 440 UCB, Boulder, CO 80309, USA, vsiddhu@protonmail.com

Sridhar Tayur
Quantum Technologies Group, Tepper School of Business, Carnegie Mellon University, Pittsburgh PA 15213, USA, stayur@cmu.edu

Abstract As the title indicates, this chapter presents a brief, self-contained introduction to five fundamental problems in Quantum Information Science (QIS) that are especially well-suited to be formulated as Semi-definite Programs (SDP). We have in mind two audiences. The primary audience comprises of Operations Research (and Computer Science) graduate students who have familiarity with SDPs, but have found it daunting to become even minimally conversant with pre-requisites of QIS. The second audience consists of Physicists (and Electrical Engineers) already knowledgeable with modeling of QIS via SDP but interested in computational tools that are applicable more generally. For both audiences, we strive for rapid access to the unfamiliar material. For the first, we provide just enough required background material (from Quantum Mechanics, treated via matrices, and mapping them in Dirac notation) and simultaneously for the second audience we recreate, computationally in Jupyter notebooks, known closed-form solutions. We hope you will enjoy this introduction and gain understanding of the marvelous connection between SDP and QIS by self-study, or as a short seminar course. Ultimately, we hope this disciplinary outreach will fuel advances in QIS through their fruitful study via SDPs.

Keywords Quantum Information Science, Semi-definite programs, Quantum entanglement, Quantum channels, Quantum state discrimination, Channel capacity

1. Introduction

Thomas Sprat, in 1667, as historian at the Royal Society of London, noted a connection between being an outsider to a trade and inventiveness:

> A glance from an angle might well reveal a new aspect of nature.

We would like to create such a trading zone through this chapter - and invite Operations Researchers and Computer Scientists - to foster innovative contributions to Quantum Information Science (QIS).

QIS spans a variety of sub-fields including quantum computing and quantum communication [68]. Quantum computing offers a novel way to perform calculations which could be faster than regular (classical) computing for several important problem classes, such as prime factorization [84]. This novelty and speed comes from utilizing properties of quantum mechanics such as superposition and entanglement [46]. Quantum communication can not only carry a new type of information using qubits ("quantum bit"), it can also be used to communicate regular (classical) information (bit) with greater privacy [12, 13]. Furthermore, quantum communication can be non-additive: two quantum communication devices

can send more together than each device used separately [93]. These potential advantages of using quantum states for information processing are often abated by noise. This noise affects quantum states and their properties important for computation and communication. To fully understand and leverage quantum technologies computation and communication, it is important to study basic properties of quantum states such as entanglement and information theoretic properties such as capacity to carry information [14]. In such studies semi-definite programs (SDPs) play a useful role. Since SDPs are a well-known tool in Operations Research (OR) and Computer Science (CS), they offer researchers in OR and CS communities a natural way to interact with QIS.

Semi-definite programs (for a brief introduction, see [71,96,98]) are an extension of linear programs (LPs) obtained by replacing element wise non-negative vector variables in LPs with positive semi-definite matrices. This replacement results in a non-linear, but convex optimization problem, which is much more general than an LP. However, this general SDP carries with it a variety of nice properties of LPs which make it possible to efficiently solve SDPs, both in theory and practice. For instance, most of the theory of duality directly extends from LPs to SDPs [11]. The simplex method for LPs [22, 80] can, in principle, be extended to SDPs [73]. For SDP constraints, one can construct cutting planes in polynomial time [18] and thus use a polynomial-time ellipsoid method [66, 83] to numerically solve an SDP. However, in practice it is often faster to use interior point methods [1, 49, 67] (such as those extending Karmarkar's interior point method [52] for LPs) to efficiently solve SDPs.

The ability to efficiently solve SDPs is not their only draw. These optimization programs naturally appear in a variety of fields including control theory [18], graph theory [33, 50], combinatorial optimization [2], and algebraic geometry [37]. SDPs in these and other engineering fields usually have real positive semi-definite matrix variables. *Complex* positive semi-definite matrices naturally appear, and play an important role in quantum mechanics, a linear theory in the physical sciences which successfully describes the physical world. It is no surprise that a variety of fundamental and applied problems in quantum theory can be re-written as SDPs [55, 99, 106]. Such re-writing has been fueled by the growth of quantum computing and information science, which study and hope to practically perform information processing using physical objects accurately described by quantum theory. Quantum information science offers an exciting and potentially fertile area where SDPs and other optimization techniques can continue to play an important role. Standard exposition of quantum mechanics often involves new notation, unitary dynamics, and other historical aspects of quantum theory. This route to learning quantum mechanics has its advantages, but it can create a barrier to entry for optimization experts working outside the area of quantum information science (QIS). The key motivation for this work is to lower this barrier and expose a broader audience to the recent SDP work in QIS.

In what follows, in Secs. 2 and 3 we provide a bare bones introduction to quantum mechanics and information theory with running examples, most using 2×2 matrices. In this introduction we not only cover basic concepts likes quantum states, measurements, Born's rule (see Sec. 2.2), entanglement, entropy, and quantum channels, but also take the opportunity to introduce *Dirac notation*, which is standard across quantum theory and QIS. Next we present five problems in QIS: quantum state discrimination (in Sec. 4), state fidelity (in Sec. 5), channel discrimination (in Sec. 6), entanglement and separability (in Sec. 7), and channel capacity (in Sec. 8). These problems, presented in order of increasing level of mathematical sophistication, by no means comprise a complete list of problems in QIS where SDPs and other optimization techniques are of value. However, they offer a strong stepping stone to continue future exploration of this type. For each problem, we provide a motivation, a crisp mathematical statement, an SDP formulation, certain special cases (sometimes with algebraic SDP solutions), numerical examples with working Python notebooks, and avenues for future exploration.

2. Quantum Mechanics: The Bare Minimum

2.1. Quantum states and Dirac notation

Quantum mechanics accurately predicts measurable properties of microscopic physical objects. While these physical objects live in the real world, they are described in quantum mechanics using complex numbers. The simplest complex number, i, is just the square root of minus one. In general, any complex number can be written as $z = x + iy$, where x, y are real numbers; that is, $x, y \in \mathbb{R}$. We say $z \in \mathbb{C}$, the space of complex numbers. The complex conjugate of z is $x - iy$ and is represented by z^*. Using z and z^* one constructs the norm, $|z| = \sqrt{z^* z} = \sqrt{x^2 + y^2}$, of a complex number.

Of main interest in quantum mechanics are tuples of complex numbers. A length d tuple of this type is just a column vector $\mathbf{v}$ in d-dimensional complex space $\mathbb{C}^d$. The inner product of a column vector $\mathbf{v}$ with $\mathbf{w}$, $\mathbf{v}^* \cdot \mathbf{w}$, resembles the ordinary dot product $\mathbf{v} \cdot \mathbf{w}$ of real vectors, except the entries of the first column vector $\mathbf{v}$ are complex conjugated. A collection of d column vectors, $\{\mathbf{v}_i\}$, where each vector has unit norm ($\mathbf{v}_i^* \mathbf{v}_i = 1$) and any distinct pair of vectors are orthogonal, that is, the inner product, $\mathbf{v}_i^* \mathbf{v}_j = 0$, for $i \neq j$, is called an *orthonormal basis* of $\mathbb{C}^d$. Using this orthonormal basis, any vector in $\mathbb{C}^d$ can be written as a linear combination, $\sum_i c_i \mathbf{v}_i$, where $c_i \in \mathbb{C}$.

In quantum physics literature, the space $\mathbb{C}^d$, its column vectors $\mathbf{v}$, and the inner product of a column $\mathbf{v}$ with another column vector $\mathbf{w}$ are denoted by $\mathcal{H}$, $|v\rangle$, and $\langle v|w\rangle$, respectively. This notation is called *Dirac* notation, where $\langle v|w\rangle$ is called a *braket*, its first half, $\langle v|$, is called a *bra* and the second half, $|w\rangle$ is called a *ket*. The ket $|v\rangle$ is represented by a column vector. Taking the transpose of this column vector and then complex conjugating each entry results in a row vector. This row vector represents the bra $\langle v|$. The multiplication of a ket $|v\rangle$ by a scalar $c \in \mathbb{C}$ is denoted as $c|v\rangle$. Any ket $|v\rangle$ that has unit inner product with itself, $\langle v|v\rangle = 1$, is called a *pure state*.

The inner product $\langle v|w\rangle$ is a complex number obtained by multiplying the row vector $\langle v|$ with the column vector $|w\rangle$. By interchanging the order of multiplication, we multiply a d-dimensional column vector $|w\rangle$ with a d-dimensional row vector $\langle v|$ to obtain the outer-product, $|w\rangle\langle v|$, which is a $d \times d$ square matrix with complex entries. This square matrix represents a linear operator. We denote the set of linear operators on $\mathcal{H}$ by $\mathcal{L}(\mathcal{H})$. The action of the operator $|w\rangle\langle v|$ on $|u\rangle \in \mathcal{H}$ is given by

$$(|w\rangle\langle v|)|u\rangle = |w\rangle(\langle v|u\rangle) = (\langle v|u\rangle)|w\rangle. \tag{1}$$

The equality on the right technically defines the action of the operator on the left. However, the middle term, obtained by removal of a parenthesis and replacement of two vertical bars $||$ between v and u with one bar $|$ is an example of slickness embedded in Dirac notation. This slickness explains the action of operators without doing a matrix calculation. For instance, the result in (1) is essentially a matrix calculation where the $(d \times d)$ matrix for $|w\rangle\langle v|$ is multiplied with a d-dimensional column vector $|u\rangle$, to obtain the outcome, $(\langle v|u\rangle)|w\rangle$.

Unlike $|w\rangle\langle v|$, not all linear operators are dyads, linear operators can be written as sums of dyads and represented by matrices. The transpose of a matrix M is denoted by M^T, and the adjoint $M^\dagger$ is obtained by complex conjugating each entry of M^T (see footnote * for commment on the notation). If N is another linear operator, then $(NM)^\dagger = M^\dagger N^\dagger$. For square matrices, those with equal numbers of rows and columns, we denote the matrix determinant by $\det(M)$. Square matrices with non-zero determinants can be inverted, and

*We have used notation which is common in physics where $*$ and † denote complex conjugate and adjoint operations, respectively. In mathematics and optimization, it is common to use $\bar{z}$ for complex conjugate of z and $*$ for adjoint operation. There is yet another combination, $\bar{z}$ for complex conjugate of z but † for adjoint operations, which can be seen in some physics, computer science and optimization literature.

the matrix inverse M^{-1} satisfies $MM^{-1} = M^{-1}M = I$, where I is the identity matrix. For any square matrix M, there is a set of non-zero vectors $\{|a_i\rangle\}$ such that M satisfies

$$M|a_i\rangle = \alpha_i|a_i\rangle. \tag{2}$$

Here complex numbers α_i are called eigenvalues and $|a_i\rangle$ are called eigenvectors. A square matrix M which commutes with its adjoint, $MM^\dagger = M^\dagger M$, is called a normal matrix. Any normal matrix M can be diagonalized using an orthonormal basis,

$$M = \sum_i \alpha_i |m_i\rangle\langle m_i|, \tag{3}$$

where the basis vector $|m_i\rangle$ is an eigenvector of M and has the complex eigenvalue α_i. Two special types of normal matrices are of particular interest in quantum mechanics. One is a unitary matrix, usually denoted by U, which satisfies $UU^\dagger = U^\dagger U = I$; another is a Hermitian matrix, O, which satisfies $O = O^\dagger$. Before proceeding forward, we illustrate the use of the adjoint operation † in Dirac notation. Suppose $|\psi\rangle$ is any ket, then $|\psi\rangle^\dagger = \langle\psi|$ and $\langle\psi|^\dagger = |\psi\rangle$. If $|\phi\rangle$ is another ket, then $(|\psi\rangle\langle\phi|)^\dagger = |\phi\rangle\langle\psi|$. For complex numbers c_0 and c_1, the adjoint of the linear combination $(c_0|\psi\rangle + c_1|\phi\rangle)^\dagger = c_0^*\langle\psi| + c_1^*\langle\phi|$. If N is a linear operator, then $\langle i|N^\dagger|j\rangle = (\langle j|N|i\rangle)^*$.

The simplest non-trivial space $\mathcal{H}$ has dimension two ($d = 2$); that is, any $|\psi\rangle \in \mathcal{H}$ can be written as a linear combination of two orthonormal vectors. It is customary to introduce a *standard* (or *computational*) basis $\{|0\rangle, |1\rangle\}$ for $\mathcal{H}$ with $d = 2$. Here $\langle 0|0\rangle = \langle 1|1\rangle = 1$ and $\langle 0|1\rangle = 0$. It is common to represent the computational basis as column vectors

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad |1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{4}$$

As mentioned earlier, the inner product $\langle 0|1\rangle$ can be obtained by multiplying each row of $|1\rangle$ with the complex conjugate of each row of $|0\rangle$. In general, the state of a two-dimensional quantum system, called a *qubit*, is given by $|\psi\rangle = c_0|0\rangle + c_1|1\rangle$ where $|c_0|^2 + |c_1|^2 = 1$; $|\psi\rangle$ can be written as a column vector

$$|\psi\rangle := \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}. \tag{5}$$

The notation $|0\rangle$ and $|1\rangle$ is intended to draw an analogy with classical bits. Just like the distinguishable states 0 and 1 of a bit, the quantum states $|0\rangle$ and $|1\rangle$ represent perfectly distinguishable states of a qubit. Like any classical analogy for a quantum system, this analogy between bits and qubits has its limitations. For instance, the linear combination $c_0|0\rangle + c_1|1\rangle$, where $|c_0|^2 + |c_1|^2 = 1, |c_0| > 0$, and $|c_1| > 0$, is a perfectly well-defined quantum state; however, there is no analogous state of a classical bit. There are particular linear combinations of the standard basis elements that are of special interest. One such linear combination is

$$|+\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and} \quad |-\rangle := \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{6}$$

Notice, $\langle +|+\rangle = \langle -|-\rangle = 1$ and $\langle +|-\rangle = 0$; as a result $\{|+\rangle, |-\rangle\}$, forms a basis of $\mathcal{H}$. This basis is sometimes called the *Hadamard basis* because it can be obtained from the computational basis, $|+\rangle = H|0\rangle$ and $|-\rangle = H|1\rangle$, using the Hadamard operator

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{7}$$

Notice the Hadamard matrix H is unitary ($HH^\dagger = H^\dagger H = I$) and Hermitian (i.e., $H = H^\dagger$); as a consequence $H^2 = I$.

Let $\{|j\rangle\}$ be the computational basis of a d-dimensional space $\mathcal{H}$. In Dirac notation, the identity operator I on $\mathcal{H}$ can be written as

$$I = \sum_{j=0}^{d-1} |j\rangle\langle j|. \tag{8}$$

The expression above is often called the *completeness relation*. This relation can be useful. For instance, suppose $|\psi\rangle$ is any ket in $\mathcal{H}$, represented by some column vector. One may use the completeness relation as follows:

$$|\psi\rangle = I|\psi\rangle = \sum_j (|j\rangle\langle j|)|\psi\rangle = \sum_j \langle j|\psi\rangle|j\rangle, \tag{9}$$

to find that $\langle j|\psi\rangle$ is simply the j^{th} entry (j starts from zero) of the column vector representing $|\psi\rangle$ in the standard basis. For instance, the column vector $|\psi\rangle$ in (5) has $c_0 = \langle 0|\psi\rangle$ and $c_1 = \langle 1|\psi\rangle$. The completeness relation can also be used to find entries of a linear operator N on $\mathcal{H}$,

$$N = INI = \Big(\sum_i |i\rangle\langle i|\Big) N \Big(\sum_j |j\rangle\langle j|\Big) = \sum_{i,j} \langle i|N|j\rangle |i\rangle\langle j| \tag{10}$$

to find that the i^{th} row and j^{th} column of N is simply $\langle i|N|j\rangle$. If d is two, then the matrix for N in the standard basis is simply

$$N = \begin{pmatrix} \langle 0|N|0\rangle & \langle 0|N|1\rangle \\ \langle 1|N|0\rangle & \langle 1|N|1\rangle \end{pmatrix}. \tag{11}$$

In general, the trace of an operator N, $\mathrm{Tr}(N)$, is simply $\sum_j \langle j|N|j\rangle$. Uses of the completeness relation, similar to those above, can show that $\mathrm{Tr}(|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle$ and $\mathrm{Tr}(A|\psi\rangle\langle\phi|) = \langle\phi|A|\psi\rangle$.

