Certifying optimality for convex quantum channel optimization problems

Bryan Coutts[1,2]   Mark Girard[1]   John Watrous[1,2,3]

[1] *Institute for Quantum Computing*
University of Waterloo, Canada

[2] *School of Computer Science*
University of Waterloo, Canada

[3] *Canadian Institute for Advanced Research*
Toronto, Canada

December 28, 2020

Abstract

We identify necessary and sufficient conditions for a quantum channel to be optimal for any convex optimization problem in which the optimization is taken over the set of all quantum channels of a fixed size. Optimality conditions for convex optimization problems over the set of all quantum measurements of a given system having a fixed number of measurement outcomes are obtained as a special case. In the case of linear objective functions for measurement optimization problems, our conditions reduce to the well-known Holevo–Yuen–Kennedy–Lax measurement optimality conditions. We illustrate how our conditions can be applied to various state transformation problems having non-linear objective functions based on the fidelity, trace distance, and quantum relative entropy.

1 Introduction

Several problems and settings that arise in quantum information theory can be expressed as *optimization problems* in which a real-valued function, defined for a class of quantum channels or measurements, is either minimized or maximized. The problem of *minimum error quantum state discrimination* [BC09], in which a quantum state randomly selected from a known ensemble of states is to be identified with the smallest possible probability of error by means of a measurement, provides a well-known example. This problem is naturally expressed as the optimization of a real-valued linear function defined on the set of all measurements with a fixed number of outcomes. Other examples arise in the study of *quantum cloning* [BDE+98, SIGA05] and the closely related notion of *quantum money* [AFG+12], where one is generally interested in knowing how well an optimally selected quantum channel can transform a single copy of a given state into multiple copies of the same state, with respect to a number of different figures of merit. Another example can be found in quantum complexity theory, in which *two-message quantum interactive proof systems* [JUW09] are naturally analyzed as optimization problems in which the objective function describes the probability that a given verifier accepts, and where the optimization is over all quantum channels of a fixed size, which describe the possible actions of a prover.

Concerning the optimization of *linear functions* defined on the set of all measurements with a fixed number of outcomes, necessary and sufficient conditions for optimality were identified by Holevo [Hol73b, Hol73a] and Yuen, Kennedy, and Lax [YKL70, YKL75]. These conditions, which are described explicitly later in this paper, are relatively easy to check; the problem of actually finding or approximating an optimal measurement, while efficiently solvable through the use of semidefinite programming [JvF02, Ip03, EMV03], is in general a more computationally involved task. These optimality conditions can be easily extended to obtain optimality conditions for real-valued linear functions defined on the set of all quantum channels transforming one quantum system to another.

We prove a generalization of these results to convex optimization problems whose objective functions are not necessarily linear. To be more precise, we consider optimization problems of the form

$$\begin{aligned} &\text{minimize} \;\; f(\Phi) \\ &\text{subject to} \;\; \Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y}), \end{aligned} \tag{1}$$

where

$$f : \mathrm{C}(\mathcal{X}, \mathcal{Y}) \rightarrow \mathbb{R} \cup \{\infty\} \tag{2}$$

is a convex function. Here, $\mathrm{C}(\mathcal{X}, \mathcal{Y})$ denotes the set of all channels (i.e., completely positive and trace-preserving linear maps) from an input system to an output system having associated complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. A channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is said to be *optimal* for the problem (1) if it is the case that $f(\Phi) \leq f(\Psi)$ for all $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$. In this paper we do not consider the difficulty of finding or approximating an optimal channel Φ for a given function f, but instead we focus only on the task of *verifying* that a given channel Φ is indeed optimal. The optimality conditions we obtain can be easily checked for differentiable functions f, and can also be used to verify optimality of channels for some non-differentiable functions. We stress that our optimality conditions are not generic optimality conditions that hold for all convex optimization problems, but rather rely on a specific structure that arises when the optimization is over all quantum channels of a fixed size.

There are, of course, situations in which one would prefer a method to find an optimal channel Φ for a chosen function f, as opposed to simply verifying the optimality of a given Φ, but the task of verifying optimality nevertheless has value for multiple reasons.

For instance, one might hypothesize that a particular channel Φ is optimal based upon an intuition concerning the function f, or upon a heuristic method, making the task of verifying optimality essentially important. The computational task of finding an optimal solution for a chosen function f might also be expensive, time-consuming, or delegated to an untrusted computer, but once this task has been performed the optimality of the solution can be verified, allowing anyone who performs the verification to trust in the optimality of the solution. Finally, there are situations in which the function f could be indeterminate in some respect, eliminating the possibility that a numerical computation could reveal an optimal solution, but potentially allowing for an optimal solution to be expressed and checked analytically. (The results of Bacon, Childs, and van Dam [BCvD05] on hidden subgroup algorithms for certain groups provide a striking example of this potential.)

Our optimality conditions are applicable to convex optimization problems in which the optimization is over all measurements having a fixed number of outcomes, as opposed to being over all channels of a fixed size. This is done through a standard correspondence between measurements and *quantum-to-classical channels*, described in the section following this one. We observe that Holevo [Hol73b, Hol73a] also derived optimality conditions for optimizations over measurements having differentiable (but not necessarily convex) objective functions. Holevo proved that these conditions are necessary for optimality, but did not prove they are sufficient (as they are not sufficient in general). When one restricts their attention to convex objective functions, our optimality conditions are equivalent to a set of intermediate conditions identified by Holevo, but not to the final set of conditions he identified.

We provide a few examples of how our optimality conditions can be applied to interesting categories of optimization problems. As a simple warm-up, we first explain how our conditions imply the Holevo–Yuen–Kennedy–Lax conditions for the optimality of measurements for minimum error state discrimination, which are easily extended to channel optimization problems having linear objective functions. We then discuss optimization problems relating to state transformations having objective functions based on fidelity, trace distance, and quantum relative entropy. We note that optimization problems of the form we consider have recently been studied in the context of *recovery measures* (i.e., the *fidelity of recovery* and generalizations of this measure) [FR15, SW15, BHOS15, CHM$^+$16, BT16, BFT17]. Expositions of this topic can be found in [Sut18] and Chapter 12 of [Wil17].

2 Background and notation

This section summarizes some concepts from convex analysis and optimization theory, narrowly focused on their applications to this paper. Further information on these topics can be found in [Roc70], [BL06], [BV04], and [MN13], for instance. We assume the reader is familiar with quantum information theory, which is covered in the books [NC00], [Wil17], and [Wat18], among others. We will, however, summarize the notion of the Choi operator of a channel, clarify its basic connection to the sorts of optimization problems we consider, and discuss the correspondence between measurements and quantum-to-classical channels, as these topics are essential to an explanation of our results. It will also be helpful to begin the section by establishing some basic notation and terminology concerning linear algebra.

Linear algebra notation and terminology

When we refer to a *complex Euclidean space*, we mean $\mathbb{C}^n$ for some positive integer n, or more generally the complex vector space consisting of vectors indexed by an arbitrary finite set in place of the index set $\{1, \ldots, n\}$. The elementary unit vectors of the space $\mathbb{C}^n$ are denoted $e_1, \ldots, e_n$. Complex Euclidean spaces will be denoted by capital calligraphic letters such as $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$.

For a complex Euclidean space $\mathcal{X}$, the space of linear operators on $\mathcal{X}$ is denoted $\mathrm{L}(\mathcal{X})$, and the identity operator on $\mathcal{X}$ is denoted $\mathbb{1}_{\mathcal{X}}$. For indices $j, k \in \{1, \ldots, n\}$, the operator $E_{j,k} \in \mathrm{L}(\mathbb{C}^n)$ is defined as $E_{j,k} = e_j e_k^*$. Equivalently, with respect to the basis $\{e_1, \ldots, e_n\}$, the matrix representation of $E_{j,k}$ has a 1 in the (j, k) entry, with all other entries 0.

The real vector space of Hermitian operators acting on a complex Euclidean space $\mathcal{X}$ is denoted $\mathrm{Herm}(\mathcal{X})$; the cone of positive semidefinite operators acting on $\mathcal{X}$ is denoted $\mathrm{Pos}(\mathcal{X})$; the set of positive definite operators acting on $\mathcal{X}$ is denoted $\mathrm{Pd}(\mathcal{X})$; and the set of density operators (i.e., positive semidefinite operators having unit trace) is denoted $\mathrm{D}(\mathcal{X})$. The Hilbert–Schmidt inner product of two Hermitian operators $X, Y \in \mathrm{Herm}(\mathcal{X})$ is given by

$$\langle X, Y \rangle = \mathrm{Tr}(XY). \tag{3}$$

For a subspace $\mathcal{V} \subseteq \mathcal{X}$ of a complex Euclidean space $\mathcal{X}$, we write $\Pi_{\mathcal{V}} \in \mathrm{Pos}(\mathcal{X})$ to denote the (orthogonal) projection operator that projects onto the subspace $\mathcal{V}$. Finally, whenever we refer to the *inverse* of a positive semidefinite operator $X \in \mathrm{Pos}(\mathcal{X})$, it should be understood that we are referring to the *Moore–Penrose pseudo-inverse* of X (i.e., the operator that acts as the inverse of X on the image of X and zero on the kernel of X).