So far we have focused on a single quantum system. Suppose there are two systems a and b in quantum states $|v\rangle_a \in \mathcal{H}_a$ and $|w\rangle_b \in \mathcal{H}_b$, respectively. Then state of the joint ab system is written as $|v\rangle_a \otimes |w\rangle_b$, a tensor product, $\otimes$, of $|v\rangle_a$ and $|w\rangle_b$. This joint state belongs to a space $\mathcal{H}_{ab} = \mathcal{H}_a \otimes \mathcal{H}_b$ obtained by taking the tensor product of $\mathcal{H}_a$ with $\mathcal{H}_b$. Suppose $|v\rangle_a$ and $|w\rangle_b$ are represented by column vectors of dimension d_a and d_b, respectively; then $|v\rangle_a \otimes |w\rangle_b$ is given by a column vector of dimension $d_a d_b$ formed by taking the *Kronecker product* of each column vector. For instance, let $d_a = d_b = 2$,

$$|v\rangle_a = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}, \quad |w\rangle_b = \begin{pmatrix} d_0 \\ d_1 \end{pmatrix}, \tag{12}$$

where $|c_0|^2 + |c_1|^2 = |d_0|^2 + |d_1|^2 = 1$, then

$$|v\rangle_a \otimes |w\rangle_b = \begin{pmatrix} c_0 \begin{pmatrix} d_0 \\ d_1 \end{pmatrix} \\ c_1 \begin{pmatrix} d_0 \\ d_1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} c_0 d_0 \\ c_0 d_1 \\ c_1 d_0 \\ c_1 d_1 \end{pmatrix}. \tag{13}$$

In general, the state $|\psi\rangle \in \mathcal{H}_{ab}$ can be written as a linear combination of an orthonormal basis of $\mathcal{H}_{ab}$. One simple orthonormal basis of this type can be constructed by taking tensor products of the computational basis elements of $\mathcal{H}_a$ and $\mathcal{H}_b$, respectively. For instance, if $d_a = 2, d_b = 2, \{|0\rangle_a, |1\rangle_a\}$ and $\{|0\rangle_b, |1\rangle_b\}$ are computational basis of $\mathcal{H}_a$ and $\mathcal{H}_b$, respectively, then $\{|0\rangle_a \otimes |0\rangle_b, |0\rangle_a \otimes |1\rangle_b, |1\rangle_a \otimes |0\rangle_b, |1\rangle_a \otimes |1\rangle_b\}$ is an orthonormal basis for $\mathcal{H}_{ab}$. This basis can be represented as follows

$$|0\rangle_a \otimes |0\rangle_b = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |0\rangle_a \otimes |1\rangle_b = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |1\rangle_a \otimes |0\rangle_b = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad |1\rangle_a \otimes |1\rangle_b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{14}$$

States of the two qubit system $\mathcal{H}_{ab}$ can be written as a linear combination of the basis above. A simple linear combination of the basis elements above is

$$|\phi\rangle = \frac{1}{2}(\sum_{i,j} |i\rangle_a \otimes |j\rangle_b). \tag{15}$$

It turns out that $|\phi\rangle$ can be written as $|+\rangle_a \otimes |+\rangle_b$ and represents the state of two qubits, each in the state $|+\rangle$. Another simple linear combination is

$$|\chi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_a \otimes |0\rangle_b + |1\rangle_a \otimes |1\rangle_b). \tag{16}$$

Unlike $|\phi\rangle$, the linear combination above cannot be written as $|v\rangle_a \otimes |w\rangle_b$ for any $|v\rangle_a$ and $|w\rangle_b$. Thus, the joint system ab is in a state that cannot be adequately described by specifying the state of each individual system a and b. Such joint states are called *entangled*. Entanglement is a key aspect of quantum theory. In general, a state $|\psi\rangle_{ab}$, given by a linear combination $M_{ij}|i\rangle_a \otimes |j\rangle_b$, is entangled if the matrix M, with entries M_{ij}, has rank greater than one.

Given two matrices, N, mapping $\mathcal{H}_a$ to itself, and M, mapping $\mathcal{H}_b$ to itself, one can define their tensor product, $N \otimes M$, a matrix from $\mathcal{H}_a \otimes \mathcal{H}_b$ to itself which acts as follows:

$$N \otimes M(|v\rangle_a \otimes |w\rangle_b) = N|v\rangle_a \otimes M|w\rangle_b. \tag{17}$$

If square matrices N and M have dimensions d_a and d_b, respectively, then the square matrix $N \otimes M$ has dimension $d_a \times d_b$. This larger square matrix is obtained by taking a Kronecker product of N and M. For instance, let $d_a = d_b = 2$,

$$N = \begin{pmatrix} N_{00} & N_{01} \\ N_{10} & N_{11} \end{pmatrix}, \quad \text{and} \quad M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}, \tag{18}$$

then $N \otimes M =$

$$\begin{pmatrix} N_{00}\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}, & N_{01}\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}, \\ N_{10}\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}, & N_{11}\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}, \end{pmatrix} = \begin{pmatrix} N_{00}M_{00} & N_{00}M_{01} & N_{01}M_{00} & N_{01}M_{01} \\ N_{00}M_{10} & N_{00}M_{11} & N_{01}M_{10} & N_{01}M_{11} \\ N_{10}M_{00} & N_{10}M_{01} & N_{11}M_{00} & N_{11}M_{01} \\ N_{10}M_{10} & N_{10}M_{11} & N_{11}M_{10} & N_{11}M_{11} \end{pmatrix}. \tag{19}$$

2.2. Measurement and Born's Rule

In quantum mechanics, *physical variables* or *observables* are represented by Hermitian operators. As stated earlier, a *Hermitian* or *self-adjoint* operator O is one that equals its adjoint $O^\dagger$. The simplest Hermitian operator is a *projector*. A projector P is both Hermitian, $P^\dagger = P$, and idempotent, $P^2 = P$. The simplest projector has rank 1 and can be written as $P = |\psi\rangle\langle\psi|/\mathrm{Tr}(|\psi\rangle\langle\psi|)$ where $|\psi\rangle$ is any ket. In general, any Hermitian operator O (representing some observable) has a *spectral decomposition* using which it can be written as the sum of orthogonal projectors,

$$O = \sum_i \lambda_i P_i \tag{20}$$

where λ_i are distinct real numbers representing distinct eigenvalues of O, the projectors $\{P_i\}$ satisfy $P_i P_j = \delta_{ij} P_j$ —that is, they are orthogonal— and $\sum_i P_i = I$— that is, $\{P_i\}$ form a *projective decomposition* of the identity. If an observable O is measured on a system with state $|\phi\rangle$, then one obtains its eigenvalue λ_i as a measurement outcome. According to Born's rule, the probability of observing the value λ_i is

$$p_i = \mathrm{Tr}(P_i|\phi\rangle\langle\phi|). \tag{21}$$

One simple observable is the identity I. Its decomposition of the form (20) contains a single projector I corresponding to the eigenvalue 1. If I is measured on a system with state $|\phi\rangle$, then one obtains its eigenvalue, 1, as a measurement outcome with probability $1 = \mathrm{Tr}(I|\phi\rangle\langle\phi|)$. Another simple observable in two dimensions is $B = |1\rangle\langle 1|$. The decomposition (20) for B takes the form

$$B = 0 \cdot |0\rangle\langle 0| + 1 \cdot |1\rangle\langle 1|. \tag{22}$$

When the observable B is measured on the state $|\phi\rangle$ in (5), one obtains two outcomes, 0 with probability $|c_0|^2 = \mathrm{Tr}(|0\rangle\langle 0||\phi\rangle\langle\phi|)$ and 1 with probability $|c_1|^2 = \mathrm{Tr}(|1\rangle\langle 1||\phi\rangle\langle\phi|)$. Notice, the normalization condition, $|c_0|^2 + |c_1|^2 = 1$, stated below (5) ensures that the probabilities sum to one. This type of measurement, which results in the measurement of B, is called a measurement in the computational basis.

3. Quantum Information Theory: The Bare Minimum

An example of classical information is learning the outcome of an unbiased coin toss. This outcome takes values from a two-letter alphabet $\mathcal{X} = \{H, T\}$, where H and T represent heads and tails. The probability that the coin toss result X takes a value $x \in \mathcal{X}$ is $p(x) := \Pr(X = x)$, where $p(H) = p(T) = 1/2$ for an unbiased coin. For a general biased coin, $p(H) = p, p(T) = 1 - p$ and $0 \le p \le 1$. Learning the coin toss outcome provides classical information because this learning removes uncertainty in the outcome. The amount of uncertainty in the outcome of a biased coin with $p(H) = p$ is captured by the *binary entropy*,

$$h(p) := -p \log_2 p - (1-p) \log_2 (1-p), \tag{23}$$

measured in bits, where $0 \log 0 := 0$. When $p = 1/2$, $h(p) = 1$ bit, a result that agrees with the usual intuition that learning the outcome of an unbiased coin provides 1 bit of information. In general, any random variable X taking values x from a finite alphabet $\mathcal{X}$ with probability $p(x)$ has *Shannon entropy*,

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x). \tag{24}$$

The entropy $H(X)$ quantifies the amount of uncertainty in the random variable X. Operationally, it represents the ultimate limit for compressing symbols x appearing with probability $p(x)$ (see [21] for additional discussion). For two random variable X and Y with joint probability mass function $p(x, y)$, the joint entropy,

$$H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y), \tag{25}$$

captures the amount of uncertainty in both X and Y taken together as a single random variable. For a given outcome x of X, the probability of obtaining y is $p(y|x) = p(x, y)/p(x)$. This probability mass has entropy,

$$H(Y|X = x) = -\sum_{y} p(y|x) \log_2 p(y|x), \tag{26}$$

representing the uncertainty in Y given x. The average value of the entropy above,

$$H(Y|X) = \sum_{x} p(x) H(Y|X = x), \tag{27}$$

is called the *conditional entropy* of Y given X. Subtracting this conditional entropy from the entropy of Y gives

$$I(X; Y) = H(Y) - H(Y|X) \tag{28}$$

the *mutual information* between Y and X. The mutual information is unchanged when X and Y in (28) are interchanged. For two probability mass functions $p(x)$ and $q(x)$, where $q(x) = 0$ only if $p(x) = 0$, one defines the *relative entropy*,

$$D(p||q) = \sum_x p(x) \log_2 \big(p(x)/q(x)\big), \tag{29}$$

a measure of how different $q(x)$ is from $p(x)$. Operationally, $D(p||q)$ quantifies the penalty of compressing symbols from a distribution $p(x)$, assuming it is $q(x)$(see [21] for additional discussion).

The notion of entropy can be generalized to the quantum world. However, prior to performing such a generalization, we discuss the quantum analog of a probability distribution, called a *density operator* or a *mixed state*. Consider a collection of N unmarked quantum systems, $n_i > 0$ of which are in the quantum state $|\psi_i\rangle$; then a quantum system chosen uniformly at random from this collection is assigned a density operator,

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \tag{30}$$

where $p_i = n_i/N$. By construction, the density operator ρ above is Hermitian, positive semi-definite ($\rho \succeq 0$ or $\langle\phi|\rho|\phi\rangle \geq 0$ for all $|\phi\rangle$), and has unit trace ($\mathrm{Tr}(\rho) = 1$). In general, any positive semi-definite operator with unit trace can be written in the form (30). In this form, if every $|\psi_i\rangle$ is some fixed state $|\chi\rangle$, then $\rho = |\chi\rangle\langle\chi|$ represents a pure quantum state $|\chi\rangle$ and $\rho^2 = |\chi\rangle\langle\chi|.|\chi\rangle\langle\chi| = \langle\chi|\chi\rangle|\chi\rangle\langle\chi| = |\chi\rangle\langle\chi| = \rho$. More generally, ρ represents a mixed state and can be written as

$$\rho = \sum_i \lambda_i |m_i\rangle\langle m_i|, \tag{31}$$

where eigenvalues λ_i are real, positive, and sum to one and $\{|m_i\rangle\}$ are orthonormal kets. Using the form of ρ in (31), one can easily obtain ρ^2 by replacing each λ_i with λ_i^2. Notice $\rho^2 = \rho$ if each $\lambda_i^2 = \lambda_i$, which happens only when $\lambda_j = 1$ for some fixed j and zero for all others— that is, $\rho = |m_j\rangle\langle m_j|$— represents a pure state. Another natural context for using density operators is to describe sub-systems of larger quantum systems. Suppose a quantum system ab, composed of two systems a and b with spaces $\mathcal{H}_a$ and $\mathcal{H}_b$, respectively, is in a pure state $|\chi\rangle \in \mathcal{H}_{ab}$. As mentioned at the end of Sec. 2, this state need not be the product of two pure states, one each on $\mathcal{H}_a$ and $\mathcal{H}_b$. The state of the a and b systems is represented by mixed states with density operators

$$\rho_a = \mathrm{Tr}_b(\rho_{ab}), \quad \text{and} \quad \rho_b = \mathrm{Tr}_a(\rho_{ab}), \tag{32}$$

respectively, where $\rho_{ab} = |\chi\rangle\langle\chi|$, Tr_a is the partial trace over $\mathcal{H}_a$; i.e., $\mathrm{Tr}_a(A \otimes R) = R\mathrm{Tr}(A)$ where $\otimes$ is the tensor product (see discussion containing (17) and (18)) and Tr_b is the partial trace over $\mathcal{H}_b$, defined similarly.

The simplest example of a density operator is a *qubit density* operator. Such density operators can be written in the Bloch parametrization,

$$\rho(\mathbf{r}) = \frac{1}{2}(I + xX + yY + zZ), \tag{33}$$

where the real three-dimensional vector $\mathbf{r} := (x, y, z)$, called the Bloch vector, has magnitude $|\mathbf{r}| = \sqrt{\mathbf{r}.\mathbf{r}}$, at most 1 and

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{34}$$

are the Pauli matrices, written in the standard basis $\{|0\rangle, |1\rangle\}$. Using the Bloch parametrization (33), any qubit density operator can be represented by its Bloch vector in a unit sphere

called the Bloch sphere (see Fig. 1). For instance the density operator $|i\rangle\langle i|$ has Bloch vector $\mathbf{r} = (0, 0, (-1)^i)$ A Bloch vector of unit length, $\mathbf{r}.\mathbf{r} = 1$, represents a pure state qubit density operator, in other words, $\rho(\mathbf{r})^2 = \rho(\mathbf{r})$. If the length of the Bloch vector is less than one, then $\rho(\mathbf{r})$ is a mixed state.

To help understand the Bloch sphere picture, we focus on the density operator in (33). This density operator can be written in the form (31)

$$\rho(\mathbf{r}) = \lambda|m_1\rangle\langle m_1| + (1-\lambda)|m_2\rangle\langle m_2|, \tag{35}$$

where $\lambda = (1 + |\mathbf{r}|)/2$ and $1 - \lambda$ are the eigenvalues of $\rho(\mathbf{r})$. Notice the eigenvalues are non-negative if and only if $|\mathbf{r}| \leq 1$. The eigenvectors $|m_1\rangle$ and $|m_2\rangle$, corresponding to eigenvalues λ and $1 - \lambda$, respectively, are normalized, $\langle m_1|m_1\rangle = \langle m_2|m_2\rangle = 1$, and orthogonal to each other, $\langle m_1|m_2\rangle = 0$, i.e., they represent orthogonal pure states. While one can write explicit expressions for these pure states, it is more useful to focus on the projectors $|m_1\rangle\langle m_1|$ and $|m_2\rangle\langle m_2|$ onto these pure states (see discussion above (31) for additional discussion on projectors). These projectors represent density operators and it is instructive to use the Bloch parametrization (33) to represent them

$$|m_1\rangle\langle m_1| = \rho(\mathbf{r}_1), \quad \text{and} \quad |m_2\rangle\langle m_2| = \rho(\mathbf{r}_2), \tag{36}$$

where $\mathbf{r}_1 := \mathbf{r}/|\mathbf{r}|$ and $\mathbf{r}_2 := -\mathbf{r}/|\mathbf{r}|$ are unit vectors. Using the above equation in (35), we find $\rho(\mathbf{r})$ is the convex combination of two pure state density operators,

$$\rho(\mathbf{r}) = \lambda\rho(\mathbf{r}_1) + (1-\lambda)\rho(\mathbf{r}_2). \tag{37}$$

In addition, one finds that the Bloch vector $\mathbf{r}$ is a convex combination, $\mathbf{r} = \lambda\mathbf{r}_1 + (1-\lambda)\mathbf{r}_2$, of two unit vectors $\mathbf{r}_1$ and $\mathbf{r}_2$. The Bloch sphere picture (see Fig. 1) provides a simple way to visualize all three density operators $\rho(\mathbf{r}), \rho(\mathbf{r}_1)$ and $\rho(\mathbf{r}_2)$.

More generally, two systems a and b, each with density operators ρ_a and ρ_b, respectively, are assigned a joint density operator

$$\rho_{ab} = \rho_a \otimes \rho_b, \tag{38}$$

where $\otimes$ represents tensor product (see discussion containing (17) and (18)).

The quantum analog of the Shannon entropy (24) is the von-Neumann entropy of a density operator ρ (31),

$$S(\rho) := -\mathrm{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log \lambda_i. \tag{39}$$

The inequality above can be derived using the eigen-decomposition (31), where the eigenvalues λ_i of ρ are strictly positive. Using this decomposition, one obtains $\log \rho := \sum_i \log \lambda_i |m_i\rangle\langle m_i|$ by simply replacing the eigenvalues λ_i with their logarithm. Multiplying $\log \rho$ with ρ gives $\rho \log \rho = \sum_i \lambda_i \log \lambda_i |m_i\rangle\langle m_i|$. The trace of this product, multiplied with minus, one gives the right side of the equality (39) For the qubit density operator in (33) the von-Neumann entropy can be easily computed using (35) as

$$S\big(\rho(\mathbf{r})\big) = h(\lambda_+) := -\big(\lambda_+ \log \lambda_+ + (1-\lambda_+)\log(1-\lambda_+)\big), \tag{40}$$

where we use $\lambda_- = 1 - \lambda_+$. Much like the Shannon entropy, the von-Neumann entropy quantifies the amount of uncertainty in the quantum state ρ. In addition, the von-Neumann entropy plays a fundamental role in a vast variety of information processing tasks. For instance, it represents the ultimate rate for compressing quantum states [81].

In practice, quantum systems are susceptible to noise. Prior to describing quantum noise, let us consider classical noise. Classical noise is often modelled by a channel $N : \mathcal{X} \to \mathcal{Y}$,

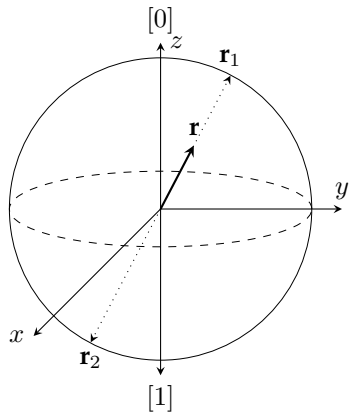

FIGURE 1. Bloch sphere in xyz Cartesian coordinates: The Bloch vector $\mathbf{r}$ represents a qubit density operator $\rho(\mathbf{r})$ (33). The other Bloch vectors, $\mathbf{r}_1$ and $\mathbf{r}_2$ (defined below (36)), have unit length. These vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are antipodes of each other and represent projectors, $|m_1\rangle\langle m_1|$ and $|m_2\rangle\langle m_2|$, respectively. Such rank-1 projectors can be written using a square bracket notation; for instance, $[0] := |0\rangle\langle 0|$.

which maps some input symbol $x \in \mathcal{X}$ to an output symbol $y \in \mathcal{Y}$ with probability $p(y|x)$ (see Sec. 8 for more details). If a channel's input symbol $\mathcal{X}$ arrives with some probability $p(x)$ then the channel maps this input distribution to an output distribution $p(y) = \sum_x p(y|x)p(x)$ over the channel outputs $\mathcal{Y}$.

Consider a simple example of a classical erasure channel E with erasure probability p. The channel's input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} = \{0, 1, e\}$. With probability p, the channel erases the input x by mapping it to an output $y = e$, otherwise with probability $1 - p$ the input is sent perfectly, i.e., the output $y = x$. This erasure channel's conditional probability distribution $p(y|x)$ is given as follows: $p(e|0) = p(e|1) = p, p(0|0) = p(1|1) = (1 - p)$, and $p(1|0) = p(0|1) = 0$. Using this conditional probability distribution, one can find the output distribution

$$p(y) = (1 - p)\, p(x)\delta_{y,x} + p\delta_{y,e}. \tag{41}$$

If $x = i$, where $i \in \{0, 1\}$, then $p(y = i) = 1 - p, p(y = 1 - i) = 0$, and $p(y = e) = p$.

The quantum analog of a channel, called a quantum channel, describes quantum noise [94]. Like its classical counterpart, a quantum channel acting on the quantum analog of a probability distribution, a density operator, maps it to another density operator. In addition, a quantum channel acting on one part of a bi-partite density operator maps the bi-partite density operator to a valid density operator. Mathematically, a quantum channel is a completely positive trace preserving (CPTP) map (see discussion below (45)). One simple example of a quantum channel is an *erasure channel*. Let a be a d_a-dimensional quantum system with space $\mathcal{H}_a = \mathcal{H}$ and let b be a $d_b = (d_a + 1)$-dimensional quantum system with space $\mathcal{H}_b = \mathcal{H} \oplus \mathcal{H}'$, where $\mathcal{H}'$ is spanned by multiples of a single ket $|e\rangle$. Then an erasure channel with erasure probability p, $\mathcal{E}_p : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$, is given by

$$\mathcal{E}_p(\rho) = (1 - p)\rho + p\mathrm{Tr}(\rho)|e\rangle\langle e|, \tag{42}$$

where the channel input ρ is sent perfectly with probability $1 - p$; otherwise, the input is erased with probability p and mapped to a fixed pure state $|e\rangle\langle e|$ orthogonal to ρ. The above equation is akin to (41) where the output probability distribution $p(y)$ was expressed in terms of the input probability distribution $p(x)$ for the classical erasure channel E. The connection with classical erasure can be made even tighter: a classical erasure channel E can emerge from a quantum erasure channel $\mathcal{E}_p$ in the following sense. Suppose a classical

input symbol $x = i$, $i \in \{0, 1\}$ is mapped to a quantum state $[i]$, then sent via $\mathcal{E}_p$, and finally measured using projectors $\{P_0 = [0], P_1 = [1], P_e = [e]\}$ corresponding to measurements outcomes $y \in \{0, 1, e\}$ respectively. Then, using the Born rule (21), the expression for $\mathcal{E}_p$ (42), and the definitions of the projectors $\{P_j\}$, one finds that $p(y|x)$, the probability that the measurement outcome is y given the input symbol is x, can be simply written as $p(y = j|x = i) = \mathrm{Tr}(\mathcal{E}_p([i])[j])$. This conditional probability is exactly the same as the one for the erasure channel E above (41).

Yet another example of a quantum channel is a *qubit depolarizing* channel $\Delta : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$ where $d_a = d_b = 2$,

$$\Delta(\rho) = \lambda\rho + \frac{(1-\lambda)}{2}\mathrm{Tr}(\rho)I, \tag{43}$$

and $-1/3 \leq \lambda \leq 1$. There is no obvious analogy between this quantum depolarizing channel and a classical channel, however the Bloch sphere picture, discussed below (33), provides a helpful way to visualize the action of the depolarizing channel. The depolarizing channel takes its qubit input ρ, with Bloch vector $\mathbf{r}$, to a qubit output $\Delta(\rho)$ with Bloch vector $\lambda\mathbf{r}$. The effect of the channel is to scale the Bloch sphere and make its length smaller (see Fig. 2 for a graphical representation). For values of $\lambda \geq 0$, this channel is often interpreted as sending its input perfectly with probability λ or replacing the input with the maximally mixed state $I/2$. There are a large variety of well-studied quantum channels [40, 106, 108]. Some common ones include the qubit dephasing channel and the (generalized) qubit amplitude damping channel [54, 109].

To describe the action of two separate quantum channels $\mathcal{B}$ and $\mathcal{B}'$ acting on systems a and a', respectively, one uses tensor products in a manner similar to those used for describing pure states and mixed state on two systems (see discussion containing (12) and (38)). Let systems a and a' be acted upon by quantum channels $\mathcal{B} : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$ and $\mathcal{B}' : \mathcal{L}(\mathcal{H}_{a'}) \to \mathcal{L}(\mathcal{H}_{b'})$, respectively, then the channel acting on the joint aa' system is the tensor product channel $\mathcal{B} \otimes \mathcal{B}'$. The tensor product channel is linear, and if ρ_a and $\rho_{a'}$ are density operators of a and a', respectively, then

$$\mathcal{B} \otimes \mathcal{B}'(\rho_a \otimes \rho_{a'}) = \mathcal{B}(\rho_a) \otimes \mathcal{B}'(\rho_{a'}). \tag{44}$$

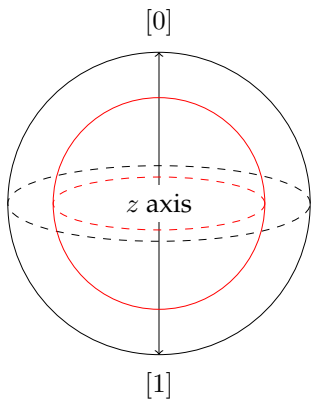

FIGURE 2. Qubit depolarizing channel: The qubit Bloch sphere (in black) is transformed into to an origin centered sphere with a shorter radius (in red) under the action of a qubit depolarizing channel (43).

Remark. When discussing the transmission of classical information across a quantum channel, one arrives at *induced classical channels* (see Ch.20 in [108]). These classical channels

$N : \mathcal{X} \to \mathcal{Y}$ arise out of quantum ones and they model the effective classical noise experienced by classical information encoded and decoded into quantum states passing through a quantum channel $\mathcal{B}$. The *capacity* $C(N)$ of any such induced channel N is obtained by maximizing the mutual information (28) between the channel output Y and input X over all possible input distributions $p(x)$. This capacity is bounded from above by the *Holevo capacity* of the channel $\mathcal{B}$, which represents the ultimate rate at which classical information can be sent across a quantum channel without entanglement at the input [41]. This is not the subject of any of the five SDP problems presented here.

In general, any quantum channel $\mathcal{B} : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$, where $\mathcal{H}_a$ and $\mathcal{H}_b$ have possibly unequal dimensions, can be written using a *Kraus decomposition* [59]

$$\mathcal{B}(A) = \sum_i K_i A K_i^\dagger, \tag{45}$$

where $A \in \mathcal{L}(\mathcal{H}_a)$, and $K_i : \mathcal{H}_a \to \mathcal{H}_b$ are Kraus operators that satisfy the relation

$$\sum_i K_i^\dagger K_i = I_a. \tag{46}$$

For an erasure channel of the form (42) acting on qubits inputs, the Kraus operators can be written as matrices

$$K_1 = \sqrt{1-p}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad K_2 = \sqrt{p}\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad K_3 = \sqrt{p}\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{47}$$

using the standard basis $\{|0\rangle, |1\rangle\}$ at the input and the basis $\{|0\rangle, |1\rangle, |e\rangle\}$ at the output. A simple calculation shows that the matrices above satisfy (46) with I_a as the 2×2 identity matrix I_2. The qubit depolarizing channel Δ (43) can also be written in the Kraus form (45) where

$$K_1 = \sqrt{1-p}I_2, \quad K_2 = \sqrt{p/3}X, \quad K_3 = \sqrt{p/3}Y, \quad K_4 = \sqrt{p/3}Z, \tag{48}$$

and $p = 3(1-\lambda)/4$. From this Kraus form, one may view $\Delta(\rho)$ as a channel that applies each of the Pauli errors X, Y, and Z with equal probability $p/3$ and applies the identity map with probability $1-p$. Using standard matrix multiplication, or the property $X^\dagger X = Y^\dagger Y = Z^\dagger Z = I_2$, one can check that the operators in (48) also satisfy the equality in (46). This equality (46) ensures that $\mathcal{B}$ is trace preserving; that is, $\mathrm{Tr}(\mathcal{B}(A)) = \mathrm{Tr}(A)$ for any operator $A \in \mathcal{L}(\mathcal{H}_a)$. Together, (46) and the Kraus decomposition (45) ensure that $\mathcal{B}$ is a completely positive trace preserving map. While a positive trace preserving map is one that maps positive semi-definite operators to positive semi-definite operators of the same trace, a CPTP map satisfies a stronger condition: for all positive semi-definite operators Λ_{ar} and any finite-dimensional space $\mathcal{H}_r$ with dimension d_r, the operator

$$\Gamma_{br} = (\mathcal{B} \otimes \mathcal{I})\Lambda_{ar}, \tag{49}$$

is positive semi-definite and has the same trace as Λ_{ar}, where $\mathcal{I}$ is the identity channel taking $\mathcal{L}(\mathcal{H}_r)$ to itself and $\mathcal{B} \otimes \mathcal{I}$ is a tensor product of two channels $\mathcal{B}$ and $\mathcal{I}$ (see discussion containing (44)) . Turns out this stronger condition is satisfied if and only if at $\mathcal{H}_r = \mathcal{H}_a$ and $\Lambda_{aa} = |\gamma\rangle\langle\gamma|$, where $|\gamma\rangle = \sum_i |i\rangle_a \otimes |i\rangle_a$ is an *unnormalized* maximally entangled state across $\mathcal{H}_a \otimes \mathcal{H}_a$, the operator Γ_{ba} is positive semi-definite and its partial trace over b is the identity on $\mathcal{H}_a$, i.e., $\mathrm{Tr}_b(\Gamma_{br}) = I_a$ [20, 48]. For this reason, an operator mapping $\mathcal{H}_{ba} := \mathcal{H}_b \otimes \mathcal{H}_a$ to itself,

$$\mathcal{J}_{ba}(\mathcal{B}) := (\mathcal{B} \otimes \mathcal{I})|\gamma\rangle\langle\gamma|, \tag{50}$$

is sometimes called the Choi-Jamiołkowski representation of the channel $\mathcal{B}$. Given two quantum channels $\mathcal{B} : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$ and $\mathcal{B}' : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$, their linear combination, $\mathcal{S} = c_0\mathcal{B} + c_1\mathcal{B}'$, $c_0, c_0 \in \mathbb{C}$, is a linear map from $\mathcal{L}(\mathcal{H}_a)$ to $\mathcal{L}(\mathcal{H}_b)$, however this map need not be a quantum channel. In general, a linear map $\mathcal{S} : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$ from operators on $\mathcal{H}_a$ to operators on $\mathcal{H}_b$ is called a *superoperator*. This superoperator represents a CPTP map if and only if its Choi-Jamiołkowski representation, $\mathcal{J}_{ba}(\mathcal{S})$, is positive semi-definite and $\mathrm{Tr}_b(\mathcal{J}_{ba}(\mathcal{S})) = I_a$.