Convex functions taking real or infinite values

Let $\mathcal{X}$ be a complex Euclidean space and let

$$f : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R} \cup \{\infty\} \tag{4}$$

be a function mapping each Hermitian operator to either a real number or to positive infinity. The *domain* of f is defined as

$$\mathrm{dom}(f) = \{X \in \mathrm{Herm}(\mathcal{X}) : f(X) \in \mathbb{R}\}. \tag{5}$$

For any function of the form $f : \mathcal{C} \to \mathbb{R}$ defined only on a subset $\mathcal{C} \subseteq \mathrm{Herm}(\mathcal{X})$, one may naturally extend f to a function of the form (4) by defining $f(X) = \infty$ whenever $X \notin \mathcal{C}$.

A function $f : \mathrm{Herm}(\mathcal{X}) \rightarrow \mathbb{R} \cup \{\infty\}$ is *convex* if $\mathrm{dom}(f)$ is a convex set and f is convex on $\mathrm{dom}(f)$. More explicitly, f is convex if for all $X, Y \in \mathrm{dom}(f)$ and $\lambda \in [0,1]$, it is the case that $\lambda X + (1-\lambda)Y \in \mathrm{dom}(f)$ and

$$f(\lambda X + (1-\lambda)Y) \leq \lambda f(X) + (1-\lambda) f(Y). \tag{6}$$

A function f of the form (4) is *proper* if $\mathrm{dom}(f) \neq \emptyset$.

The *indicator function* of a set $\mathcal{C} \subseteq \mathrm{Herm}(\mathcal{X})$ is the function

$$I_{\mathcal{C}} : \mathrm{Herm}(\mathcal{X}) \rightarrow \mathbb{R} \cup \{\infty\} \tag{7}$$

defined as

$$I_{\mathcal{C}}(X) = \begin{cases} 0 & X \in \mathcal{C} \\ \infty & X \notin \mathcal{C} \end{cases} \tag{8}$$

for all $X \in \mathrm{Herm}(\mathcal{X})$. It is evident that $\mathrm{dom}(I_{\mathcal{C}}) = \mathcal{C}$ for every set $\mathcal{C} \subseteq \mathrm{Herm}(\mathcal{X})$, and if $\mathcal{C}$ is a convex set then $I_{\mathcal{C}}$ is a convex function.

Subdifferentials

Let f be a proper function of the form (4) and let $X \in \mathrm{dom}(f)$. The *subdifferential* of f at X is the set defined as

$$\partial f(X) = \{Z \in \mathrm{Herm}(\mathcal{X}) \,:\, f(Y) - f(X) \geq \langle Z, Y - X \rangle \text{ for all } Y \in \mathrm{dom}(f)\}. \tag{9}$$

A key property of the subdifferential of a proper function, which follows trivially from the definition of the subdifferential, is its relation to global minima: $X \in \mathrm{dom}(f)$ is a global minimizer of f if and only if $0 \in \partial f(X)$. Two additional properties of subdifferentials that are relevant to this paper are the following:

1. If f is convex and differentiable at $X \in \mathrm{dom}(f)$ then $\partial f(X) = \{\nabla f(X)\}$.
2. If $\|\cdot\|$ is any norm on $\mathrm{Herm}(\mathcal{X})$ and $f(X) = \|X\|$ for all $X \in \mathrm{Herm}(\mathcal{X})$, then
$$\partial f(X) = \{Y : \langle Y, X \rangle = \|X\|,\ \|Y\|_* \leq 1\}, \tag{10}$$
where $\|Y\|_* = \sup\{\langle Y, X \rangle : \|X\| \leq 1\}$ is the dual norm to $\|\cdot\|$. (See Example 14.4.2 in [Lan13].)

We will make use of the following proposition, which presents a variant of the *chain rule for subdifferentials*. In the statement of this proposition, relint denotes the *relative interior* of a set, and for a linear map $\Lambda : \mathrm{Herm}(\mathcal{Y}) \rightarrow \mathrm{Herm}(\mathcal{X})$, the map

$$\Lambda^* : \mathrm{Herm}(\mathcal{X}) \rightarrow \mathrm{Herm}(\mathcal{Y}) \tag{11}$$

denotes the adjoint map to Λ, which is the uniquely determined linear map that satisfies

$$\langle X, \Lambda(Y) \rangle = \langle \Lambda^*(X), Y \rangle \tag{12}$$

for all $X \in \mathrm{Herm}(\mathcal{X})$ and $Y \in \mathrm{Herm}(\mathcal{Y})$. While the following result can be easily derived from basic rules of convex analysis, we note that this particular result is stated and proved as Theorem 7.2 in [MN17].

Proposition 1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $f : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R} \cup \{\infty\}$ be a convex function, and let $g : \mathrm{Herm}(\mathcal{Y}) \to \mathrm{Herm}(\mathcal{X})$ be an affine linear map, meaning that*

$$g(Y) = \Lambda(Y) + A \tag{13}$$

for all $Y \in \mathrm{Herm}(\mathcal{Y})$, for some choice of a linear map $\Lambda : \mathrm{Herm}(\mathcal{Y}) \to \mathrm{Herm}(\mathcal{X})$ and an operator $A \in \mathrm{Herm}(\mathcal{X})$.

1. *For every $Y \in \mathrm{Herm}(\mathcal{Y})$, it is the case that*

$$\partial(f \circ g)(Y) \supseteq \Lambda^*(\partial f(g(Y))). \tag{14}$$

2. *If $\mathrm{im}(\Lambda) \cap \mathrm{relint}(\mathrm{dom}(f)) \neq \varnothing$, then*

$$\partial(f \circ g)(Y) = \Lambda^*(\partial f(g(Y))) \tag{15}$$

for every $Y \in \mathrm{Herm}(\mathcal{Y})$.

Lastly, we present the subdifferentials of some indicator functions. The subdifferential of the indicator function of a set $\mathcal{C} \subseteq \mathrm{Herm}(\mathcal{X})$ at an operator $X \in \mathrm{Herm}(\mathcal{X})$ may be expressed as

$$\partial I_{\mathcal{C}}(X) = \begin{cases} \{Z \in \mathrm{Herm}(\mathcal{X}) \,:\, \langle Z, X\rangle \geq \langle Z, Y\rangle \text{ for all } Y \in \mathcal{C}\} & \text{if } X \in \mathcal{C} \\ \varnothing & \text{otherwise.} \end{cases} \tag{16}$$

The set described in (16) is sometimes referred to as the *normal cone* of $\mathcal{C}$ at X. (For further discussion of normal cones and subdifferentials of indicator functions, see Section 2.2 in [MN13] and Chapter 23 in [Roc70].) For a positive semidefinite operator $X \in \mathrm{Pos}(\mathcal{X})$ it is straightforward to verify that

$$\partial I_{\mathrm{Pos}(\mathcal{X})}(X) = \{Z \,:\, -Z \in \mathrm{Pos}(\mathcal{X}),\, \langle Z, X\rangle = 0\} = \{Z \,:\, -Z \in \mathrm{Pos}(\mathcal{X}),\, ZX = 0\}. \tag{17}$$

Let $\Lambda : \mathrm{Herm}(\mathcal{X}) \to \mathrm{Herm}(\mathcal{Y})$ be a linear map, let $A \in \mathrm{Herm}(\mathcal{Y})$ be an operator, and define the set

$$\mathcal{K} = \{X \in \mathrm{Herm}(\mathcal{X}) \,:\, \Lambda(X) = A\}. \tag{18}$$

(This is the pre-image of A under Λ.) For an operator $X \in \mathcal{K}$, it holds that

$$\partial I_{\mathcal{K}}(X) = \{Z \in \mathrm{Herm}(\mathcal{X}) \,:\, \langle Z, Y\rangle = 0 \text{ for all } Y \in \ker(\Lambda)\} = \mathrm{im}(\Lambda^*). \tag{19}$$

(See for example Proposition 2.12 in [MN13].)

Convex optimization problems

Convex optimization problems in quantum information theory often have the following general form, for some choice of a convex function $f : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R} \cup \{\infty\}$ and a convex set $\mathcal{C} \subseteq \mathrm{Herm}(\mathcal{X})$:

$$\begin{aligned} &\text{minimize} && f(X) \\ &\text{subject to} && X \in \mathcal{C}. \end{aligned} \tag{20}$$

An operator $X \in \mathcal{C} \cap \mathrm{dom}(f)$ is said to be *optimal* for the optimization problem (20) if $f(X) \leq f(Y)$ for all $Y \in \mathcal{C}$. As we will see below, convenient conditions for optimality can be given if it is the case that

$$\mathrm{relint}\big(\mathrm{dom}(f)\big) \cap \mathrm{relint}(\mathcal{C}) \neq \varnothing. \tag{21}$$

A (constrained) convex optimization problem of the form (20) can be considered as an unconstrained convex optimization problem by minimizing $f(X) + I_{\mathcal{C}}(X)$ over all Hermitian operators $X \in \mathrm{Herm}(\mathcal{X})$. The domain of the function $f + I_{\mathcal{C}}$ is given by

$$\mathrm{dom}(f + I_{\mathcal{C}}) = \mathrm{dom}(f) \cap \mathcal{C}. \tag{22}$$

As was mentioned previously, an operator $X \in \mathrm{dom}(f) \cap \mathcal{C}$ is a global minimizer of $f(X) + I_{\mathcal{C}}(X)$ if and only if $0 \in \partial(f + I_{\mathcal{C}})(X)$. If the condition in (21) holds and both f and $\mathcal{C}$ are convex, a subdifferential sum-rule (see, e.g. Theorem 23.8 of [Roc70]) implies that

$$\partial(f + I_{\mathcal{C}})(X) = \partial f(X) + \partial I_{\mathcal{C}}(X) \tag{23}$$

for all $X \in \mathrm{dom}(f) \cap \mathcal{C}$. For this reason, a characterization of the subdifferential set $\partial I_{\mathcal{C}}(X)$ for a convex set $\mathcal{C}$ can be useful in identifying necessary and sufficient conditions for optimality.