The Kraus form (45) ensures that a map is CPTP, but it can be also used to interpret properties of a quantum channel. For instance, if each Kraus operator in (45) has rank 1, then $\mathcal{B}$ becomes an *entanglement breaking* channel [43]. Such channels have the property that any entangled channel input ρ_{ra} is mapped to an unentangled output (50) (see Sec. 7 for additional discussion about entanglement). One simple example of an entanglement breaking channel is a qubit channel $\mathcal{E}$ with two Kraus operators:

$$K_1 = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad K_2 = |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{51}$$

Each operator above has rank 1. To express $\mathcal{E}$ using these operators in (51), one uses and simplifies an equation of the form (45) to obtain

$$\mathcal{E}(\rho) = |0\rangle\langle 0|\mathrm{Tr}(\rho|0\rangle\langle 0|) + |1\rangle\langle 1|\mathrm{Tr}(\rho|1\rangle\langle 1|). \tag{52}$$

4. Problem 1: Quantum State Discrimination

The concept of projective measurements, discussed in Sec. 2.2, and quantum channels, discussed at the end of Sec. 3, can be combined to obtain a more general measurement scheme described mathematically as a positive operator value measure (POVM). Consider a quantum channel $\mathcal{B} : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b) \otimes \mathcal{L}(\mathcal{H}_e)$, of the form

$$\mathcal{B}(\rho) = \sum_i K_i \rho K_i^\dagger \otimes |i\rangle\langle i|, \tag{53}$$

where $\otimes$ represents tensor product (see discussion containing (17) and (18)). Using Born's rule, a projective measurement on one half of the channel output $\mathcal{H}_e$ in the computational basis of $\mathcal{H}_e$ results in an outcome i with probability

$$p_i = \mathrm{Tr}\big(\mathcal{B}(\rho)(I_b \otimes |i\rangle\langle i|)\big). \tag{54}$$

Using standard linear algebra along with the definition $E_i := K_i^\dagger K_i$, we obtain

$$p_i = \mathrm{Tr}(\rho E_i). \tag{55}$$

The collection of operators $\{E_i\}$ are called a POVM. These operators are positive semi-definite and sum to the identity I_a; that is,

$$E_j = E_j^\dagger \succeq 0, \quad \text{and} \quad \sum_i E_i = I_a. \tag{56}$$

Any general measurement on a quantum system a can be described using a POVM $\{E_i\}$. Associated with each E_i is a measurement outcome i which occurs with probability (55).

A general setup for the quantum state discrimination problem can be obtained as follows. Suppose a random variable X takes one of n values i with probability p_i. When $X = i$, a d-dimensional quantum state σ_i is prepared. The key task in quantum state discrimination

is to measure the prepared state and predict i with high probability. If the random variable Y predicts X, then we wish to maximize the success probability:

$$p_s := \sum_i p_i \Pr(Y = i | X = i). \tag{57}$$

Suppose $\{E_j\}$ is a POVM that describes the measurement; then the conditional probability

$$\Pr(Y = j | X = i) = \mathrm{Tr}(E_j \sigma_i). \tag{58}$$

Using the above equation, the maximum success probability (57) over all POVMs is obtained as the optimum value p_s^* of the semi-definite program,

$$\begin{aligned} \text{maximize } & \sum_i p_i \mathrm{Tr}(E_i \sigma_i) \\ \text{subject to } & E_j \succeq 0, \qquad \forall 1 \le j \le n, \\ \text{and } & \sum_{j=1}^{n} E_j = I. \end{aligned} \tag{59}$$

Consider a simple case when $n = d = 2$, $p_1 = q, p_2 = 1 - q$,

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{60}$$

In this case, the SDP in (59) admits an algebraic solution, $E_1 = \sigma_1, E_2 = I - E_1$ and $p_s = 1$, which is independent of p. Such algebraic solutions exist for any $d \ge 2$ and $2 \le n \le d$ when

$$\mathrm{Tr}(\sigma_i \sigma_j) = \delta_{ij} \tag{61}$$

for all $1 \le i, j \le n$; that is, when σ_i are pairwise orthogonal to one another. In such cases $E_i = \sigma_i$ for all $1 \le i \ne n - 1$ and $E_n = I - \sum_i E_i$ is a solution to (59) with optimum value $p_s^* = 1$, independent of p_i. An interesting case where the SDP in (59) can be solved algebraically is $n = 2$ and arbitrary d. In this case, the SDP in (59) reduces to

$$\begin{aligned} \text{maximize } & \frac{1}{2}(1 + \mathrm{Tr}\big(F(p_1 \sigma_1 - p_2 \sigma_2)\big) \\ \text{subject to } & -I \preceq F \preceq I, \end{aligned} \tag{62}$$

where we have introduced the operator $F := E_1 - E_2$. The above SDP has an optimum value

$$p^* = \frac{1}{2}(1 + ||p_1 \sigma_1 - p_2 \sigma_2||_1), \tag{63}$$

where $||A||_1 = \mathrm{Tr}\sqrt{A^\dagger A}$ is the operator 1-norm, also called the nuclear norm, and $\sqrt{A^\dagger A}$, square root of a positive semi-definite matrix, is an operator obtained by replacing the eigenvalues of $A^\dagger A$ with their square root. The optimum value is often called the Helstrom bound [39]. In general, the SDP in (59) cannot be solved analytically; however, one can use numerical SDP solvers. For using such numerical solvers, we reformulate the SDP in (59) as follows

$$\begin{aligned} \text{maximize } & \sum_{i=1}^{n-1} p_i \mathrm{Tr}(E_i \sigma_i) + (1 - \sum_{i=1}^{n-1} p_i) \mathrm{Tr}\big((I - \sum_{j=1}^{n-1} E_j)\sigma_n\big) \\ \text{subject to } & I - \sum_{j=1}^{n-1} E_j \succeq 0, \\ \text{and } & E_j \succeq 0, \end{aligned} \tag{64}$$

where $1 \leq j \leq n-1$. Numerical solutions to the SDP above can be obtained using solvers in open-source packages. For instance, one may use a Python interface, PICOS [79], and an open-source solver, CVXOPT [3]. Using these numerical tools we formulate the SDP above with $n = d = 2$, $p \in [0,1]$ chosen randomly, and σ_i defined in (60). We find almost perfect agreement between the numerically obtained objective value p_s^n and the algebraic value $p^*s = 1$ stated below (61). In other examples with fixed $n = 2, d = 6$, and randomly chosen $p \in [0,1]$, σ_1, and σ_2, we find good agreement, $|p_s^n - p_s*| \simeq O(10^{-9})$, between the numerical value p_s^n and the true value, p_s*, computed using (63). These and additional examples are available along with this chapter (see Notebook 1 in [90]).

Discrimination of quantum states, and quantum hypothesis testing, is a vast [5,38,39,51, 56,69,72] and active sub-field of quantum information science. Here, we have touched the surface of this field by introducing certain special cases. For solving these special cases, we illustrate the use of open-source numerical tools. The interested reader may find a variety of other resources and open problems in these reviews [6,7,19] and references therein.

5. Problem 2: Quantum State Fidelity

Classical objects of different types are perfectly distinguishable. On the other hand, quantum objects in two different quantum states are not always perfectly distinguishable. This motivates a basic question, given two different states of a quantum system: how similar are these states to each other? One measure of similarity between quantum states is *fidelity*. Consider pure states $|\psi\rangle$ and $|\phi\rangle$; these are simply unit vectors in some space $\mathcal{H}_a$. The magnitude of the overlap between these vectors,

$$F(\psi, \phi) = |\langle\psi|\phi\rangle|, \tag{65}$$

is defined as the fidelity between the pure states $|\psi\rangle$ and $|\phi\rangle$. When the pure states $|\psi\rangle$ and $|\phi\rangle$ are the same, their fidelity $F(\psi,\phi)$ is one; when $|\psi\rangle$ and $|\phi\rangle$ are orthogonal, $F(\psi,\phi) = 0$, and in general, $0 \leq F(\psi,\phi) \leq 1$.

While fidelity between pure states is straightforward to define, quantum systems cannot always be described by pure states. In general, quantum systems are described by mixed states. However, a quantum system a in some mixed state ρ can always be viewed as a sub-system of two systems a and r in some pure state $|\psi\rangle$. More precisely, let $\mathcal{H}_a$ and $\mathcal{H}_r$ describe the spaces of a and r, respectively, and then any ρ describing a can be obtained as

$$\rho = \mathrm{Tr}_r(|\psi\rangle\langle\psi|), \tag{66}$$

where $|\psi\rangle \in \mathcal{H}_{ar}$. The state $|\psi\rangle$ above is called a purification of ρ and r, the purifying system. If ρ has a spectral decomposition $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$ then its purification has the form $|\psi\rangle = \sum_i \sqrt{\lambda_i}|e_i\rangle \otimes U|e_i\rangle$ where U is any unitary matrix on $\mathcal{H}_r$ and $\otimes$ represents tensor product (see discussion containing (12) and (13)). Clearly, every choice of U gives a different purification of the same state ρ and two different purifications are related by a unitary on the purifying system $\mathcal{H}_r$. To define the fidelity between two, possibly mixed states ρ and σ of a system a, we can consider the maximal fidelity between their purifications,

$$F(\rho,\sigma) = \max_{\psi,\phi} |\langle\psi|\phi\rangle|, \tag{67}$$

where $|\psi\rangle$ and $|\phi\rangle$ purify ρ and σ respectively, using the same purifying system r. Uhlmann's theorem [97] (see Th.9.2.1 in [108] for a short proof) shows that the fidelity defined above reduces to the simple form

$$F(\rho,\sigma) = ||\sqrt{\rho}\sqrt{\sigma}||_1. \tag{68}$$

When ρ and σ are pure states $|\psi\rangle\langle\psi|$ and $|\phi\rangle\langle\phi|$, respectively, the fidelity expression (68) simply reduces to (65). When ρ is a mixed state but $\sigma = |\psi\rangle\langle\psi|$ is a pure state then the fidelity (68) is simply

$$F(\rho,\sigma) = \langle\psi|\rho|\psi\rangle. \tag{69}$$

Suppose ρ and σ are mixed states that are diagonal in the same basis $|x\rangle$. It is convenient to write these states as follow

$$\rho = \sum_x p(x)|x\rangle\langle x|, \quad \text{and} \quad \sigma = \sum_x q(x)|x\rangle\langle x| \tag{70}$$

where $p(x)$ and $q(x)$ are probability distributions. The fidelity between these diagonal operators above

$$F(\rho,\sigma) = \sum_x \sqrt{p(x)}\sqrt{q(x)}, \tag{71}$$

is simply the Bhattacharya overlap [16] between the classical probability distributions $p(x)$ and $q(x)$.

Fidelity $F(\rho,\sigma)$ (68) is given by the optimum value of these primal and dual semi-definite programs [57, 105],

Primal: (72) Dual: (73)

$$\begin{aligned} &\text{maximize } \frac{1}{2}\mathrm{Tr}(\Lambda+\Lambda^\dagger) && \text{minimize } \frac{1}{2}\big(\mathrm{Tr}(\rho Y)+\mathrm{Tr}(\sigma Z)\big)\\ &\text{subject to } \begin{pmatrix}\rho & \Lambda\\ \Lambda^\dagger & \sigma\end{pmatrix} \succeq 0, && \text{subject to } \begin{pmatrix} Y & -I\\ -I & Z\end{pmatrix} \succeq 0.\\ &\quad \text{and } \Lambda \in \mathcal{L}(\mathcal{H}_a); \end{aligned}$$

The primal and dual SDPs can be solved numerically. When ρ and σ are randomly chosen 4-dimensional pure states we find optimum values F_p and F_d, for the primal and dual SDP objectives respectively. These are in good agreement with each other, and with F computed using (65). In particular, the maximum absolute difference between any pair of these three values is $O(10^{-9})$. This absolute difference remains typically small, $O(10^{-5})$, when ρ is a random mixed state and σ is a random pure state, each 3-dimensional. To compute this difference numerically we solve the primal and dual SDP above and find F using (69). A similar computation, using 10-dimensional mixed states ρ and σ chosen randomly and F computed using (71), shows good numerical agreement. Typically the maximum pairwise difference between all three values F_p, F_d, and F, is $O(10^{-7})$. A short tutorial helping perform these computations is available along with this chapter (see Notebook 2 in [90]).

While the fidelity function (68) can be computed using standard numerical algebra libraries, the SDP formulation (73) for computing fidelity has additional utilities. One utility is the use of the formulation (73) to show that a variety of generalizations of the fidelity function [34, 53, 111, 112] can also be computed efficiently via an SDP. Some of these generalizations play a useful rule in security analysis of quantum protocols. Another utility of the SDP formulation is to the quantum channel discrimination problem discussed next.

6. Problem 3: Quantum Channel Discrimination

Consider a protocol with two parties, Alice and Bob, where Alice prepares a quantum state ρ_a and hands it to Bob. Upon receiving ρ_a, Bob generates a random variable X, which takes the value 1 with probability t and 2 with probability $1-t$. When $X=i$, Bob applies the channel $\mathcal{B}_i : \hat{\mathcal{H}}_a \to \hat{\mathcal{H}}_b$ and obtains a state $\mathcal{B}_i(\rho_a)$. This new state is returned to Alice, whose task is to measure it and correctly predict i. Alice knows a description of each fixed channel $\mathcal{B}_i$ and controls the state ρ_a, but is unaware of the random value i. By varying ρ_a the maximum probability with which Alice can correctly predict i by measuring $\mathcal{B}_i(\rho_a)$ is

$$q^* = \frac{1}{2}(1+\max_{\rho_a} ||t\mathcal{B}_1(\rho_a) - (1-t)\mathcal{B}_2(\rho_a)||_1). \tag{74}$$

Using

$$\mathcal{D} := t\mathcal{B}_1 - (1-t)\mathcal{B}_2, \tag{75}$$

one may rewrite the above expression,

$$q^* = \frac{1}{2}(1 + \max_{\rho_a} ||\mathcal{D}(\rho_a)||_1). \tag{76}$$

One can show that (see discussion below Def. 3.37 in [106])

$$\max_{\rho_a} ||\mathcal{D}(\rho_a)||_1 = ||\mathcal{D}||_1, \tag{77}$$

where the 1-norm $||\mathcal{D}||_1$ of the map $\mathcal{D}$ is the maximum 1-norm of the operator $||\mathcal{D}(X)||_1$ where $||X||_1 \leq 1$.

In the protocol above, instead of preparing a state ρ_a, Alice can prepare a (possibly entangled) state ρ_{ar} on $\mathcal{H}_a \otimes \mathcal{H}_r$, tensor product of $\mathcal{H}_a$ with $\mathcal{H}_r$ (see discussion containing (12) and (38)), a space of some auxiliary system with dimension d_r of Alice's choosing. Alice can then send the a sub-system of ρ_{ar} to Bob. Bob applies the map $\mathcal{B}_i$ to this sub-system and returns the sub-system to Alice. The final state with Alice is given by $\mathcal{B}_i \otimes \mathcal{I}_r(\rho_{ar})$, where $\mathcal{I}_r$ is the identity channel on $\mathcal{H}_r$. By varying ρ_{ar} and d_r, the maximum probability with which Alice can correctly predict i by measuring $\mathcal{B}_i \otimes \mathcal{I}_r(\rho_{ar})$ is

$$s^* = \frac{1}{2}(1 + \sup_{\rho_{ar}, d_r} ||\mathcal{D} \otimes \mathcal{I}_r(\rho_{ar})||_1). \tag{78}$$

One can show that (see discussion below Def. 3.43 in [106])

$$\sup_{\rho_{ar}, d_r} ||\mathcal{D} \otimes \mathcal{I}_r(\rho_{ar})||_1 = ||\mathcal{D}||_\diamond, \tag{79}$$

where the diamond norm (also called the completely bounded trace norm) $||\mathcal{D}||_\diamond := ||\mathcal{D} \otimes \mathcal{I}_a||_1$. The equality above demonstrates that the dimension of the auxiliary system $\mathcal{H}_r$ chosen by Alice does not need to be larger than d_a, the dimension of system a being sent from Alice to Bob. In addition, since $1/2 \leq s^* \leq 1$, $||\mathcal{D}||_\diamond$ for any $\mathcal{D}$ of the form (75) is at most 1.