Optimizing over Choi operators of channels

For complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the set of completely positive, trace-preserving linear maps (i.e., channels) from $\mathrm{L}(\mathcal{X})$ to $\mathrm{L}(\mathcal{Y})$ is denoted $\mathrm{C}(\mathcal{X}, \mathcal{Y})$. Given a linear map $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$, its Choi representation $J(\Phi) \in \mathrm{L}(\mathcal{Y} \otimes \mathcal{X})$ is defined as

$$J(\Phi) = \sum_{j,k=1}^{n} \Phi(E_{j,k}) \otimes E_{j,k}, \tag{24}$$

assuming $\mathcal{X} = \mathbb{C}^n$. (An analogous definition is used for index sets other than $\{1, \ldots, n\}$.) Through this representation, the set of channels is isomorphic to the set

$$J(\mathrm{C}(\mathcal{X}, \mathcal{Y})) = \{X \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X}) \,:\, \mathrm{Tr}_{\mathcal{Y}}(X) = \mathbb{1}_{\mathcal{X}}\} \subseteq \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X}). \tag{25}$$

This set is convex, and it is helpful to observe that its relative interior is

$$\mathrm{relint}\big(J(\mathrm{C}(\mathcal{X}, \mathcal{Y}))\big) = \{X \in \mathrm{Pd}(\mathcal{Y} \otimes \mathcal{X}) \,:\, \mathrm{Tr}_{\mathcal{Y}}(X) = \mathbb{1}_{\mathcal{X}}\}. \tag{26}$$

The action of a map Φ can be recovered from its Choi representation by the relation

$$\Phi(X) = \mathrm{Tr}_{\mathcal{X}}((\mathbb{1}_{\mathcal{Y}} \otimes X^{\mathsf{T}})J(\Phi)) \tag{27}$$

for all operators $X \in \mathrm{L}(\mathcal{X})$, where X^{T} denotes the transpose of X.

An optimization problem of the form

$$\begin{aligned} &\text{minimize} \;\; g(\Phi) \\ &\text{subject to} \;\; \Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y}), \end{aligned} \tag{28}$$

where

$$g : \mathrm{C}(\mathcal{X}, \mathcal{Y}) \to \mathbb{R} \cup \{\infty\} \tag{29}$$

is a given function, can equivalently be expressed as

$$\begin{aligned} &\text{minimize} \;\; f(X) \\ &\text{subject to} \;\; X \in J(\mathrm{C}(\mathcal{X}, \mathcal{Y})), \end{aligned} \tag{30}$$

where

$$f : J(\mathrm{C}(\mathcal{X}, \mathcal{Y})) \to \mathbb{R} \cup \{\infty\} \tag{31}$$

is defined as $f(J(\Phi)) = g(\Phi)$ for all $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$. Although the formulations (28) and (30) are equivalent, it will be convenient for us to focus primarily on the formulation (30). The results we obtain can, however, easily be adapted to the formulation (28).

Subdifferentials of the indicator function of the set of Choi operators of channels

The following proposition, which follows from a straightforward application of the rules of subdifferentials, provides a useful characterization of the subdifferential of the indicator function of $J(\mathrm{C}(\mathcal{X}, \mathcal{Y}))$ at every point $X \in J(\mathrm{C}(\mathcal{X}, \mathcal{Y}))$. While we suspect the result is not new, we are not aware of a reference for it, and have included its proof for completeness.

Proposition 2. *For complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and for $X \in J(\mathrm{C}(\mathcal{X}, \mathcal{Y}))$, it holds that*

$$\partial I_{J(\mathrm{C}(\mathcal{X},\mathcal{Y}))}(X) = \{-Y - \mathbb{1}_{\mathcal{Y}} \otimes Z \,:\, Z \in \mathrm{Herm}(\mathcal{X}),\; Y \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X}),\; YX = 0\}. \tag{32}$$

Proof. Define $\mathcal{K} = \{K \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X}) \,:\, \mathrm{Tr}_{\mathcal{Y}}(K) = \mathbb{1}_{\mathcal{X}}\}$. It is the case that

$$J(\mathrm{C}(\mathcal{X}, \mathcal{Y})) = \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X}) \cap \mathcal{K}, \tag{33}$$

and therefore

$$I_{J(\mathrm{C}(\mathcal{X},\mathcal{Y}))} = I_{\mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X})} + I_{\mathcal{K}}. \tag{34}$$

As $\mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$ and $\mathcal{K}$ are both convex and

$$\mathrm{relint}(\mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})) \cap \mathrm{relint}(\mathcal{K}) = \{P \in \mathrm{Pd}(\mathcal{Y} \otimes \mathcal{X}) \,:\, \mathrm{Tr}_{\mathcal{Y}}(P) = \mathbb{1}_{\mathcal{X}}\} \neq \varnothing, \tag{35}$$

one has that

$$\partial I_{J(\mathrm{C}(\mathcal{X},\mathcal{Y}))}(X) = \partial I_{\mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X})}(X) + \partial I_{\mathcal{K}}(X). \tag{36}$$

Finally, from (17) it holds that

$$\partial I_{\mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X})}(X) = \{-Y \,:\, Y \in \mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X}),\ XY = 0\} \tag{37}$$

and from (19) it holds that

$$\partial I_{\mathcal{K}}(X) = \{\mathbb{1}_{\mathcal{Y}} \otimes Z \,:\, Z \in \mathrm{Herm}(\mathcal{X})\}. \tag{38}$$

This implies the proposition. □

The Lagrange dual problem for channel optimization

Consider a channel optimization problem expressed in the following form that is equivalent to (30):

$$\begin{aligned} \text{minimize}\ \ & f(X) \\ \text{subject to}\ \ & \mathrm{Tr}_{\mathcal{Y}}(X) = \mathbb{1}_{\mathcal{X}}, \\ & X \in \mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X}). \end{aligned} \tag{39}$$

One may then formulate the associated Lagrange dual problem:

$$\begin{aligned} \text{maximize}\ \ & g(Y,Z) \\ \text{subject to}\ \ & Y \in \mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X}), \\ & Z \in \mathrm{Herm}(\mathcal{X}), \end{aligned} \tag{40}$$

where

$$g(Y,Z) = \mathrm{Tr}(Z) + \inf_{X\in\mathrm{Herm}(\mathcal{Y}\otimes\mathcal{X})} \bigl(f(X) - \langle X, Y + \mathbb{1}_{\mathcal{Y}} \otimes Z\rangle\bigr) \tag{41}$$

for all $Y \in \mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X})$ and $Z \in \mathrm{Herm}(\mathcal{X})$.

The optimal value of the Lagrange dual problem (40) is a lower-bound for the optimal value of the original problem (39) (even if the function f is not convex), which is a property known as weak duality. *Slater's theorem* implies that if f is convex, and there exists a positive definite operator $X \in \mathrm{Pd}(\mathcal{Y}\otimes\mathcal{X})$ such that $\mathrm{Tr}_{\mathcal{Y}}(X) = \mathbb{1}_{\mathcal{X}}$ and $X \in \mathrm{relint}(\mathrm{dom}(f))$, then the problems (39) and (40) have the same optimal value—a property known as *strong duality*.

Quantum measurements as channels

Optimizations over the set of all measurements having a fixed number of outcomes can be expressed as optimizations over channels, as we now explain. Consider first a measurement on a complex Euclidean space $\mathcal{X}$ that has m possible measurement outcomes, and is described by measurement operators $\{P_1, \ldots, P_m\} \subseteq \text{Pos}(\mathcal{X})$. Letting $\mathcal{Y} = \mathbb{C}^m$, this measurement can be represented by the channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ defined as

$$\Phi(X) = \sum_{k=1}^{m} \langle P_k, X \rangle \, E_{k,k} \tag{42}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Any channel expressible in this form is called a *quantum-to-classical channel*. An equivalent condition to a channel taking the form (42) is that its Choi operator takes the form

$$J(\Phi) = \sum_{k=1}^{m} E_{k,k} \otimes P_k^{\mathsf{T}}. \tag{43}$$

Next, for each $k \in \{1, \ldots, m\}$, define a linear map $\Xi_k : \text{Herm}(\mathcal{Y} \otimes \mathcal{X}) \to \text{Herm}(\mathcal{X})$ as

$$\Xi_k(X) = \big((e_k^* \otimes \mathbb{1}_{\mathcal{X}}) X (e_k \otimes \mathbb{1}_{\mathcal{X}})\big)^{\mathsf{T}}, \tag{44}$$

and observe that one has

$$P_k = \Xi_k(J(\Phi)) \tag{45}$$

for the quantum-to-classical channel Φ given by (42). In words, for a measurement represented by operators $\{P_1, \ldots, P_m\}$, the linear map Ξ_k allows for the recovery of the operator P_k from the Choi operator $J(\Phi)$ of the quantum-to-classical channel associated with that measurement. A function $g(P_1, \ldots, P_m)$ of these measurement operators can therefore be expressed as a function

$$f(J(\Phi)) = g\big(\Xi_1(J(\Phi)), \ldots, \Xi_m(J(\Phi))\big) \tag{46}$$

of the Choi operator $J(\Phi)$. Note that if $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is an arbitrary (i.e., not necessarily quantum-to-classical) channel, then the operators $P_1, \ldots, P_m$ defined by (45) will still necessarily satisfy $P_1, \ldots, P_m \in \text{Pos}(\mathcal{X})$ and $P_1 + \cdots + P_m = \mathbb{1}_{\mathcal{X}}$, and therefore represent a valid measurement. (In general, the same measurement is given by a continuum of channels Φ, including the quantum-to-classical channel described earlier.) An optimization problem of the form

$$\begin{aligned} \text{minimize} \quad & g(P_1, \ldots, P_m) \\ \text{subject to} \quad & P_1, \ldots, P_m \in \text{Pos}(\mathcal{X}) \\ & P_1 + \cdots + P_m = \mathbb{1}_{\mathcal{X}} \end{aligned} \tag{47}$$

can therefore be expressed equivalently as follows:

$$\begin{aligned} \text{minimize} \quad & f(X) \\ \text{subject to} \quad & X \in J(\mathrm{C}(\mathcal{X}, \mathcal{Y})) \end{aligned} \tag{48}$$

for f defined from g as in (46).