The probabilities q and s, defined in (76) and (78), respectively, can be obtained from the 1-norm and the diamond norm of the map $\mathcal{D}$ defined in (75). As mentioned below (79) and (77), for any map $\mathcal{D} : \hat{\mathcal{H}}_a \to \hat{\mathcal{H}}_b$, both the 1-norm $||\mathcal{D}||_1$ and the diamond norm $||\mathcal{D}||_\diamond$ are defined as the maximum value of some convex function. Computing the maximum of a convex function is a non-trivial problem. However, in the special case of the diamond norm, such a maximization can be reframed as a semi-definite program. Consider a linear superoperator $\mathcal{D} : \hat{\mathcal{H}}_a \to \hat{\mathcal{H}}_b$ with Choi-Jamiołkowski representation $\mathcal{J}_{ba}(\mathcal{D})$ (for definition, see (50)). The diamond norm of $\mathcal{D}$ is the optimal value of the

Primal SDP: (80)

$$\begin{aligned}
\text{maximize } & \frac{1}{2}\left(\mathrm{Tr}(\mathcal{J}_{ba}(\mathcal{D})X) + \mathrm{Tr}(\mathcal{J}_{ba}(\mathcal{D})X)^*\right) \\
\text{subject to } & \begin{pmatrix} I_b \otimes \rho_a & X \\ X^\dagger & I_b \otimes \sigma_a \end{pmatrix} \succeq 0, \\
& \mathrm{Tr}(\rho_a) = 1, \\
& \mathrm{Tr}(\sigma_a) = 1, \\
\text{and } & X \in \mathcal{L}(\mathcal{H}_a),
\end{aligned}$$

and

Dual SDP: (81)

$$\begin{aligned}&\text{minimize } \frac{1}{2}\big(||\mathrm{Tr}_b(N_{ab})||_\infty + ||\mathrm{Tr}_b(M_{ab})||_\infty\big)\\ &\text{subject to } \begin{pmatrix} N_{ba} & -\mathcal{J}_{ba}(\mathcal{D}) \\ -\mathcal{J}_{ba}(\mathcal{D})^\dagger & M_{ba}\end{pmatrix} \succeq 0,\end{aligned}$$

where $||A||_\infty$ denotes the infinity norm, also called the spectral norm of the operator $A \in \hat{\mathcal{H}}$ [104].

This norm is dual to the nuclear norm. It is the minimum real μ for which $A \preceq \mu I$. Using this characterization of the infinity norm, the dual semi-definite program above can be re-written as

$$\begin{aligned}&\text{minimize } \frac{1}{2}(\mu+\nu)\\ &\text{subject to } \begin{pmatrix} N_{ba} & -\mathcal{J}_{ba}(\mathcal{D}) \\ -\mathcal{J}_{ba}(\mathcal{D})^\dagger & M_{ba}\end{pmatrix} \succeq 0,\\ &\qquad \mathrm{Tr}_b(N_{ba}) \preceq \mu I_a,\\ &\qquad \mathrm{Tr}_b(M_{ba}) \preceq \nu I_a,\\ &\text{and } \mu,\nu \in \mathbb{R} \qquad .\end{aligned}$$

These SDPs above can be derived using a connection between the fidelity function and the diamond norm. The connection, together with a simplification of the SDP (73) for the fidelity function, can be used to arrive at the above SDP (for details see [105]).

Let us consider some simple examples where the diamond norm can be computed algebraically in closed form. Our first example is the diamond norm of any quantum channel $\mathcal{B} : \hat{\mathcal{H}}_a \to \hat{\mathcal{H}}_b$. Using the definition of the diamond norm $||\mathcal{B}||_\diamond$, one can easily show that any quantum channel has diamond norm one. Our second example considers the protocol between Alice and Bob discussed at the beginning of this section. In that protocol, suppose $d_a = d_b = 2$; in other words, Alice and Bob exchange qubit states. In addition, consider a concrete case where $\mathcal{B}_1$ is simply the identity channel $\mathcal{I}$ and $\mathcal{B}_2$ is the qubit depolarizing channel Δ in (43), and each is equally probable, i.e., $t = 1/2$; thus $\mathcal{D}$ in (75) takes the form

$$\mathcal{D}(\rho) = \frac{1}{2}\big(\mathcal{B}_1(\rho) - \mathcal{B}_2(\rho)\big) = \frac{(1-\lambda)}{2}(\rho - \mathrm{Tr}(\rho)\frac{I}{2}). \tag{82}$$

Using the form of $\mathcal{D}$ in (82), we find (see App. A for details)

$$||\mathcal{D}||_1 = (1-\lambda)/2, \quad \text{and} \quad ||\mathcal{D}||_\diamond = 3(1-\lambda)/4. \tag{83}$$

The norms above, along with equations (76) and (78), give

$$q^* = (3-\lambda)/4 \quad \text{and} \quad s^* = (7-3\lambda)/8 > q^*. \tag{84}$$

Since s^* is larger than q^*, we find that Alice can increase the probability of correctly distinguishing Δ from $\mathcal{I}$ using a qubit auxiliary system. This increase comes from the possibility of Alice sending to Bob one half of a joint system, in state ρ_{aa}. Later, system b returned by Bob to Alice results in a joint state ρ_{ba} which Alice may measure jointly.

Consider an extreme case where $\lambda = -1/3$ [103]. Suppose Alice prepares a pure entangled state, $\rho_{aa} = |\psi^+\rangle\langle\psi^+|$, and

$$|\psi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{85}$$

One-half of this entangled state is sent to Bob. If Bob applies the identity channel $\mathcal{I}$, the state ρ_{ba} remains unchanged. If Bob applies Δ, the state ρ_{ba} becomes orthogonal to $|\psi^+\rangle\langle\psi^+|$. This orthogonal state can be distinguished perfectly by doing a joint measurement on ba (see discussion accompanying (61)). As a result, we get $s^* = 1$. Without using such

entangled inputs, the maximum probability of distinguishing $\mathcal{I}$ from Δ is $q^* = 5/6$. We test these findings numerically. In particular, we compute the diamond norm of $\mathcal{D}$ in (82) by solving the primal and dual SDPs, (80) and (81). The optimum value of both SDPs is in good agreement, and their difference is typically $O(10^{-10})$. Using this value, we obtain the numerical estimate s_n^* (78) of s^* (84); their absolute difference is typically small $O(10^{-11})$ too. Comparing s_n^* with q^* reveals s_n^* is larger and thus entanglement helps discriminate quantum channels $\mathcal{I}$ and Δ.

Remark. The above example can be generalized to a d-dimensional quantum system called a *qudit*. In this higher-dimensional case, $\mathcal{B}_1$ is the qudit identity channel $\mathcal{I}_d$ and $\mathcal{B}_2$ is the qudit depolarizing channel,

$$\Delta_d(\rho) = \lambda\rho + (1-\lambda)\mathrm{Tr}(\rho)\frac{I}{d}. \tag{86}$$

In (75) setting $t = 1/2, \mathcal{B}_1 = \mathcal{I}_d$ and $\mathcal{B}_2 = \Delta_d(\rho)$ results in a superoperator

$$\mathcal{D}(\rho) = \frac{1-\lambda}{2}(\rho - \mathrm{Tr}(\rho)\frac{I}{d}), \tag{87}$$

whose 1 norm and diamond norm are (see App. A for details)

$$||\mathcal{D}||_1 = (1-\lambda)\frac{d-1}{d}, \quad \text{and} \quad ||\mathcal{D}||_\diamond = (1-\lambda)\frac{d^2-1}{d^2}, \tag{88}$$

respectively. Using the above equations, along with (76) and (78), one finds that $s^* > q^*$ and the difference $s^* - q^*$ scales as $O(1/d)$.

In our third channel discrimination example (based on Ex. 3.36 in [106]), we again consider the protocol between Alice and Bob discussed at the beginning of this section. In this protocol Alice and Bob exchange qudits. With equal probability, i.e., $t = 1/2$, Bob chooses one of two Werner-Holevo channels

$$\mathcal{B}_1(\rho) = \frac{1}{d+1}\big(\mathrm{Tr}(\rho)I + \rho^T\big) \quad \text{or} \quad \mathcal{B}_2(\rho) = \frac{1}{d-1}\big(\mathrm{Tr}(\rho)I - \rho^T\big), \tag{89}$$

where the transpose is done in the standard basis. Channels $\mathcal{B}_1$ and $\mathcal{B}_2$ have Choi-Jamiołkowski representations,

$$\mathcal{J}_{ba}(\mathcal{B}_1) = \frac{1}{d+1}(I_b \otimes I_a + S_{ba}), \quad \text{and} \quad \mathcal{J}_{ba}(\mathcal{B}_2) = \frac{1}{d-1}(I_b \otimes I_a - S_{ba}), \tag{90}$$

respectively, where

$$S_{ba}|i\rangle \otimes |j\rangle = |j\rangle \otimes |i\rangle \tag{91}$$

is the swap operator acting on $\mathcal{H}_b \otimes \mathcal{H}_a$ (spaces $\mathcal{H}_a$ and $\mathcal{H}_b$ have equal dimension d). The operators $(I_b \otimes I_a + S_{ba})/2$ and $(I_b \otimes I_a - S_{ba})/2$ equal the projector onto the symmetric and anti-symmetric sub-spaces of $\mathcal{H}_b \otimes \mathcal{H}_a$, respectively. For these Werner-Holevo channels, $\mathcal{D}$ in (75) takes the form

$$\mathcal{D}(\rho) = \frac{d}{d^2-1}\big(\rho^T - \mathrm{Tr}(\rho)\frac{I}{d}\big). \tag{92}$$

The operator norm $||\mathcal{D}||_1 = 2/(d+1)$; using (74), we get

$$q^* = \frac{1}{2} + \frac{1}{d+1}. \tag{93}$$

As stated below (78), $||\mathcal{D}||_\diamond$ is at most 1 and equals the maximum value of $||\mathcal{D} \otimes \mathcal{I}(X)||_1$ where $||X||_1 \leq 1$. Let X be a projector onto the maximally entangled states on two qudits,

$$|\phi_d\rangle := \frac{1}{\sqrt{d}}\sum_i |i\rangle_a \otimes |i\rangle_a, \tag{94}$$

then

$$||\mathcal{D}\otimes\mathcal{I}(X)||_1 = ||\frac{1}{d}\big(\mathcal{J}(\mathcal{B}_1)_{ba} - \mathcal{J}(\mathcal{B}_2)_{ba}\big)|| = 1. \tag{95}$$

As a result, $||\mathcal{D}||_\diamond = 1$. From (78) and (79) it follows that

$$s^* = 1. \tag{96}$$

A numerical estimate s_n^* of s^* is found by directly using the definition of s^* (78) and by solving the primal and dual SDPs, (80) and (81) with $\mathcal{D}$ in (92) and $d = 3$. This numerical estimate is in reasonably good agreement with the value of s^* stated above.

Notice that for any d, there is a gap between s^* above and q^* in (93) which is at least $1/2$ and scales as $O(1/d)$. The strategy that allows us to obtain $s^* = 1$ makes use of entanglement. In particular the maximally entangled state $|\phi_d\rangle\langle\phi_d|$ (94) can be sent by Alice to Bob. If Bob applies $\mathcal{B}_1$ then Bob's joint state with the auxilliary system is $\mathcal{J}_{ba}(\mathcal{B}_1)/\mathrm{Tr}(\mathcal{J}_{ba}(\mathcal{B}_1))$, otherwise Bob applies $\mathcal{B}_2$, creating the joint state $\mathcal{J}_{ba}(\mathcal{B}_2)/\mathrm{Tr}(\mathcal{J}_{ba}(\mathcal{B}_2))$. Notice these two joint states are orthogonal to each other, and thus can be distinguished perfectly by doing a joint measurement on $\mathcal{H}_{ba}$ (see discussion accompanying (61)). Notice that measuring the joint system is necessary, if Alice throws away the reference system a and keeps only b; then the states received by Alice from Bob would be identical, regardless of the channels $\mathcal{B}_1$ and $\mathcal{B}_2$ applied by Bob.

In the special cases discussed above, we provided algebraic expressions for the diamond norm. When such algebraic expressions are not available, one needs to use numerical techniques to a solve semi-definite program and obtain the diamond norm. A short tutorial on solving these semi-definite programs using Python packages PICOS and solvers in CVXOPT/MOSEK is available along with this chapter (see Notebook 3 in [90]).

The SDP formulation of the diamond norm has several virtues. It shows that a variety of bounds on the ability of quantum channels to send information can be computed efficiently [42, 100]. Not only bounds, one can use the SDP formulation to obtain computable measures of entanglement [102], a non-trivial problem in entanglement theory. The SDP formulation has also aided the use of diamond norm in quantum error correction [47, 64] and compressed sensing [58].

7. Problem 4: Quantum Entanglement and Separability

Entanglement is a patently non-classical aspect of quantum states. Any bipartite quantum state on $\mathcal{H}_a\otimes\mathcal{H}_b$ is said to be entangled if the state's density operator ρ_{ab} cannot be expressed as a convex combination,

$$\rho_{ab} = \sum_i p_i|\psi_i\rangle\langle\psi_i|\otimes|\phi_i\rangle\langle\phi_i|, \tag{97}$$

of pure product states $|\psi_i\rangle\otimes|\phi_i\rangle$ on the tensor product (see discussion containing (38) for the definition) of $\mathcal{H}_a$ and $\mathcal{H}_b$; here $p_i \geq 0$ and $\sum_i p_i = 1$. A quantum state with a density operator of the form in (97) is called a separable state. In what follows, we discuss criteria for the separability of quantum states. When ρ_{ab} is a pure state, i.e., $\rho_{ab}^2 = \rho_{ab}$, then criteria for separability of ρ_{ab} is relatively easy (see discussion at the end of Sec. 2.1). To discuss a criterion of separability that applies more generally, we need a notion called the *extension* of a density operator.

Consider a density operator ρ_{ab} on $\mathcal{H}_{ab}$. For any integer $n \geq 1$, let $\mathcal{H}_{bk}$ be a space of a dimension equal to that of $\mathcal{H}_b$ and

$$\mathcal{H}_B := \bigotimes_{j=1}^{k}\mathcal{H}_{bj}, \tag{98}$$

where $\bigotimes$ is our notation for tensor product (defined in the discussion containing (12) and (13)) of several spaces $\mathcal{H}_{b1}, \mathcal{H}_{b2}, \cdots, \mathcal{H}_{bk}$. For $k = 1$, $\mathcal{H}_B = \mathcal{H}_{b1}$ and a density operator ρ_{aB} on $\mathcal{H}_{aB}$ is an extension of ρ_{ab} if $\rho_{aB} = \rho_{ab}$. For $k > 1$, ρ_{aB} is called an extension of ρ_{ab} if the partial trace ρ_{ab1} of ρ_{aB} over $\mathcal{H}_{b2} \otimes \mathcal{H}_{b3} \otimes \cdots \otimes \mathcal{H}_{bk}$ is ρ_{ab}; in other words,

$$\rho_{ab} = \rho_{ab1} = \mathrm{Tr}_{b2,b3,\ldots,bn}(\rho_{aB}). \tag{99}$$

An extension ρ_{aB} of ρ_{ab} is called symmetric if swapping any space $\mathcal{H}_{bi}$ with $\mathcal{H}_{bj}$ in $\mathcal{H}_B$, where $1 \leq i < j \leq n$, has no effect on ρ_{aB}; that is,

$$\rho_{aB} = \Pi_{bibj} \rho_{aB} \Pi_{bibj}, \tag{100}$$

where the swap operator Π_{bibj} on $\mathcal{H}_a \otimes \mathcal{H}_B$ simply extends the usual swap operator S_{bibj} (91) on $\mathcal{H}_{bi} \otimes \mathcal{H}_{bj}$ to $\mathcal{H}_a \otimes \mathcal{H}_B$. This extension applies S_{bibj} to $\mathcal{H}_{bi} \otimes \mathcal{H}_{bj}$ and the identity to all other spaces.

A quantum state ρ_{ab} is separable if and only if it has a symmetric extension for all $k \geq 1$ [29, 75, 107]. Further, one can show that if ρ_{ab} has a symmetric extension for some k, then it has a symmetric extension for all $k' > k$. These powerful symmetric extension results provide a straightforward route to check if a given density operator ρ_{ab} is separable or entangled. This route is to pick an integer $k \geq 1$ and frame a constraint satisfaction problem which simply checks if the linear constraints (99) and (100) can be satisfied by a unit-trace positive semi-definite operator ρ_{aB}. For each k, this problem can also be framed as a SDP. If this SDP is infeasible, then ρ_{ab} does not have a symmetric extension for that k and hence ρ_{ab} must be entangled. On the other hand if the SDP is feasible, then ρ_{ab} may still be entangled.