3 Optimality conditions for convex channel optimization

In this section, we present our main general result regarding optimality conditions for convex optimization problems over quantum channels. As suggested in the previous section, it is convenient to associate quantum channels with their Choi representations, and to consider optimization problems of the form

$$\begin{array}{ll} \text{minimize} & f(X) \\ \text{subject to} & X \in J(\mathrm{C}(\mathcal{X},\mathcal{Y})) \end{array} \tag{49}$$

for various choices of convex functions

$$f : \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X}) \rightarrow \mathbb{R} \cup \{\infty\}. \tag{50}$$

Optimality conditions for such problems can be translated to optimality conditions for problems of the form (1), as will be illustrated in the section following this one.

Theorem 3. *Let* $f : \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X}) \rightarrow \mathbb{R} \cup \{\infty\}$ *be a convex function such that*

$$\mathrm{relint}\big(\mathrm{dom}(f)\big) \cap \mathrm{relint}\big(J\big(\mathrm{C}(\mathcal{X},\mathcal{Y})\big)\big) \neq \emptyset, \tag{51}$$

and let $X \in J\big(\mathrm{C}(\mathcal{X},\mathcal{Y})\big)$ *be the Choi representation of a channel such that* $X \in \mathrm{dom}(f)$*. The following statements are equivalent:*

1. *The operator* X *is optimal for the optimization problem* (49)*.*
2. *There exists an operator* $H \in \partial f(X)$ *satisfying*

$$\mathrm{Tr}_{\mathcal{Y}}(HX) \in \mathrm{Herm}(\mathcal{X}) \quad \textit{and} \quad H \geq \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HX). \tag{52}$$

Proof. An operator X is optimal for the problem (49) if and only if

$$0 \in \partial(f + I_{J(\mathrm{C}(\mathcal{X},\mathcal{Y}))})(X) = \partial f(X) + \partial I_{J(\mathrm{C}(\mathcal{X},\mathcal{Y}))}(X). \tag{53}$$

By the characterization of the subdifferential of the indicator function $I_{J(\mathrm{C}(\mathcal{X},\mathcal{Y}))}$ given by Proposition 2, it follows that X is optimal for the optimization problem (49) if and only if there exist operators $Y \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$, $Z \in \mathrm{Herm}(\mathcal{X})$, and $H \in \partial f(X)$ satisfying

$$YX = 0 \quad \text{and} \quad H - Y - \mathbb{1}_{\mathcal{Y}} \otimes Z = 0. \tag{54}$$

Assume first that statement 2 holds, and define $Y \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$ and $Z \in \mathrm{Herm}(\mathcal{X})$ as

$$Y = H - \mathbb{1}_{\mathcal{Y}} \otimes Z \quad \text{and} \quad Z = \mathrm{Tr}_{\mathcal{Y}}(HX). \tag{55}$$

As X and Y are both positive operators and

$$\langle Y, X \rangle = \langle H - \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HX), X \rangle = \langle H, X \rangle - \langle \mathrm{Tr}_{\mathcal{Y}}(HX), \mathbb{1}_{\mathcal{X}} \rangle = 0, \tag{56}$$

it follows that $YX = 0$. The conditions in (54) are therefore satisfied, implying that statement 1 holds.

Now assume that statement 1 holds, so that there exist operators $Y \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$, $Z \in \mathrm{Herm}(\mathcal{X})$, and $H \in \partial f(X)$ such that $YX = 0$ and $H - Y - \mathbb{1}_{\mathcal{Y}} \otimes Z = 0$. It follows that $HX = (\mathbb{1}_{\mathcal{Y}} \otimes Z)X$ and therefore

$$\mathrm{Tr}_{\mathcal{Y}}(HX) = \mathrm{Tr}_{\mathcal{Y}}\big((\mathbb{1}_{\mathcal{Y}} \otimes Z)X\big) = Z\,\mathrm{Tr}_{\mathcal{Y}}(X) = Z, \tag{57}$$

which is a Hermitian operator. Moreover, one has $H - \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HX) = Y$, which is positive semidefinite by assumption. The conditions in (52) are therefore satisfied, which implies that statement 2 holds, completing the proof. □

To make use of Theorem 3, one requires that the relative interior of the domain of the objective function contains at least one point in the relative interior of the set of channels. This requirement is precisely Slater's condition for this optimization problem, which guarantees strong duality with the corresponding dual problem (40). This regularity condition is automatically satisfied if f is continuous at at least one point in the set of Choi representations of channels. In particular, if f is differentiable at X then one may take the operator H in Theorem 3 to be $H = \nabla f(X)$.

Corollary 4. *Let $f : \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X}) \to \mathbb{R} \cup \{\infty\}$ be a convex function, let $X \in J(\mathrm{C}(\mathcal{X}, \mathcal{Y}))$ be the Choi representation of a channel, and assume that f is differentiable at X. The operator X is an optimal solution to the optimization problem (49) if and only if*

$$\mathrm{Tr}_{\mathcal{Y}}(\nabla f(X)X) \in \mathrm{Herm}(\mathcal{X}) \quad \textit{and} \quad \nabla f(X) \geq \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(\nabla f(X)X). \tag{58}$$

Proof. As f is differentiable at X, one has that $\partial f(X) = \{\nabla f(X)\}$. Furthermore, f must be finite in some neighborhood around X and thus $\mathrm{relint}(\mathrm{dom}(f)) \cap \mathrm{relint}(J(\mathrm{C}(\mathcal{X}, \mathcal{Y}))) \neq \emptyset$ must hold. The result now follows from Theorem 3. □

We remark that the optimality conditions represented by this corollary appear to be a special feature of problems involving an optimization over channels. In essence, for differentiable convex quantum channel optimization problems, every optimal primal solution X determines an optimal dual solution consisting of operators $Z = \mathrm{Tr}_{\mathcal{Y}}(\nabla f(X)X)$ and $Y = \nabla f(X) - \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(\nabla f(X)X)$.

It is natural to ask if there is an approximate version of the implication that statement 2 implies statement 1 in Theorem 3. That is, if the requirements (52) hold approximately for some $H \in \partial f(X)$, then is X necessarily close to being optimal? The following theorem demonstrates that this is indeed the case, which allows for bounds to be placed on the optimal value of such optimization problems in the case the conditions in (52) cannot be verified exactly.

Theorem 5. *Let $f : \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X}) \to \mathbb{R} \cup \{\infty\}$ be a function, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel such that $J(\Phi) \in \mathrm{dom}(f)$, and let $H \in \partial f(X)$. It is the case that*

$$f(J(\Phi)) \leq \inf_{\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})} f(J(\Psi)) + \varepsilon \dim(\mathcal{X}), \tag{59}$$

where

$$\varepsilon = \inf_{P \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})} \big\| H - \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) - P \big\|_{\infty}. \tag{60}$$

Proof. Because the spectral norm is a continuous function, a compactness argument implies that there must exist a positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$ for which

$$\varepsilon = \big\| H - \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) - P \big\|_{\infty}. \tag{61}$$

We will let such a P be fixed for the remainder of the proof.

Define a Hermitian operator

$$Z = \frac{1}{2} \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) + \frac{1}{2} \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi))^{*} - \varepsilon \mathbb{1}_{\mathcal{X}}, \tag{62}$$

and define

$$Y = H - \mathbb{1}_{\mathcal{Y}} \otimes Z. \tag{63}$$

Also define

$$A = H - \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) - P, \tag{64}$$

so that $\|A\|_{\infty} = \varepsilon$, and therefore

$$\left\| \frac{1}{2} A + \frac{1}{2} A^{*} \right\|_{\infty} \leq \varepsilon. \tag{65}$$

It is the case that

$$Y = P + \frac{A + A^{*}}{2} + \varepsilon \mathbb{1}_{\mathcal{Y}} \otimes \mathbb{1}_{\mathcal{X}} \geq P, \tag{66}$$

and therefore $Y \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$.

Now consider the Lagrange dual problem (40) associated with the minimization of f over all channels in $\mathrm{C}(\mathcal{X}, \mathcal{Y})$. As Z is Hermitian and Y is positive semidefinite, (Y, Z) is a dual feasible solution to this dual problem. As $H \in \partial f(J(\Phi))$, one has that

$$f(X) - \langle H, X \rangle \geq f(J(\Phi)) - \langle H, J(\Phi) \rangle \tag{67}$$

for every $X \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$. Therefore, when the dual objective function g is evaluated at (Y, Z), we find that

$$\begin{aligned} g(Y, Z) &= \mathrm{Tr}(Z) + \inf_{X \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})} \big(f(X) - \big\langle X, Y + \mathbb{1}_{\mathcal{Y}} \otimes Z \big\rangle \big) \\ &= \langle H, J(\Phi) \rangle - \varepsilon \dim(\mathcal{X}) + \inf_{X \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})} \big(f(X) - \langle H, X \rangle \big) \\ &\geq f(J(\Phi)) - \varepsilon \dim(\mathcal{X}). \end{aligned} \tag{68}$$

The theorem follows by weak duality. □

We note that the previous theorem does not require the function f to be convex, or for the associated optimization problem to satisfy the conditions of Slater's theorem. However, having knowledge of the subdifferential $\partial f(X)$ at an operator X requires knowledge of the global behavior of the function f when f is not convex, so the usefulness of Theorem 5 may be limited when f is not convex.

Remark 6. Theorem 5 implies that every feasible primal solution X to the optimization problem in (30) for which $\partial f(X)$ is non-empty provides both an upper bound *and* a lower bound to the optimal value. Indeed, from (59) one has that

$$f(X) - \varepsilon \leq \inf_{Y \in J(\mathrm{C}(\mathcal{X},\mathcal{Y}))} f(Y) \leq f(X) \tag{69}$$

provided that one can compute the value of ε as defined in (60) for some $H \in \partial f(X)$.

4 Applications

In this section we apply the optimality conditions given by Theorem 3 to a few categories of examples, including the simple case of channel optimization problems having linear objective functions and three variants of problems involving quantum state transformations. In these examples we will make use of various facts concerning differentiation for functions mapping between spaces of Hermitian operators; a short discussion of this topic, along with a lemma that is needed for one of the examples, can be found in Appendix A. Additionally, we remark that Theorem 7 can be used to prove that the generalized fidelity of recovery is multiplicative.

4.1 Linear objective functions

We will begin by considering the simple case in which the objective function f in Theorem 3 is linear. In this situation, the optimization problem (49) may be rewritten as

$$\begin{aligned} &\text{minimize} && \langle H, X \rangle \\ &\text{subject to} && X \in J(\mathrm{C}(\mathcal{X},\mathcal{Y})) \end{aligned} \tag{70}$$

for some choice of a Hermitian operator $H \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$. The subdifferential of the function $f(X) = \langle H, X \rangle$ is given by $\partial f(X) = \{H\}$ for all $X \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$. By Theorem 3 it follows that the Choi operator $X = J(\Phi)$ of a channel $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ is optimal for the problem (70) if and only if

$$\mathrm{Tr}_{\mathcal{Y}}(HX) \in \mathrm{Herm}(\mathcal{X}) \quad \text{and} \quad H \geq \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HX). \tag{71}$$

We observe that this optimality criterion can alternatively be obtained through semidefinite programming duality and complementary slackness. (See, for instance, Exercise 3.5

of [Wat18], observing that the inequality is reversed in that exercise because the optimization problem is expressed as a maximization rather than a minimization.)

The problem of *minimum error state discrimination*, which was mentioned in the introduction, is a special case in which the function f in the optimization problem (49) is linear. Consider an ensemble of states, which represents the random selection of one of a finite number of quantum states according to a given probability distribution. Formally speaking, an ensemble is described by a collection $\{\rho_1, \ldots, \rho_n\} \subseteq \mathrm{D}(\mathcal{X})$ of density operators together with a probability vector $p = (p_1, \ldots, p_n)$. The problem of minimum-error state discrimination seeks a measurement on the system represented by the space $\mathcal{X}$ that identifies, with the minimum possible probability of error, a state chosen randomly according to this ensemble.

For a given choice of a measurement, represented by operators $\{P_1, \ldots, P_n\} \subseteq \mathrm{Pos}(\mathcal{X})$, the error probability incurred by this measurement can be expressed as

$$\sum_{k=1}^{n} \langle \rho - p_k \rho_k, P_k \rangle, \tag{72}$$

where

$$\rho = \sum_{k=1}^{n} p_k \rho_k \tag{73}$$

is the average state of the ensemble. A minimization of the error probability (72) over all measurements $\{P_1, \ldots, P_n\}$ can be represented as an optimization of the form (70) by letting $\mathcal{Y} = \mathbb{C}^n$ and setting

$$H = \sum_{k=1}^{n} E_{k,k} \otimes (\rho - p_k \rho_k)^{\mathsf{T}}. \tag{74}$$

The measurement described by $\{P_1, \ldots, P_n\}$, which may alternatively be represented by a quantum-to-classical channel whose Choi operator is defined as

$$X = \sum_{k=1}^{n} E_{k,k} \otimes P_k^{\mathsf{T}}, \tag{75}$$

is optimal for the minimization of the error probability (72) if and only of the conditions (71) hold. These conditions can be simplified by first calculating that

$$\mathrm{Tr}_{\mathcal{Y}}(HX) = \sum_{k=1}^{n} (\rho - p_k \rho_k)^{\mathsf{T}} P_k^{\mathsf{T}} = \rho^{\mathsf{T}} - \sum_{k=1}^{n} p_k \rho_k^{\mathsf{T}} P_k^{\mathsf{T}}, \tag{76}$$

then observing that this operator is Hermitian if and only if the operator

$$\sum_{k=1}^{n} p_k P_k \rho_k \tag{77}$$

is Hermitian, and finally noting that the inequality $H \geq \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HX)$ is equivalent to

$$\sum_{j=1}^{n} E_{j,j} \otimes p_j \rho_j \leq \mathbb{1}_{\mathcal{Y}} \otimes \sum_{k=1}^{n} p_k P_k \rho_k, \tag{78}$$

which may alternatively be expressed as

$$\sum_{k=1}^{n} p_k P_k \rho_k \geq p_j \rho_j \quad \text{(for all } j = 1, \ldots, n). \tag{79}$$

The conditions that the operator (77) is Hermitian and satisfies the inequalities (79) comprise the Holevo–Yuen–Kennedy–Lax conditions for measurement optimality [Hol73b, Hol73a, YKL70, YKL75].

4.2 State transformation

The second category of channel optimization problems we consider involves channels that transform one state into another in such a way that the distance to a target state is minimized. One may consider any number of specific measures of distance in such a problem; we will analyze measures based on the fidelity, trace distance, and quantum relative entropy. For each of these measures, we will consider the situation in which two bipartite states, $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ and $\sigma \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, are given. The optimization problem to be considered is to minimize the distance (or maximize the similarity) between the states $\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)$ and σ, with respect to the measure under consideration, over all possible channels $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$.

Optimization problems of this sort were suggested in [SW15] (see Remark 6) in the special case that $\mathcal{Y} = \mathcal{X} \otimes \mathcal{W}$ and $\rho = \mathrm{Tr}_{\mathcal{W}}(\sigma)$, so that the aim of the channel Φ is to "recover" the original state σ from the portion of it represented by ρ. The resulting quantities are known as the *fidelity of recovery*, *relative entropy of recovery*, and so on. These and related measures have been studied in a number of recent papers (such as [FR15, BHOS15, CHM+16, BT16, BFT17]), and the more general situation in which ρ and σ are arbitrary states arises naturally in this study. (For example, [BT16] uses the term *generalized fidelity of recovery* in the situation mentioned above in the case of the fidelity.)

As a further application of the methods introduced in this paper, it is straightforward to show (by applying the result of Theorem 7) that the generalized fidelity of recovery is multiplicative. This fact has already been shown in [BT16], where it was shown by examining the fidelity of recovery as a semidefinite program. A detailed proof using Theorem 7 is presented in Appendix B.

When analyzing optimality conditions for these problems, it will be helpful to refer to the *evaluation map* corresponding to the operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$. This is the uniquely determined completely positive map $\Psi_\rho \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ that satisfies

$$\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho\big)(J(\Phi)) = \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho) \tag{80}$$

for every complex Euclidean space $\mathcal{Y}$ and every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$. The relationship between the map Ψ_ρ and the state ρ is closely related to the Choi representation of maps: assuming for the moment that $\mathcal{X} = \mathbb{C}^n$, one has that

$$\rho = \sum_{j,k=1}^{n} E_{j,k} \otimes \Psi_\rho(E_{j,k}). \tag{81}$$

That is, up to swapping the tensor factors corresponding to $\mathcal{X}$ and $\mathcal{Z}$, the state ρ is the Choi operator of the map Ψ_ρ.

Objective functions based on the fidelity

For positive semidefinite operators $P, Q \in \mathrm{Pos}(\mathcal{X})$, one defines the fidelity between P and Q as

$$\mathrm{F}(P, Q) = \left\| \sqrt{P}\sqrt{Q} \right\|_1 = \mathrm{Tr}\sqrt{\sqrt{P}Q\sqrt{P}}. \tag{82}$$

The first variant of the optimal state transformation problem we will consider is as follows:

$$\begin{aligned} &\text{minimize} \quad -\mathrm{F}\big(\sigma, (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\big) \\ &\text{subject to} \quad \Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y}). \end{aligned} \tag{83}$$

The fidelity function is jointly concave (see for example Corollary 3.26 in [Wat18]), and therefore is concave in each of its arguments, from which it follows that this problem is a convex optimization problem. It is possible to express this optimization problem as a semidefinite program, as is demonstrated in [BFT17].