Another criterion for checking if a state ρ_{ab} is separable is the positive under partial transpose (PPT) criterion [74]. Notice $\mathcal{T}_b$, a transpose with respect to $\mathcal{H}_b$ of the separable state of ρ_{ab} in (97), results in

$$\mathcal{T}_b(\rho_{ab}) = \sum_i p_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_i\rangle\langle\phi_i|^T, \tag{101}$$

where the superscript T represents transpose in the standard basis of $\mathcal{H}_b$. For any separable state ρ_{ab}, the operator $\mathcal{T}_b(\rho_{ab})$ above is positive semi-definite; thus, one says any separable state ρ_{ab} is PPT. On the other hand, if for some state σ_{ab} the partial transpose $\mathcal{T}_b(\sigma_{ab})$ is not PPT, then σ_{ab} cannot be separable; that is, σ_{ab} is entangled if it is not PPT. For $d_a = 2$ and $d_b = 2, 3$, this PPT criterion is both necessary and sufficient [44]. But generally, this PPT criterion is necessary but not sufficient. However, this PPT criterion can be combined with the previously discussed necessary and sufficient condition for a separable state to have a symmetric extension for all k. This combination results in the *PPT symmetric extension* criterion for separability [26]. In this criterion, for any $k \geq 1$, the symmetric extension ρ_{aB} of ρ_{ab} must also be PPT, where the partial transpose is taken with respect to each of the k spaces $\mathcal{H}_{b1}$, $\mathcal{H}_{b1b2}$, $\ldots$, $\mathcal{H}_{b1b2\ldots bk}$; that is,

$$\mathcal{T}_{b1b2\ldots bj}(\rho_{aB}) \succeq 0, \tag{102}$$

for all $1 \leq j \leq k$.

The PPT symmetric extension criterion also provides a route to check if a density operator ρ_{ab} is separable or entangled. In this route, one picks an integer k and first formulates an SDP for finding a symmetric extension ρ_{aB} and then adds linear PPT constraints (102) to this SDP. These additional PPT constraints can turn an SDP that was feasible to one that isn't. Such an infeasible SDP indicates the absence of a PPT symmetric extension, and thus the presence of entanglement. As a result, the PPT symmetric extension-based route to checking entanglement can be strictly better than the usual symmetric extension route.

The SDP formulations discussed so far were about checking feasibility. From a numerical standpoint, it is convenient to reformulate these feasibility problems as SDP optimization problems. For any $k \geq 1$, the SDP optimization problems arising from the PPT symmetric extension criterion can be written as

$$\begin{aligned} &\text{minimize } \mu \\ &\text{subject to } \rho_{ab} = \mathrm{Tr}_{b2,b3,\ldots bk}(\rho_{aB}), \\ &\qquad \rho_{aB} = \Pi_{bibj}\rho_{aB}\Pi_{bibj}, \qquad 1 \leq i < j \leq k, \\ &\qquad \mathrm{Tr}(\rho_{aB}) = 1, \\ &\qquad \rho_{aB} + \mu I_{aB} \succeq 0, \\ &\text{and } \mathcal{T}_{b1b2\ldots bj}(\rho_{aB}) + \mu I_{aB} \succeq 0, \quad 1 \leq j \leq k. \end{aligned} \tag{103}$$

Dropping the last constraint in the SDP above results in an SDP arising from the symmetric extension criterion. For some fixed k, if ρ_{ab} does not have a PPT symmetric extension then the optimum value of the above SDP, μ^*, is strictly positive. This strict positivity implies that ρ_{ab} is entangled. On the other hand, if μ^* is zero or less than zero, then ρ_{ab} has a PPT symmetric extension and this ρ_{ab} may or may not be entangled.

The simplest example of the SDP in (103) occurs for $k = 1$. In this case $\mathcal{H}_B = \mathcal{H}_{b1} = \mathcal{H}_b$, and thus for any state ρ_{ab}, the SDP above can be reduced to the form

$$\begin{aligned} &\text{minimize } \mu \\ &\text{subject to } \mathcal{T}_b(\rho_{ab}) + \mu I_{ab} \succeq 0. \end{aligned} \tag{104}$$

Here, μ^* is simply obtained by putting a negative sign in front of the smallest eigenvalue of $\mathcal{T}_b(\rho_{ab})$, the partial transpose of ρ_{ab} with respect to $\mathcal{H}_b$. Suppose $d_a = d_b = 2$, and $\rho_{ab} = |\chi\rangle\langle\chi|$, where $|\chi\rangle$ is defined in (16); then one may write

$$\rho_{ab} = \frac{1}{2}(|0\rangle\langle 0|_a \otimes |0\rangle\langle 0|_b + |1\rangle\langle 0|_a \otimes |1\rangle\langle 0|_b + |0\rangle\langle 1|_a \otimes |0\rangle\langle 1|_b + |1\rangle\langle 1|_a \otimes |1\rangle\langle 1|_b,), \tag{105}$$

or express ρ_{ab} above as a 4×4 matrix in the $|i\rangle_a \otimes |j\rangle_b$ basis as

$$\rho_{ab} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{106}$$

The partial transpose of ρ_{ab} with respect to $\mathcal{H}_b$,

$$\mathcal{T}_b(\rho_{ab}) = \frac{1}{2}(|0\rangle\langle 0|_a \otimes |0\rangle\langle 0|_b + |1\rangle\langle 0|_a \otimes |0\rangle\langle 1|_b + |0\rangle\langle 1|_a \otimes |1\rangle\langle 0|_b + |1\rangle\langle 1|_a \otimes |1\rangle\langle 1|_b), \tag{107}$$

can also be written in a matrix form,

$$\mathcal{T}_b(\rho_{ab}) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{108}$$

again using the $|i\rangle_a \otimes |j\rangle_b$ basis. The smallest eigenvalue of the above matrix is $-1/2$. As a result, the optimum value of the SDP (104), μ^*, is $1/2$. A strictly positive μ^* value indicates that $\rho_{ab} = |\chi\rangle\langle\chi|$ is an entangled state. This entanglement was already discussed below (16), the SDP approach above merely confirms this fact. Next, let $d_a = d_b = 3$ and ρ_{ab} be a state described in [45],

$$\rho_{ab} = \frac{2}{7}|\psi^+\rangle\langle\psi^+| + \frac{\alpha}{7}\sigma_+ + \frac{5-\alpha}{7}S_{ab}\sigma_+ S_{ab}, \tag{109}$$

where $0 \leq \alpha \leq 5$, $|\psi^+\rangle = \frac{1}{\sqrt{3}}(|00\rangle + |11\rangle + |22\rangle)$, $\sigma_+ = \frac{1}{3}(|01\rangle\langle 01| + |12\rangle\langle 12| + |20\rangle\langle 20|)$, and S_{ab} is the swap operator in (91). Replacing α with $5-\alpha$ in (109) is equivalent to swapping the $\mathcal{H}_a$ and $\mathcal{H}_b$ spaces. Such a swap has no effect on the solution to the SDP (104) and does not change whether ρ_{ab} is entangled or separable. Thus, we restrict ourselves to $0 \leq \alpha \leq 5/2$. Using a procedure similar to the one described above, one can compute $\mathcal{T}_b(\rho_{ab})$ and its smallest eigenvalue, $\alpha^* = (5 - \sqrt{4\alpha^2 - 20\alpha + 41})/42$. We know $\mu^* = -\alpha^*$. It is easy to check that μ^* is strictly positive for $0 \leq \alpha < 1$, and negative for $1 \leq \alpha \leq 5/2$. Thus we conclude, that for $\alpha < 1$, the qudit state ρ_{ab} in (109) is entangled. For other values of $1 \leq \alpha \leq 5/2$, μ^* is negative and one cannot conclude if ρ_{ab} is entangled or separable. A path forward to check entanglement for these other values of α is to solve a larger SDP by setting $k = 2$ in (103).

For $k = 2$, $\mathcal{H}_B = \mathcal{H}_{b1} \otimes \mathcal{H}_{b2}$, $\rho_{ab1} = \rho_{ab}$, and the SDP in (103) can be reduced to the form

$$\begin{aligned} &\text{minimize } \mu \\ &\text{subject to } \rho_{ab1} = \mathrm{Tr}_{b2}(\rho_{aB}), \\ &\qquad \rho_{aB} = \Pi_{b1b2}\rho_{aB}\Pi_{b1b2}, \\ &\qquad \mathrm{Tr}(\rho_{aB}) = 1, \\ &\qquad \rho_{aB} + \mu I_{aB} \succeq 0, \\ &\qquad \mathcal{T}_{b1}(\rho_{aB}) + \mu I_{aB} \succeq 0, \\ &\quad \text{and } \mathcal{T}_{b1b2}(\rho_{aB}) + \mu I_{aB} \succeq 0. \end{aligned} \tag{110}$$

To solve this SDP, we use open-source numerical packages. A short tutorial using Python packages PICOS and CVXOPT is available along with this chapter (see Notebook 4 in [90]). Using these packages, for the above SDP we find the optimal value μ^*. This value is strictly positive for $1 \leq \alpha < 2$ and thus ρ_{ab} in (109) is entangled for $1 \leq \alpha < 2$. This entanglement was not found by the SDP in (104) and it demonstrates that the SDP for $k = 2$ is strictly better at detecting entanglement than the SDP for $k = 1$. The finding that ρ_{ab} in (109) is entangled for $0 \leq \alpha < 2$ is consistent with [45], where the entanglement of ρ_{ab} was first discussed. That discussion also considers the parameter range $2 \leq \alpha \leq 5/2$, where ρ_{ab} in (109) is shown to be separable.

The methods for checking separability discussed here are often called an SDP hierarchy. The hierarchy discussed here is based on the work of [26]. This is not the only hierarchy; a number of other hierarchies have been studied [15,35,36,65]. While we have only provided a brief introduction, the theory of quantum entanglement is an active area of study with a variety of open problems (see references in and citation to [46]).

8. Problem 5: Quantum Channel Capacity

Information is processed via physical media. However, because physical media introduce noise, it is natural to ask the amount of noiseless information that can be sent across some noisy medium. To answer such questions, one constructs an abstract model for the noisy medium. This model is called a noisy communication channel. A classical channel sends distinguishable input symbols to distinguishable output symbols. Suppose the input symbols come from a discrete set, an alphabet $\mathcal{X}$, and the output symbols come from a possibly different alphabet $\mathcal{Y}$. A discrete memoryless channel N takes an input $x \in \mathcal{X}$ to an output $y \in \mathcal{Y}$ with probability $p(y|x)$. The input and output may be considered random variables X and Y, respectively, and the channel simply takes X to Y. This channel N is called memoryless because any output $Y = y$ only depends on the current channel input $X = x$ and not on a prior input.

Noise introduced by a channel can be corrected by error-correcting codes that encode and decode information across multiple channel uses. The rate of error correction across multiple channel uses is captured by the notion of an *achievable rate*. Roughly speaking, an

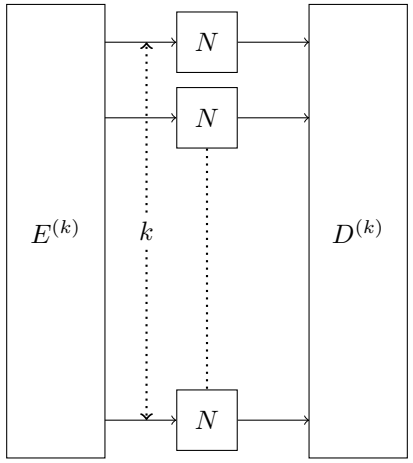

FIGURE 3. Encoding, $E^{(k)}$, and decoding, $D^{(k)}$, classical information across k uses of a classical channel N.

encoding $E^{(k)}$ and decoding $D^{(k)}$ over k joint uses of a channel N (see Fig. 3) that sends kR bits with vanishing error as $k \mapsto \infty$ is said to have an achievable rate R. The maximum possible achievable rate is called the channel capacity $C(N)$. Achievable rates and channel capacity are fundamental quantities in information theory; Shannon [82] provided a simple way to compute an achievable rate for any given channel N with conditional probability $p(y|x)$. This rate, which we call the *channel mutual information*, is simply given by

$$C^{(1)}(N) = \max_{p(x)} I(X:Y), \tag{111}$$

where $p(x)$ is a probability distribution over input symbols x, and $I(X;Y)$ is the mutual information (28) between the input X and output Y. For any fixed N —that is, fixed $p(y|x)$— the mutual information is concave in $p(x)$ (see Th.2.7.4 in [21]). As a result, $C^{(1)}(N)$ can be computed efficiently using tools from convex optimization [4, 17]. In addition, the channel mutual information is additive: for any two channels N and N' used together the channel mutual information $C^{(1)}(N \times N')$ is simply the sum $C^{(1)}(N) + C^{(1)}(N')$. The channel capacity $C(N)$ can be written in terms of the channel mutual information as a limit

$$C(N) = \lim_{k \mapsto \infty} \frac{1}{k} C^{(1)}(N^{\times k}), \tag{112}$$

where $N^{\times k}$ represents k joint uses of N. This limit greatly simplifies due to additivity, $C(N) = C^{(1)}(N)$, a remarkable *single-letter* expression. From this expression, the capacity of any channel N to send information over infinitely many channel uses is given by the maximum mutual information between the input and output of a single use of the channel.

Both information and the physical medium carrying this information can be modelled using quantum mechanics. In this model, noise introduced by physical medium is described by a quantum channel. A fundamental question in quantum information theory is to understand the maximum rate at which noiseless quantum information can be sent across a noisy quantum channel (for instance see [13]). This question is answered in a way analogous to the one above used by Shannon to understand the capacity of a classical channel.

Noise introduced by a quantum channel $\mathcal{B}$ can be corrected by using quantum error-correcting codes that encode and decode information across quantum channels. Roughly speaking, a quantum code with encoding $\mathcal{E}^{(k)}$ and decoding $\mathcal{D}^{(k)}$ over k joint uses of a channel $\mathcal{B}$ (see Fig. 4), which sends kR qubits with vanishing error as $k \mapsto \infty$, is said to have

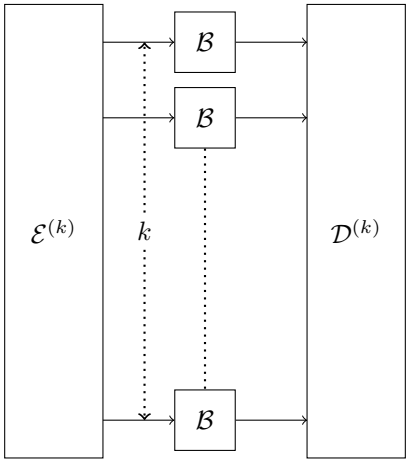

FIGURE 4. Encoding, $\mathcal{E}^{(k)}$, and decoding, $\mathcal{D}^{(k)}$, quantum information across k uses of a quantum channel $\mathcal{B}$.

an achievable rate R. The maximum possible achievable rate of this type is defined to be the quantum capacity $\mathcal{Q}(\mathcal{B})$. The quantum analog of the channel mutual information $C^{(1)}$ is the *channel coherent information* of a channel $\mathcal{B} : \hat{\mathcal{H}}_a \to \hat{\mathcal{H}}_b$ [8],

$$\mathcal{Q}^{(1)}(\mathcal{B}) = \max_{\rho_a} I_c(\mathcal{B}, \rho_a), \tag{113}$$

where $I_c(\mathcal{B}, \rho_a) := S\big(\mathcal{B}(\rho_a)\big) - S\big(\mathcal{B}^c(\rho_a)\big)$ is the *entropy bias* or the *coherent information* of a channel $\mathcal{B}$ at ρ and $\mathcal{B}^c : \hat{\mathcal{H}}_a \to \hat{\mathcal{H}}_c$ is the complementary channel of $\mathcal{B}$ and $\mathcal{H}_c$ the complementary output space, sometimes called the environment of $\mathcal{B}$. Recall any channel $\mathcal{B}$ has a Kraus decomposition (45), written using d_c Kraus operators $K_i : \mathcal{H}_a \to \mathcal{H}_b$; the complementry channel $\mathcal{B}^c$ can be defined using a Kraus decomposition,

$$\mathcal{B}^c(A) = \sum_{j=1}^{d_b} L_j A L_j^\dagger, \tag{114}$$

where the Kraus operators $L_j : \mathcal{H}_a \to \mathcal{H}_c$ have matrix elements $[L_j]_{ki} = [K_i]_{jk}$.