The following theorem establishes optimality conditions for a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ in the optimization problem (83) under the assumption that the operator $\mathrm{Tr}_{\mathcal{X}}(\rho)$ is positive definite. We note that the theorem statement does not actually require ρ and σ to have unit trace—they can be arbitrary positive semidefinite operators, but we nevertheless use the letters ρ and σ to make the connection to the optimization problem (83) clear.

Theorem 7. *Let $\rho \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z})$ and $\sigma \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$ be positive semidefinite operators, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and assume that $\mathrm{Tr}_{\mathcal{X}}(\rho)$ is a positive definite operator. A channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is optimal for the optimization problem (83) if and only if the following two conditions are met:*

1. $\mathrm{im}(\sigma) \subseteq \mathrm{im}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\big)$.
2. *The operator*

$$H = -\frac{1}{2}(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*)\Big(\sqrt{\sigma}\big(\sqrt{\sigma}\,(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\sqrt{\sigma}\big)^{-\frac{1}{2}}\sqrt{\sigma}\Big) \tag{84}$$

satisfies

$$\mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) \in \mathrm{Herm}(\mathcal{X}) \quad \text{and} \quad H \geq \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)). \tag{85}$$

(As per the convention mentioned in Section 2, the inverse in (84) refers to the Moore–Penrose pseudo-inverse in case σ does not have full rank.)

Remark 8. The theorem assumes that $\mathrm{Tr}_{\mathcal{X}}(\rho)$ has full rank, as this assumption allows for a cleaner theorem statement. It is, however, straightforward to apply the theorem to a situation in which $\mathrm{Tr}_{\mathcal{X}}(\rho)$ does not have full rank. Specifically, for an arbitrary choice of ρ and σ, one may take $B \in \mathrm{L}(\mathcal{V}, \mathcal{Z})$ to be an isometry for which BB^* is the projection onto the image of $\mathrm{Tr}_{\mathcal{X}}(\rho)$, and then observe that by replacing ρ and σ with $(\mathbb{1}_{\mathcal{X}} \otimes B^*)\rho(\mathbb{1}_{\mathcal{X}} \otimes B)$ and $(\mathbb{1}_{\mathcal{Y}} \otimes B^*)\sigma(\mathbb{1}_{\mathcal{Y}} \otimes B)$, respectively, an equivalent problem is obtained that satisfies the assumptions of the theorem.

Remark 9. The argument of the expression in (84) has the form of an *operator geometric mean*, which is defined as follows. For positive definite operators $P, Q \in \mathrm{Pd}(\mathcal{X})$ on a complex Euclidean space $\mathcal{X}$, their geometric mean is the operator $P \,\#\, Q \in \mathrm{Pd}(\mathcal{X})$ defined as

$$P \,\#\, Q = \sqrt{P}\Big(\sqrt{P^{-1}}Q\sqrt{P^{-1}}\Big)^{\frac{1}{2}}\sqrt{P}.$$

If one takes $A \in \mathrm{L}(\mathcal{W}, \mathcal{Y} \otimes \mathcal{Z})$ to be an isometry for which $AA^* = \Pi_{\mathrm{im}(\sigma)}$, the operator H in (84) may be expressed as

$$H = -\frac{1}{2}(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*)\Big(A\Big((A^*\sigma A) \,\#\, (A^*YA)^{-1}\Big)A^*\Big) \tag{86}$$

where $Y = (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)$. If the operator σ is positive definite (in which case $A = \mathbb{1}_{\mathcal{Y}\otimes\mathcal{Z}}$), this expression simplifies to

$$H = -\frac{1}{2}(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*)\big(\sigma \,\#\, Y^{-1}\big). \tag{87}$$

For further discussion of the operator geometric mean see Section 4.1 in [Bha07].

Proof of Theorem 7. Let $r = \mathrm{rank}(\sigma)$, let $\mathcal{W} = \mathbb{C}^r$, and let $A \in \mathrm{L}(\mathcal{W}, \mathcal{Y} \otimes \mathcal{Z})$ be any isometry for which $AA^* = \Pi_{\mathrm{im}(\sigma)}$ (the projection onto the image of σ). Define a function

$$g : \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{Z}) \rightarrow \mathbb{R} \cup \{\infty\} \tag{88}$$

as

$$g(Y) = \begin{cases} -\,\mathrm{F}(A^*\sigma A, A^*YA) & \text{if } A^*YA \in \mathrm{Pos}(\mathcal{W}) \\ \infty & \text{otherwise,} \end{cases} \tag{89}$$

for all $Y \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{Z})$, and observe that $g(Y) = -\,\mathrm{F}(\sigma, Y)$ for every $Y \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$.

For a given operator $Y \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{Z})$ satisfying $A^*YA \in \mathrm{Pos}(\mathcal{W})$, there are two cases for the subdifferential $\partial g(Y)$.

Case 1: A^*YA is positive definite. In this case g is differentiable at Y, and

$$\nabla g(Y) = -\frac{1}{2}A\left((A^*\sigma A) \,\#\, (A^*YA)^{-1}\right)A^* \tag{90}$$

which follows from Lemma 13 (as stated and proved in Appendix A). It therefore follows that

$$\partial g(Y) = \left\{-\frac{1}{2}A\left((A^*\sigma A) \,\#\, (A^*YA)^{-1}\right)A^*\right\}. \tag{91}$$

Case 2: A^*YA is not positive definite. In this case, $\partial g(Y) = \varnothing$, which also follows from Lemma 13.

Next, define

$$\Lambda = \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho, \tag{92}$$

and observe that

$$(g \circ \Lambda)(J(\Phi)) = -\mathrm{F}\big(\sigma, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)\big) \tag{93}$$

for every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$. It is the case that $\Lambda(\mathbb{1}_\mathcal{Y} \otimes \mathbb{1}_\mathcal{X}) = \mathbb{1}_\mathcal{Y} \otimes \mathrm{Tr}_\mathcal{X}(\rho)$, which is positive definite by assumption. As $\mathrm{Pd}(\mathcal{Y} \otimes \mathcal{Z}) \subset \mathrm{relint}(\mathrm{dom}(g))$, it follows that

$$\mathrm{im}(\Lambda) \cap \mathrm{relint}(\mathrm{dom}(g)) \neq \varnothing, \tag{94}$$

and therefore, by Proposition 1,

$$\partial(g \circ \Lambda)(X) = \Lambda^*(\partial g(\Lambda(X))) \tag{95}$$

for every $X \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$.

The theorem now follows from Theorem 3. In greater detail, if Φ is optimal for the problem (83), there must exist an operator $H \in \partial(g \circ \Lambda)(J(\Phi))$ such that

$$\mathrm{Tr}_\mathcal{Y}(HJ(\Phi)) \in \mathrm{Herm}(\mathcal{X}) \quad \text{and} \quad H \geq \mathbb{1}_\mathcal{Y} \otimes \mathrm{Tr}_\mathcal{Y}(HJ(\Phi)). \tag{96}$$

Case 1 described above must therefore hold when $Y = \Lambda(J(\Phi))$, for otherwise the subdifferential $\partial(g \circ \Lambda)(J(\Phi))$ would be empty. It follows that

$$H = -\frac{1}{2}(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*)\big(A\big((A^*\sigma A) \,\#\, (A^*(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)A)^{-1}\big)A^*\big) \tag{97}$$

and straightforward manipulation reveals that

$$H = -\frac{1}{2}(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*)\Big(\sqrt{\sigma}\big(\sqrt{\sigma}\,(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\sqrt{\sigma}\big)^{-\frac{1}{2}}\sqrt{\sigma}\Big) \tag{98}$$

(where, as always, we interpret the inverse as the Moore–Penrose pseudoinverse in the case when the given operators are not positive definite). The second condition in the statement of the theorem now follows.

Conversely, if the two conditions in the statement of the theorem hold, it follows that $H \in \partial(g \circ \Lambda)(J(\Phi))$, and moreover that

$$\mathrm{Tr}_\mathcal{Y}(HJ(\Phi)) \in \mathrm{Herm}(\mathcal{X}) \quad \text{and} \quad H \geq \mathbb{1}_\mathcal{Y} \otimes \mathrm{Tr}_\mathcal{Y}(HJ(\Phi)). \tag{99}$$

One concludes by Theorem 3 that Φ is optimal for the optimization problem (83). □

An interesting special case of the optimization problem (83) is when the states ρ and σ take the form

$$\rho = \sum_{k=1}^{n} p_k \rho_k \otimes E_{k,k} \quad \text{and} \quad \sigma = \sum_{k=1}^{n} p_k \sigma_k \otimes E_{k,k} \tag{100}$$

for a probability vector $(p_1, \ldots, p_n)$ and collections of states $\{\rho_1, \ldots, \rho_n\} \subseteq \mathrm{D}(\mathcal{X})$ and $\{\sigma_1, \ldots, \sigma_n\} \subseteq \mathrm{D}(\mathcal{Y})$. The objective function simplifies in this case to

$$-\sum_{k=1}^{n} p_k \, \mathrm{F}(\sigma_k, \Phi(\rho_k)), \tag{101}$$

and therefore the optimization concerns the average fidelity with which a channel Φ maps each ρ_k to σ_k. The operator

$$H = -\frac{1}{2}\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*\big)\Big(\sqrt{\sigma}\Big(\sqrt{\sigma}\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)\sqrt{\sigma}\Big)^{-\frac{1}{2}}\sqrt{\sigma}\Big) \tag{102}$$

simplifies in this case to

$$H = -\frac{1}{2}\sum_{k=1}^{n} p_k \sqrt{\sigma_k}\Big(\sqrt{\sigma_k}\Phi(\rho_k)\sqrt{\sigma_k}\Big)^{-\frac{1}{2}}\sqrt{\sigma_k} \otimes \rho_k^{\mathsf{T}}. \tag{103}$$