Unlike $I(X:Y)$, $I_c(\mathcal{B}, \rho_a)$ is not necessarily concave in the input ρ_a for fixed $\mathcal{B}$. As a result, despite the fundamental importance of $\mathcal{Q}^{(1)}(\mathcal{B})$, methods for computing $\mathcal{Q}^{(1)}$ are limited. While $C^{(1)}$ is additive, its quantum analog $\mathcal{Q}^{(1)}$ is non-additive; that is, for two channels $\mathcal{B}$ and $\mathcal{B}'$ used together, the coherent information of the joint channel satisfies an inequality

$$\mathcal{Q}^{(1)}(\mathcal{B} \otimes \mathcal{B}') \geq \mathcal{Q}^{(1)}(\mathcal{B}) + \mathcal{Q}^{(1)}(\mathcal{B}'), \tag{115}$$

which can be strict [9, 10, 25, 61, 86, 87, 92, 93], where $\mathcal{B} \otimes \mathcal{B}'$ is the tensor product of two channels (see the discussion containing (44) for definition).

The quantum capacity $\mathcal{Q}(\mathcal{B})$ can be written in terms of the channel coherent information as a limit [23, 63, 85]

$$\mathcal{Q}(\mathcal{B}) = \lim_{k \mapsto \infty} \frac{1}{k} \mathcal{Q}^{(1)}(\mathcal{B}^{\otimes k}), \tag{116}$$

where $\mathcal{B}^{\otimes k}$ denotes k tensor products of $\mathcal{B}$ with itself. This expression for the quantum capacity requires computing a limit over multiple uses of the same channel. In general, this *multi-letter* expression for $\mathcal{Q}$ is intractable to compute because of non-additivity. As a result, $\mathcal{Q}^{(1)}$ is always a lower bound on $\mathcal{Q}$, but it need not equal $\mathcal{Q}$. For a special class of channels called *degradable channels* $\mathcal{Q}^{(1)} = \mathcal{Q}$ [24]. A channel $\mathcal{B}$ is said to be degradable if

there is another channel $\mathcal{C}$ such that $\mathcal{C} \circ \mathcal{B} = \mathcal{B}^c$, and $\mathcal{B}^c$, the complement of $\mathcal{B}$, is called anti-degradable. For any two (anti) degradable channels $\mathcal{B}$ and $\mathcal{B}'$, the inequality in (115) is an equality [60]; that is, additivity holds. Such additivity simplifies the multi-letter expression for $\mathcal{Q}$ in (116) to a single-letter formula:

$$\mathcal{Q}(\mathcal{B}) = \mathcal{Q}^{(1)}(\mathcal{B}), \tag{117}$$

where $\mathcal{B}$ is a degradable or anti-degradable channel. For a degradable channel, $\mathcal{Q}^{(1)}(\mathcal{B})$ is relatively easy to compute: the entropy bias $\Delta(\mathcal{B}, \rho_a)$ for a degradable channel $\mathcal{B}$ is a concave function of ρ [110] and can be maximized using tools from convex optimization [30,76].

Given the role of degradable channels in simplifying the discussion of quantum capacities, it is natural to ask if approximate notions of degradable channels can approximately simplify the discussion of quantum capacities. One such approximate notion is that of an ϵ-degradable channel [95]. A channel $\mathcal{B}$ is ϵ-degradable if there is another channel $\mathcal{C}$ such that

$$||\mathcal{C} \circ \mathcal{B} - \mathcal{B}^c||_\diamond = \epsilon. \tag{118}$$

If ϵ is zero, then $\mathcal{B}$ is degradable and (117) holds. If ϵ is not zero, then (117) gets modified to

$$\mathcal{Q}^{(1)}(\mathcal{B}) \leq \mathcal{Q}(\mathcal{B}) \leq \mathcal{Q}^{(1)}(\mathcal{B}) + \epsilon \log(d_c - 1)/2 + h(\epsilon/2) + \epsilon \log(d_c) + (1 + \frac{\epsilon}{2}) h\big(\epsilon/(2+\epsilon)\big), \tag{119}$$

where d_c is the output dimension of the complementary channel $\mathcal{B}^c$. The smallest ϵ for which (118) holds, $\epsilon_{\mathcal{B}}$, can be found using a semi-definite program (SDP),

$$\begin{aligned} &\text{minimize } 2\mu \\ &\text{subject to } \mathrm{Tr}_c(Z_{ca}) \preceq \mu I_a, \\ &\qquad \mathrm{Tr}_c\big(\mathcal{J}_{cb}(\mathcal{C})\big) = I_b, \\ &\qquad Z_{ca} \succeq \mathcal{J}_{ca}(\mathcal{B}^c) - \mathcal{J}_{ca}(\mathcal{C} \circ \mathcal{B}), \\ &\qquad Z_{ca} \succeq 0, \\ &\quad \text{and } \mathcal{J}_{cb}(\mathcal{C}) \succeq 0. \end{aligned} \tag{120}$$

In this SDP, μ is a real variable while Z_{ca} and $\mathcal{J}_{cb}(\mathcal{C})$ are positive semi-definite variables. These variables satisfy linear constraints. In particular the constraint $Z_{ca} \succeq \mathcal{J}_{ca}(\mathcal{B}^c) - \mathcal{J}_{ca}(\mathcal{C} \circ \mathcal{B})$ is linear in $\mathcal{J}_{cb}(\mathcal{C})$ because $\mathcal{J}_{ca}(\mathcal{C} \circ \mathcal{B})$ is linear in $\mathcal{J}_{cb}(\mathcal{C})$ and can be written using $\mathcal{J}_{ca}(\mathcal{B})$ (see App. B). The operators $\mathcal{J}_{ca}(\mathcal{B}^c)$ and $\mathcal{J}_{cb}(\mathcal{B})$ are both constants which only depend on the channel $\mathcal{B}$.

For any channel $\mathcal{B}$, the SDP (120) can be solved numerically to compute $\epsilon_{\mathcal{B}}$. However, $\epsilon_{\mathcal{B}}$ alone is not sufficient to evaluate the bound (119) on the quantum capacity of $\mathcal{Q}(\mathcal{B})$. In addition, one requires $\mathcal{Q}^{(1)}(\mathcal{B})$. As stated earlier, methods for computing $\mathcal{Q}^{(1)}(\mathcal{B})$, obtained from solving a non-convex optimization problem (113), are limited. This limitation makes it non-trivial to evaluate the bound (119) for an abritrary channel $\mathcal{B}$. However, there are special well studied channels for which $\mathcal{Q}^{(1)}(\mathcal{B})$ is known [25,32,88,109]. These include channels that are degradable and have $\epsilon_{\mathcal{B}} = 0$, but also include channels that are not degradable.

One simple example of a degradable channel is the qubit dephasing channel with dephasing probability q [109],

$$\mathcal{F}_q(\rho) = (1-q)\rho + qZ\rho Z^\dagger, \tag{121}$$

where Z is the Pauli matrix defined in (34). The coherent information of this channel, $\mathcal{Q}^{(1)}(\mathcal{F}_q)$, is simply $I_c(\mathcal{F}_q, I/2) = 1 - h(q)$. Since $\mathcal{F}_q$ channel is degradable, $\epsilon_{\mathcal{F}_q} = 0$. One may verify this by solving the SDP (120) to obtain a numerical value $\epsilon^*_{\mathcal{F}_q}$ of $\epsilon_{\mathcal{F}_q}$. We find the absolute difference between $\epsilon^*_{\mathcal{F}_q}$ and $\epsilon_{\mathcal{F}_q}$ to be small, $O(10^{-10})$. Using $\epsilon_{\mathcal{F}_q}$, an equation of the form (119) for $\mathcal{F}_q$ gives $\mathcal{Q}^{(1)}(\mathcal{F}_q) \leq \mathcal{Q}(\mathcal{F}_q) \leq \mathcal{Q}^{(1)}(\mathcal{F}_1)$. These inequalities simply state that the quantum capacity of $\mathcal{F}_q$ equals its coherent information, $1 - h(q)$.

Another simple example of a degradable channel is the erasure channel. Recall, the erasure channel $\mathcal{E}_p$ with erasure probability p acting on a d_a-dimensional input is defined in (42). We consider the case where $0 \leq p \leq 1/2$. For these values of p, the erasure channel is degradable, $\epsilon_{\mathcal{E}_p} = 0$ and agrees with its numerical estimate $\epsilon^*_{\mathcal{E}_q}$ up to $O(10^{-9})$. This estimate is obtained by solving the SDP (120) for a qubit erasure channel with erasure probability chosen randomly between zero and one half. The channel coherent information, $\mathcal{Q}^{(1)}(\mathcal{E}_p)$, simply equals $I_c(\mathcal{E}_p, I/d_a) = (1-2p)\log d_a$ (see Sec.4 in [89]). These results, together with (119), imply that $\mathcal{Q}^{(1)}(\mathcal{E}_p) = \mathcal{Q}(\mathcal{E}_p)$ for $p \leq 1/2$. This equality also holds for $1/2 < p \leq 1$, where $\mathcal{E}_p$ is anti-degradability.

As our final example, we consider the qubit depolarizing channel, Δ, with Kraus decomposition (48),

$$\Delta(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z), \tag{122}$$

where $0 \leq p \leq 1$. The coherent information of this channel is

$$\mathcal{Q}^{(1)}(\Delta) = \max(0, 1 - h(p) - p\log_2 3). \tag{123}$$

The qubit depolarizing channel Δ is not degradable, except at $p = 0$. The smallest ϵ for which equation (118) holds must be found by numerically solving an SDP (120). This numerical solution, along with (119), can be used to compute bounds on the quantum capacity of Δ. For instance, when the depolarizing parameter $p \simeq .026$, solving the SDP (120) gives a numerical value $\epsilon^*_{\mathcal{D}} \simeq 3.8 \times 10^{-3}$ of $\epsilon_{\mathcal{D}}$. Using this value, one finds that $|\mathcal{Q}(\Delta) - \mathcal{Q}^{(1)}(\Delta)| \leq 5.1 \times 10^{-2}$. For these and other numerics of this type, a short Python based notebook accompanies chapter (see Notebook 5 in [90]). This same notebook has additional examples consisting of the qubit dephasing and erasure channels.

So far we have discussed ϵ-degradable channels, which is one approximate notion of degradability. This notion allows one to find computable bounds on the quantum capacity of channels with known coherent information. Another notion for approximate degradability is ϵ-close degradable channels [95]. A given channel $\mathcal{B}$ is ϵ-close degradable if there is a degradable channel $\mathcal{M}$ which is ϵ close to $\mathcal{B}$ in diamond norm distance, i.e., $||\mathcal{B} - \mathcal{M}||_\diamond = \epsilon$. If $\mathcal{B}$ is ϵ-close degradable, then its quantum capacity can be bounded as follows

$$|\mathcal{Q}(\mathcal{B}) - \mathcal{Q}^{(1)}(\mathcal{M})| \leq \epsilon \log(d_b) + (2+\epsilon)h(\frac{\epsilon}{2+\epsilon}). \tag{124}$$

In the expression above $\mathcal{Q}^{(1)}(\mathcal{M})$ can be computed using tools from convex optimization because $\mathcal{M}$ is degradable. On the other hand, there is no known way to efficiently compute the smallest ϵ for which a given channel $\mathcal{B}$ is ϵ-close degradable. Such computations could potentially lead to new and possibly tighter bounds (124) on the quantum capacity of channels. Finding such bounds is an active area of fundamental research in quantum Shannon theory [27,28,42,70,91,100,101].

9. Concluding Remarks

The SDP models were chosen here so that they can be accompanied by working numerical examples to aid in learning, as a companion to closed-form solutions possible in special situations. However, practical use of SDPs for QIS can easily become numerically intractable for current solvers as the number of quantum systems involved increase (a quantum system consisting of k qubits has dimension $d = 2^k$ and it introduces $O(2^k)$ variables, which is exponential in k.) This can make the SDPs in Secs. 4,5,6, and 7 numerically intractable for larger k. This is not the only manner in which SDPs in QIS become intractable. As one example, the quantum query complexity of a Boolean function of n-bits can be computed using an SDP that has $O(2^n)$ variables(see [77,78] and references therein). Another reason is non-linearity: certain optimization problems involving the von-Neumann entropy, a non-linear function, can be approximated using an SDP [30,31], with numerical effort increasing with precision, sometimes prohibitively. We hope that this introduction to QIS through these starter problems provides a rapid and accessible gateway to new researchers.

Acknowledgments

VS gratefully acknowledges support from NSF CAREER Award CCF 1652560 and NSF grant PHY 1915407, thanks Mark M. Wilde for helpful comments, and thanks to Maximilian Stahlberg for support with PICOS.

Appendix

A. Norms of Superoperators

In Sec. 6 we defined a superoperator $\mathcal{D}$ (87), which can we written as,

$$\mathcal{D}(\rho) = \frac{(1-\lambda)}{2}(\mathcal{I}(\rho) - \mathcal{T}(\rho)), \tag{125}$$

where $\mathcal{I}(\rho) = \rho$ and $\mathcal{T}(\rho) = \mathrm{Tr}(\rho)\frac{I}{d}$. Here we compute the l_1-norm and the diamond norm of this superoperator. The l_1-norm of $\mathcal{D}$ is given by (77)

$$||\mathcal{D}||_1 = \max_{\rho} ||\mathcal{D}(\rho)||_1. \tag{126}$$

The set of density operators is convex and $||.||_1$, the $1-$norm of an operator, is a convex function. As a result, the optimum value of the convex maximization problem above is achieved at an extreme point of the set of density operator. These extreme points are projectors onto pure states, i.e., $|\psi\rangle\langle\psi|$ where $\langle\psi|\psi\rangle = 1$. Thus

$$||\mathcal{D}||_1 = \max_{|\psi\rangle} ||\mathcal{D}(|\psi\rangle\langle\psi|)||_1. \tag{127}$$

A simple calculation shows that for $\mathcal{D}$ in (125), $\mathcal{D}(|\psi\rangle\langle\psi|)$ is independent of $|\psi\rangle$ and equals $(1-\lambda)(d-1)/d$.

To compute the diamond norm of $\mathcal{D}$ in (125) we use a technique similar to the one employed in [62] to compute $||\mathcal{D}||_\diamond$ for $d = 2$. Using (78), we can bound $||\mathcal{D}||_\diamond$ from below by $||\mathcal{I}_d \otimes \mathcal{D}(X)||_1$, where $||X||_1 \leq 1$. Letting $X = |\phi_d\rangle\langle\phi_d|$, where

$$|\phi_d\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle \otimes |i\rangle \tag{128}$$

is the maximally entangled state on two qudits, gives

$$||\mathcal{D}||_\diamond \geq ||\mathcal{D}(|\phi\rangle\langle\phi|) \otimes \mathcal{I}_d||_1 = (1-\lambda)\frac{d^2-1}{d^2}. \tag{129}$$

This lower bound can be matched by an upper bound. To obtain this upper bound notice $\mathcal{D}$ in (125) is proportional to $\mathcal{N} = \mathcal{I} - \mathcal{T}$, the difference of two quantum channels. The diamond norm of such a difference is given by the SDP [104]:

$$\begin{aligned} &\text{minimize } 2\mu \\ &\text{subject to } \mu I_a \succeq \mathrm{Tr}_b(Z_{ba}), \\ &\qquad Z_{ba} \succeq \mathcal{J}_{ba}(\mathcal{N}), \\ &\qquad \text{and } Z_{ab} \succeq 0. \end{aligned} \tag{130}$$

Notice $Z_{ba} = \frac{d^2-1}{d}|\phi_d\rangle\langle\phi_d|$ is a feasible solution of this SDP with $\mu = \frac{d^2-1}{d^2}$, thus $||\mathcal{N}||_\diamond \leq 2\frac{d^2-1}{d^2}$. Since $||\mathcal{D}||_\diamond = \frac{(1-\lambda)}{2}||N||_\diamond$, we get a matching upper bound to (129).