An analogous criterion for optimality can also be found when the figure of merit is the square of the fidelity, which is also jointly concave (see for example Property 9.2.2 in [Wil17]). The corresponding optimization problem is the following:

$$\begin{aligned} &\text{minimize} \quad -\sum_{k=1}^{m} p_k \, \mathrm{F}\big(\sigma_k, \Phi(\rho_k)\big)^2 \\ &\text{subject to} \quad \Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y}). \end{aligned} \tag{104}$$

By differentiating the objective function in (104), one finds that a channel Φ is optimal precisely when the same conditions in Theorem 7 are met, but with $H \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$ given by

$$H = -\sum_{k=1}^{n} p_k \, \mathrm{F}\big(\sigma_k, \Phi(\rho_k)\big)\sqrt{\sigma}\Big(\sqrt{\sigma}\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)\sqrt{\sigma}\Big)^{-\frac{1}{2}}\sqrt{\sigma} \otimes \rho_k^{\mathsf{T}}. \tag{105}$$

Objective functions based on trace distance

Next we consider an optimization problem that is analogous to (83), but based on the trace distance rather than the fidelity:

$$\begin{aligned} &\text{minimize} \quad \big\| \sigma - \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho) \big\|_1 \\ &\text{subject to} \quad \Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y}). \end{aligned} \tag{106}$$

The objective function of this optimization problem is convex (with respect to Φ), so we may use Theorem 3 to obtain optimality conditions for this problem. We note that, similar to its variant based on the fidelity described above, this optimization problem can be represented as a semidefinite program.

For this optimization problem, the optimality conditions we obtain may not be efficiently checkable. We do, however, obtain an efficiently checkable condition that is sufficient for optimality, and we conjecture that this condition is also a necessary for optimality.

Corollary 10. *Let $\rho \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z})$ and $\sigma \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$ be positive semidefinite operators, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The channel Φ is optimal for the state transformation problem in (106) if and only if there exists an operator $Y \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{Z})$ with $\|Y\|_\infty = 1$ such that the following conditions are satisfied:*

1. *It is the case that $\left\| \sigma - (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho) \right\|_1 = \left\langle Y, \sigma - (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho) \right\rangle$.*
2. *The operator*
$$H = \left(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^* \right)(Y) \tag{107}$$
satisfies
$$\mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) \in \mathrm{Herm}(\mathcal{X}) \quad \textit{and} \quad H \geq \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}\big(HJ(\Phi)\big). \tag{108}$$

Corollary 10 follows directly from Theorem 3 and applying the rules of subdifferentiation presented in Section 2. Indeed, operators of the form in (107) are precisely the elements of the subdifferential of the objective function in (106) at $J(\Phi)$. As a generalization, one can replace the trace norm $\|\cdot\|_1$ in the statement of Corollary 10 with any other norm on operators, and replace $\|\cdot\|_\infty$ with the corresponding dual norm.

In the event that the operator $\sigma - (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)$ arising in Corollary 10 has no zero eigenvalues, there is a unique choice of Y for which the first condition of the theorem holds. Specifically, if

$$\sigma - (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho) = \sum_{k=1}^{m} \lambda_k \Pi_k \tag{109}$$

is a spectral decomposition where each λ_k is nonzero, then the unique operator Y satisfying condition 1 in Corollary 10 is given by

$$Y = \sum_{k=1}^{m} \mathrm{sign}(\lambda_k)\, \Pi_k. \tag{110}$$

In this case it is sufficient for the second condition to be checked for this unique choice of Y, yielding an efficiently checkable optimality criterion. However, if it is the case that $\sigma - (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)$ has one or more zero eigenvalues, then the first condition holds for a continuum of choices of Y, and from Corollary 10 we conclude only that the optimality of Φ is equivalent to the existence of at least one such choice of Y for which the second statement in the theorem hold.

It is reasonable, though, to view the operator Y defined by (110), where now it is to be understood that $\mathrm{sign}(0) = 0$, as a natural selection of an operator through which optimality may be verified. We conjecture, based on numerical evidence, that this choice yields an efficiently checkable necessary and sufficient optimality condition.

Conjecture 11. *Let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ and $\sigma \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$ be density operators, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Let*

$$\sigma - (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho) = \sum_{k=1}^{m} \lambda_k \Pi_k \tag{111}$$

be a spectral decomposition, and define

$$Y = \sum_{k=1}^{m} \operatorname{sign}(\lambda_k)\, \Pi_k, \tag{112}$$

where $\operatorname{sign}(\alpha) = 1$ *and* $\operatorname{sign}(-\alpha) = -1$ *for all* $\alpha > 0$ *and* $\operatorname{sign}(0) = 0$. *The channel* Φ *is optimal for the state transformation problem in* (106) *if and only if the operator*

$$H = \big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*\big)(Y) \tag{113}$$

satisfies

$$\operatorname{Tr}_{\mathcal{Y}}(HJ(\Phi)) \in \operatorname{Herm}(\mathcal{X}) \quad \textit{and} \quad H \geq \mathbb{1}_{\mathcal{Y}} \otimes \operatorname{Tr}_{\mathcal{Y}}\big(HJ(\Phi)\big). \tag{114}$$

Objective functions based on relative entropy

Finally, we consider a variant of the optimal state transformation problem based on the quantum relative entropy. For positive semidefinite operators $P, Q, \in \operatorname{Pos}(\mathcal{X})$, the quantum relative entropy of P with respect to Q is defined as

$$\mathrm{D}(P\|Q) = \begin{cases} \operatorname{Tr}(P\log(P)) - \operatorname{Tr}(P\log(Q)) & \text{if } \operatorname{im}(P) \subseteq \operatorname{im}(Q) \\ \infty & \text{otherwise.} \end{cases}$$

The specific variant of the problem to be considered is

$$\begin{aligned} &\text{minimize} && \mathrm{D}\big(\sigma \big\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\big) \\ &\text{subject to} && \Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y}). \end{aligned} \tag{115}$$

The relative entropy is jointly convex, which implies that it is convex in its second argument, and therefore the problem above is a convex optimization problem. Because the relative entropy can be approximated through the use of semidefinite programming [FF18, FSP18], it is possible to efficiently approximate the optimization problem (115) on a computer.

Theorem 12. *Let $\rho \in \operatorname{Pos}(\mathcal{X} \otimes \mathcal{Z})$ and $\sigma \in \operatorname{Pos}(\mathcal{Y} \otimes \mathcal{Z})$ be positive semidefinite operators, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and assume that $\operatorname{Tr}_{\mathcal{X}}(\rho)$ is a positive definite operator. A channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is optimal for the optimization problem* (115) *if and only if the following two conditions are met:*

1. $\operatorname{im}(\sigma) \subseteq \operatorname{im}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\big)$.

2. *The operator*

$$H = -\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*\big)\big(D\log\big(\Pi\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)\Pi\big)(\sigma)\big) \tag{116}$$

satisfies

$$\operatorname{Tr}_{\mathcal{Y}}(HJ(\Phi)) \in \operatorname{Herm}(\mathcal{X}) \quad \textit{and} \quad H \geq \mathbb{1}_{\mathcal{Y}} \otimes \operatorname{Tr}_{\mathcal{Y}}(HJ(\Phi)). \tag{117}$$

Here, Π denotes the projection onto the image of σ and $D\log(P)$ denotes the differential operator of the logarithm function at the operator P (as described in (166) at the end of Appendix A).

Proof. If the first condition does not hold for a given channel Φ, then the objective function in (115) takes an infinite value. However, by the assumption that $\operatorname{Tr}_{\mathcal{X}}(\rho)$ is positive definite, one has that the channel

$$\Omega(X) = \frac{\operatorname{Tr}(X)\mathbb{1}_{\mathcal{Y}}}{\dim(\mathcal{Y})} \tag{118}$$

yields a finite value for the same objective function, implying that Φ is not optimal. If Φ is optimal, the first condition must therefore hold. It remains to prove that if Φ satisfies the first condition, then Φ is optimal if and only if the second condition holds.