B. Transfer Matrix and Choi-Jamiołkowski representation

Let $\mathcal{H}$ be a d-dimensional complex space and $\mathcal{L}(\mathcal{H})$ be the space of linear operators on $\mathcal{H}$. Given two operators A and B in $\mathcal{L}(\mathcal{H})$, their Frobenius inner product is

$$\langle A, B\rangle := \mathrm{Tr}(A^\dagger B). \tag{131}$$

Using this inner product, one can define an orthonormal basis of d^2 operators, $\{|i\rangle\langle j|\}$, for $\mathcal{L}(\mathcal{H})$. Using such a basis, a channel superoperator $\mathcal{B} : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$ can be written in matrix form,

$$\mathcal{B}(|k\rangle\langle l|_a) = \sum_{i,j} T(\mathcal{B})_{kl,ij}|i\rangle\langle j|_b, \tag{132}$$

where the complex numbers

$$T(\mathcal{B})_{kl,ij} = \langle \mathcal{B}(|k\rangle\langle l|_a), |i\rangle\langle j|_b \rangle, \tag{133}$$

form the entries of a $d_b^2 \times d_a^2$ matrix $T(\mathcal{B})$, sometimes called the *transfer matrix* of the channel $\mathcal{B}$. The rows and columns of $T(\mathcal{B})$ are indexed by $0 \leq k, l \leq d_b - 1$ and $0 \leq i, j \leq d_a - 1$ respectively. This transfer matrix $T(\mathcal{B}) : \mathcal{H}_a \otimes \mathcal{H}_a \to \mathcal{H}_b \otimes \mathcal{H}_b$ is related to the channel's Choi-Jamiołkowski representation $\mathcal{J}_{ba}(\mathcal{B}) \in \mathcal{H}_{ba}$ (defined in (50)) as follows

$$\langle kl|T(\mathcal{B})|ij\rangle = \langle ki|\mathcal{J}_{ba}(\mathcal{B})|lj\rangle, \tag{134}$$

where we use the notation $|ij\rangle := |i\rangle \otimes |j\rangle$.

Let $\mathcal{B} : \mathcal{L}(\mathcal{H}_a) \to \mathcal{L}(\mathcal{H}_b)$ and $\mathcal{C} : \mathcal{L}(\mathcal{H}_b) \to \mathcal{L}(\mathcal{H}_c)$ be two channels with transfer matrices $T(\mathcal{C})$ and $T(\mathcal{B})$, respectively. Using the two channels in series results in a third channel $\mathcal{C} \circ \mathcal{B}$ with transfer matrix $T(\mathcal{C} \circ \mathcal{B})$. This matrix is simply the product of the individual transfer matrices and is given by

$$T(\mathcal{C} \circ \mathcal{B}) = T(\mathcal{C})T(\mathcal{B}), \tag{135}$$

Using the above equation, along with the relation (134) one can show that $\mathcal{J}_{ca}(\mathcal{C} \circ \mathcal{B})$, the Choi-Jamiołkowski representation of $\mathcal{C} \circ \mathcal{B}$, is just a linear function of $\mathcal{J}_{ba}(\mathcal{B})$ and $\mathcal{J}_{cb}(\mathcal{C})$.

References

[1] F. Alizadeh. *Combinatorial optimization with interior point methods and semi-definite matrices.* PhD thesis, University of Minnesota, USA. UMI Order No. GAX92-07776, 1992.

[2] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13-51, 1995.

[3] M. Andersen, J. Dahl, and L. Vandenberghe. CVXOPT: Convex optimization, 2021.

[4] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14-20, 1972.

[5] K. M. R. Audenaert, J. Calsamiglia, R. M. Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum Chernoff bound. *Physical Review Letters*, 98:160501, 2007.

[6] J. Bae and L.-C. Kwek. Quantum state discrimination and its applications. *Journal of Physics A: Mathematical and Theoretical*, 48(8):083001, 2015.

[7] S. M. Barnett and S. Croke. Quantum state discrimination. *Adv. Opt. Photon.*, 1(2):238-278, 2009.

[8] H. Barnum, M. A. Nielsen, and B. Schumacher. Information transmission through a noisy quantum channel. *Physical Review A*, 57:4153-4175, 1998.

[9] J. Bausch and F. Leditzky. Quantum codes from neural networks. *New Journal of Physics*, 22(2):023005, 2020.

[10] J. Bausch and F. Leditzky. Error thresholds for arbitrary Pauli noise. *SIAM Journal on Computing*, 50(4):1410–1460, 2021.

[11] R. Bellman and K. Fan. On systems of linear inequalities in Hermitian matrix variables. *Proc. Sympos. Pure Math.*, VII:1-11, 1963.

[12] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing*, page 175, India, 1984.

[13] C. H. Bennett and P. W. Shor. Quantum information theory. *IEEE Transactions on Information Theory*, 44(6):2724-2742, 1998.

[14] C. H. Bennett and P. W. Shor. Quantum channel capacities. *Science*, 303(5665):1784-1787, 2004.

[15] M. Berta, O. Fawzi, and V. B. Scholz. Quantum bilinear optimization. *SIAM Journal on Optimization*, 26(3):1529–1564, 2016.

[16] A. K. Bhattacharya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of Calcutta Mathematical Society*, 35:99-110, 1943.

[17] R. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460-473, 1972.

[18] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics, 1994.

[19] A. Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401-424, 2000.

[20] M. -D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285 - 290, 1975.

[21] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Ltd, 2001.

[22] G. B. Dantzig and M. N. Thapa. *Linear Programming 1*. Springer-Verlag New York, 1997.

[23] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44-55, 2005.

[24] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256(2):287-303, 2005.

[25] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum-channel capacity of very noisy channels. *Physical Review A*, 57:830-839, 1998.

[26] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2), 022308, 2004.

[27] K. Fang and H. Fawzi. Geometric rényi divergence and its applications in quantum channel capacities. *Communications in Mathematical Physics*, 384(3):1-63, 2021.

[28] M. Fanizza, F. Kianvash, and V. Giovannetti. Quantum flags and new bounds on the quantum capacity of the depolarizing channel. *Physical Review Letters*, 125:020503, 2020.

[29] M. Fannes, J. T. Lewis, and A. Verbeure. Symmetric states of composite systems. *Letters in Mathematical Physics*, 15(3):255-260, 1988.

[30] H. Fawzi and O. Fawzi. Efficient optimization of the quantum relative entropy. *Journal of Physics A: Mathematical and Theoretical*, 51(15):154003, 2018.

[31] H. Fawzi, J. Saunderson, and P. A. Parrilo. Semidefinite approximations of the matrix logarithm. *Foundations of Computational Mathematics*, 19(2):259-296, 2019.

[32] J. Fern and K. B. Whaley. Lower bounds on the nonzero capacity of Pauli channels. *Physical Review A*, 78:062335, 2008.

[33] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.

[34] G. Gutoski, A. Rosmanis, and J. Sikora. Fidelity of quantum strategies with applications to cryptography. *Quantum*, 2:89, 2018.

[35] A. W. Harrow, A. Natarajan, and X. Wu. An improved semidefinite programming hierarchy for testing entanglement. *Communications in Mathematical Physics*, 352(3):881–904, 2017.

[36] A W. Harrow, A. Natarajan, and X. Wu. Limitations of semidefinite programs for separable states and entangled games. *Communications in Mathematical Physics*, 366(2):423–468, 2019.

[37] J. D. Hauenstein, A. C. Liddell, S. McPherson, and Y. Zhang. Numerical algebraic geometry and semidefinite programming. *Results in Applied Mathematics*, 11:100166, 2021.

[38] M. Hayashi. *Quantum Hypothesis Testing and Discrimination of Quantum States*, pages 95-153. Springer Berlin Heidelberg, Berlin, Heidelberg, 2017.

[39] C. W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231-252, 1969.

[40] A. S. Holevo. *Quantum Systems, Channels, Information*. De Gruyter, 2012.

[41] A. S. Holevo and V. Giovannetti. Quantum channels and their entropic characteristics. *Reports on Progress in Physics*, 75(4):046001, 2012.

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

[43] M. Horodecki, P. W. Shor, and M. B. Ruskai. Entanglement breaking channels. *Reviews in Mathematical Physics*, 15(06):629-641, 2003.

[44] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1-8, 1996.

[45] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and quantum communication. *arXiv:quant-ph*, 0109124, 2001.

[46] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81:865-942, 2009.

[47] P. S. Iyer and D. Poulin. A small quantum computer is needed to optimize fault-tolerant *Quantum Sci. Technol.* 3 030504, 2018

[48] A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275-278, 1972.

[49] A. Kamath and N. Karmarkar. *A Continuous Approach to Compute Upper Bounds in Quadratic Maximization Problems With Integer Constraints:*, pages 125-140. Princeton University Press, 2014.

[50] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998.

[51] V. Kargin. On the Chernoff bound for efficiency of quantum hypothesis testing. *The Annals of Statistics*, 33(2):959-976, 2005.

[52] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373-395, 1984.

[53] V. Katariya and M. M. Wilde. Geometric distinguishability measures limit quantum channel estimation and discrimination. *Quantum Information Processing*, 20(2):78, 2021.

[54] S. Khatri, K. Sharma, and M M. Wilde. Information-theoretic aspects of the generalized amplitude-damping channel. *Physical Review A*, 102:012401, 2020.

[55] S. Khatri and M. M. Wilde. Principles of quantum communication theory: A modern approach. *arXiv:* 2011.04672, 2020

[56] A. S. Kholevo. On asymptotically optimal hypothesis testing in quantum statistics. *Theory of Probability & Its Applications*, 23(2):411-415, 1979.

[57] N. Killoran. *Entanglement quantification and quantum benchmarking of optical communication devices*. PhD thesis, University of Waterloo, 2012.

[58] M. Kliesch, R. Kueng, J. Eisert, and D. Gross. Improving compressed sensing with the diamond norm. *IEEE Transactions on Information Theory*, 62(12):7445-7463, 2016.

[59] K. Kraus. *States, Effects, and Operations Fundamental Notions of Quantum Theory*. Springer, Berlin, Heidelberg, 1983.

[60] F. Leditzky, N. Datta, and G. Smith. Useful states and entanglement distillation. *IEEE Transactions on Information Theory*, 64(7):4689-4708, 2018.

[61] F. Leditzky, D. Leung, and G. Smith. Dephrasure channel and superadditivity of coherent information. *Physical Review Letters*, 121:160501, 2018.

[62] F. Leditzky, D. Leung, and G. Smith. Quantum and private capacities of low-noise channels. *Physical Review Letters*, 120:160503, 2018.

[63] S. Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55:1613-1622, 1997.

[64] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Physical Review A*, 85:042311, 2012.

[65] M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[66] A. S. Nemirovsky and D. B. Yudin. Informational complexity and efficient methods for solving complex extremal problems. *Matekon*, 13:25-45, 1977.

[67] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[68] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[69] M. Nussbaum and A. Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *The Annals of Statistics*, 37(2):1040 - 1057, 2009.

[70] Y. Ouyang. Channel covariance, twirling, contraction and some upper bounds on the quantum capacity. *Quantum Information and Computation*, 14:0917-0936, 2014.

[71] M. Overton and H. Wolkowicz. Semidefinite programming. *Mathematical Programming*, 77(1):105-109, 1997.

[72] K. R. Parthasarathy. On Consistency of the Maximum Likelihood Method in Testing Multiple Quantum Hypotheses. In T. Hida, R. L. Karandikar, H. Kunita, B. S. Rajput, S. Watanabe, J. Xiong, editors, *Stochastics in Finite and Infinite Dimensions*, pages 361-377, Birkhäuser Boston, Boston, MA, 2001.

[73] G. Pataki. Cone-LP's and semidefinite programs: Geometry and a simplex-type method. In W. H. Cunningham, S. T. McCormick, and M. Queyranne, editors, *Integer Programming and Combinatorial Optimization*, pages 162-174, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg.

[74] A. Peres. Separability criterion for density matrices. *Physical Review Letters*, 77:1413-1415, 1996.

[75] G. Raggio and R. Werner. Quantum statistical mechanics of general mean field systems. *Helvetica Physica Acta*, 62:980, 03 1989.

[76] N. Ramakrishnan, R. Iten, V. B. Scholz, and M. Berta. Computing quantum channel capacities. *IEEE Transactions on Information Theory*, 67(2):946-960, 2021.

[77] B. W. Reichardt. Least span program witness size equals the general adversary lower bound on quantum query complexity. Technical Report TR10-057, Electronic Colloquium on Computational Complexity, 2010.

[78] B. W. Reichardt. Span programs and quantum query algorithms. Technical Report TR10-110, Electronic Colloquium on Computational Complexity, 2010.

[79] G. Sagnol and M. Stahlberg. PICOS: A Python interface to conic optimization solvers. *Journal of Open Source Software*, 7(70), 3915, 2022.

[80] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.

[81] B. Schumacher. Quantum coding. *Physical Review A*, 51:2738-2747, 1995.

[82] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379-423, 1948.

[83] N. Z. Shor. Cut-off method with space extension in convex programming problems. *Cybernetics*, 13(1):94-96, 1977.

[84] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[85] P. W. Shor. Quantum error correction, https://www.msri.org/workshops/203/schedules/1181, 2002.

[86] V. Siddhu. Leaking information to gain entanglement. *arXiv*:2011.15116, 2020.

[87] V. Siddhu. Entropic singularities give rise to quantum transmission. *Nature Communications*, 12(1):5750, 2021.

[88] V. Siddhu and R. B. Griffiths. Degradable quantum channels using pure-state to product-of-pure-state isometries. *Physical Review A*, 94:052331, 2016.

[89] V. Siddhu and R. B. Griffiths. Positivity and nonadditivity of quantum capacities using generalized erasure channels. *IEEE Transactions on Information Theory*, 67(7):4533-4545, 2021.

[90] V. Siddhu and S. Tayur. SDP-Quantum-OR. https://github.com/vsiddhu/SDP-Quantum-OR, 2021.

[91] G. Smith and J.A. Smolin. Additive extensions of a quantum channel. In *Information Theory Workshop, 2008. ITW '08. IEEE*, pages 368-372, 2008.

[92] G. Smith and J A. Smolin. Degenerate quantum codes for Pauli channels. *Physical Review Letters*, 98:030501, 2007.

[93] G. Smith and J. Yard. Quantum communication with zero-capacity channels. *Science*, 321(5897):1812-1815, 2008.

[94] E. C. G. Sudarshan, P. M. Mathews, and J. Rau. Stochastic dynamics of quantum-mechanical systems. *Physical Review*, 121:920-924, 1961.

[95] D. Sutter, V. B. Scholz, A. Winter, and R. Renner. Approximate degradable quantum channels. *IEEE Transactions on Information Theory*, 63(12):7832-7844, 2017.

[96] M. J. Todd. Semidefinite optimization. *Acta Numerica*, 10:515–560, 2001.

[97] A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273-279, 1976.

[98] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49-95, 1996.

[99] X. Wang. *Semidefinite optimization for quantum information*. PhD thesis, University of Technology Sydney, 2018.

[100] X. Wang and R. Duan. A semidefinite programming upper bound of quantum capacity. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1690-1694, 2016.

[101] X. Wang, K. Fang, and R. Duan. Semidefinite programming converse bounds for quantum communication. *IEEE Transactions on Information Theory*, 65(4):2583-2592, 2018.

[102] X. Wang and M. M. Wilde. Exact entanglement cost of quantum states and channels under PPT-preserving operations.*arXiv*:1809.09592, 2018.

[103] J. Watrous. Distinguishing quantum operations having few Kraus operators. *Quantum Info. Comput.*, 8(8):819–833, 2008.

[104] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11):217-238, 2009.

[105] J. Watrous. Simpler semidefinite programs for completely bounded norms. *Chicago Journal of Theoretical Computer Science*, Article 08:01-19, 2013.

[106] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 1st Edition, 2018.

[107] R. F. Werner. An application of bell's inequalities to a quantum state extension problem. *Letters in Mathematical Physics*, 17(4):359-363, 1989.

[108] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2nd Edition, 2017.

[109] M. M. Wolf and D. P.érez-García. Quantum capacities of channels with small environment. *Physical Review A*, 75:012303, 2007.

[110] J. Yard, P. Hayden, and I. Devetak. Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions. *IEEE Transactions on Information Theory*, 54(7):3091-3113, 2008.

[111] H. Yuan and C.-H. F. Fung. Fidelity and Fisher information on quantum channels. *New Journal of Physics*, 19(11):113039, 2017.

[112] L. Zhang, L. Chen, and K. Bu. Fidelity between one bipartite quantum state and another undergoing local unitary dynamics. *Quantum Information Processing*, 14(12):4715-4730, 2015.