Let r be the rank of σ, let $\mathcal{W} = \mathbb{C}^r$, and let $A \in \mathrm{L}(\mathcal{W}, \mathcal{Y} \otimes \mathcal{Z})$ be an isometry that satisfies $AA^* = \Pi_{\mathrm{im}(\sigma)}$. Define a linear map $\Xi : \operatorname{Herm}(\mathcal{Y} \otimes \mathcal{X}) \rightarrow \operatorname{Herm}(\mathcal{W})$ as

$$\Xi(X) = A^*\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho\big)(X)A \tag{119}$$

for all $X \in \operatorname{Herm}(\mathcal{Y} \otimes \mathcal{X})$. For every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ it is the case that

$$\mathrm{D}\big(\sigma \big\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\big) = \mathrm{D}\big(A^*\sigma A \big\| \Xi(J(\Phi))\big). \tag{120}$$

With this observation in mind, define a function $f : \operatorname{Herm}(\mathcal{Y} \otimes \mathcal{X}) \rightarrow \mathbb{R} \cup \{\infty\}$ as

$$f(X) = \begin{cases} \mathrm{D}\big(A^*\sigma A \big\| \Xi(X)\big) & \text{if } \Xi(X) \in \operatorname{Pd}(\mathcal{W}) \\ \infty & \text{otherwise,} \end{cases} \tag{121}$$

so that a given channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is optimal for the problem (115) if and only if $J(\Phi)$ is optimal for the problem

$$\begin{aligned} &\text{minimize} && f(X) \\ &\text{subject to} && X \in J(\mathrm{C}(\mathcal{X}, \mathcal{Y})). \end{aligned} \tag{122}$$

The function f is differentiable at every operator $X \in \operatorname{Herm}(\mathcal{Y} \otimes \mathcal{X})$ in its domain, with its gradient being

$$\begin{aligned} \nabla f(X) &= -\Xi^*\big(D\log(\Xi(X))(A^*\sigma A)\big) \\ &= -\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*\big)\big(D\log\big(\Pi\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho\big)(X)\Pi\big)(\sigma)\big). \end{aligned} \tag{123}$$

As f is differentiable at every point in its domain, it follows that

$$\partial f(J(\Phi)) = \{\nabla f(J(\Phi))\} \tag{124}$$

for every $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ for which $\mathrm{im}(\sigma) \subseteq \mathrm{im}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\big)$. For a given channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ for which $\mathrm{im}(\sigma) \subseteq \mathrm{im}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\big)$, it therefore follows from Theorem 3 that Φ is optimal if and only if the operator

$$H = -\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho^*\big)\big(D \log\big(\Pi\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)\Pi\big)(\sigma)\big) \tag{125}$$

satisfies

$$\mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) \in \mathrm{Herm}(\mathcal{X}) \quad \text{and} \quad H \geq \mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)), \tag{126}$$

which is the second condition in the statement of the theorem. □

Acknowledgments

We thank Jamie Sikora for helpful comments and discussions and Mark Wilde for valuable suggestions. This research was supported by Canada's NSERC.

A Gradients and subdifferentials of functions on matrices

Results in the main body of this have required the computation of gradients and subdifferentials for various functions mapping Hermitian operators to the real numbers. In this appendix we provide details on these computations.

Definitions and basic results

It is appropriate to begin with some basic definitions. Throughout this discussion, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are arbitrary complex Euclidean spaces.

Suppose that

$$f : \mathrm{Herm}(\mathcal{X}) \rightarrow \mathrm{Herm}(\mathcal{Y}) \tag{127}$$

is a partial function, meaning that it may only be defined on some subset of inputs $X \in \mathrm{Herm}(\mathcal{X})$. The function f is *(Fréchet) differentiable* at $X \in \mathrm{Herm}(\mathcal{X})$ if there exists a linear map

$$\Phi : \mathrm{Herm}(\mathcal{X}) \rightarrow \mathrm{Herm}(\mathcal{Y}) \tag{128}$$

for which the equation

$$\lim_{Z \to 0} \frac{\|f(X+Z) - f(X) - \Phi(Z)\|}{\|Z\|} = 0 \tag{129}$$

is satisfied. If there does exist such a map, it must be unique, and we denote it by $Df(X)$. Whenever f is differentiable at X, it must be the case that

$$Df(X)(Z) = \frac{\mathrm{d}}{\mathrm{d}t} f(X + tZ)\Big|_{t=0} \tag{130}$$

for all choices of $Z \in \mathrm{Herm}(\mathcal{X})$. In the special case that $\mathcal{Y} = \mathbb{C}$, which is equivalent to f taking the form

$$f : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R}, \tag{131}$$

one has that

$$Df(X)(Z) = \langle \nabla f(X), Z \rangle \tag{132}$$

for all $Z \in \mathrm{Herm}(\mathcal{X})$, assuming f is differentiable at X.

The *chain rule* for differentiation states that if

$$f : \mathrm{Herm}(\mathcal{X}) \to \mathrm{Herm}(\mathcal{Y}) \quad \text{and} \quad g : \mathrm{Herm}(\mathcal{Y}) \to \mathrm{Herm}(\mathcal{Z}), \tag{133}$$

f is differentiable at X, and g is differentiable at $Y = f(X)$, then

$$D(g \circ f)(X)(Z) = Dg(f(X))(Df(X)(Z)) \tag{134}$$

for all $Z \in \mathrm{Herm}(\mathcal{X})$.

Affine linear functions

A function

$$f : \mathrm{Herm}(\mathcal{X}) \to \mathrm{Herm}(\mathcal{Y}) \tag{135}$$

is *affine linear* if there exists a linear map

$$\Phi : \mathrm{Herm}(\mathcal{X}) \to \mathrm{Herm}(\mathcal{Y}) \tag{136}$$

and an operator $Y \in \mathrm{Herm}(\mathcal{Y})$ such that

$$f(X) = \Phi(X) + Y \tag{137}$$

for all $X \in \mathrm{Herm}(\mathcal{X})$. Every such function is differentiable at every $X \in \mathrm{Herm}(\mathcal{X})$, with its derivative given by

$$Df(X) = \Phi. \tag{138}$$

For an arbitrary function

$$g : \mathrm{Herm}(\mathcal{Y}) \to \mathrm{Herm}(\mathcal{Z}), \tag{139}$$

one therefore finds that

$$D(g \circ f)(X)(Z) = Dg(f(X))(\Phi(Z)), \tag{140}$$

provided that g is differentiable at $f(X)$, and if g takes the form

$$g : \mathrm{Herm}(\mathcal{Y}) \to \mathbb{R}, \tag{141}$$

then it is the case that

$$\nabla(g \circ f)(X) = \Phi^*(\nabla g(f(X))). \tag{142}$$

Real-valued functions extended to Hermitian operators

If $f : \mathbb{R} \to \mathbb{R}$ is a function, then it may be extended to a function of the form

$$g : \mathrm{Herm}(\mathcal{X}) \to \mathrm{Herm}(\mathcal{X}) \tag{143}$$

in a standard way: for any choice of $X \in \mathrm{Herm}(\mathcal{X})$, one considers the spectral decomposition

$$X = \sum_{k=1}^{m} \lambda_k \Pi_k \tag{144}$$

of X, then defines

$$g(X) = \sum_{k=1}^{m} f(\lambda_k)\Pi_k. \tag{145}$$

(It is typical that this extended function is given the same name as the original function on the real numbers, but for the sake of clarity we have introduced a distinct name for the extended function.) Naturally, if f is defined only on a subset of $\mathbb{R}$, then g is defined for all X whose eigenvalues are contained in the domain of f.

The function g is differentiable at every Hermitian operator whose eigenvalues correspond to differentiable points of the function f. The derivative of g can be described explicitly by first defining a function

$$h(\alpha, \beta) = \begin{cases} \frac{f(\alpha)-f(\beta)}{\alpha-\beta} & \text{if } \alpha \neq \beta \\ f'(\alpha) & \text{if } \alpha = \beta \end{cases} \tag{146}$$

for every pair of points (α, β) for which f is differentiable at both α and β. (The function h is sometimes called the *first divided difference* of f, although this terminology is sometimes limited to the case that $\alpha \neq \beta$.) In terms of this function, the derivative of g at an operator X having a spectral decomposition (144) is

$$Dg(X)(Z) = \sum_{j,k=1}^{m} h(\lambda_j, \lambda_k)\Pi_j Z \Pi_k \tag{147}$$

for every $Z \in \mathrm{Herm}(\mathcal{X})$ (see for example Theorem V.3.3 in [Bha97] and Theorem 3.25 in [HP14]).

Gradients of functions involving the fidelity

Let $\mathcal{Y}$ be a complex Euclidean space, and consider the function $f : \mathrm{Herm}(\mathcal{Y}) \to \mathbb{R} \cup \{\infty\}$ defined as

$$f(Y) = \begin{cases} -\operatorname{Tr}\sqrt{Y} & \text{if } Y \in \mathrm{Pos}(\mathcal{Y}) \\ \infty & \text{otherwise.} \end{cases} \tag{148}$$

This function is differentiable at every positive definite operator $Y \in \mathrm{Pd}(\mathcal{Y})$, with its gradient being

$$\nabla f(Y) = -\frac{1}{2}Y^{-\frac{1}{2}}. \tag{149}$$

One way to verify this expression is to first consider the function $g(Y) = \sqrt{Y}$, defined for every positive semidefinite operator $Y \in \mathrm{Pos}(\mathcal{Y})$, and to use the formula (147) to conclude that

$$Dg(Y)(Z) = \sum_{j,k=1}^{m} \frac{\Pi_j Z \Pi_k}{\sqrt{\lambda_j} + \sqrt{\lambda_k}}, \tag{150}$$

provided that Y is positive definite and has spectral decomposition

$$Y = \sum_{k=1}^{m} \lambda_k \Pi_k \tag{151}$$

(see also Example 3.26 in [HP14]). The equation (149) follows from the chain rule.

Lemma 13. *Let $\mathcal{X}$ be a complex Euclidean space, let $P \in \mathrm{Pd}(\mathcal{X})$ be a positive definite operator, and define a function $g : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R} \cup \{\infty\}$ as*

$$g(X) = \begin{cases} -\,\mathrm{F}(P, X) & \text{if } X \in \mathrm{Pos}(\mathcal{X}) \\ \infty & \text{otherwise.} \end{cases} \tag{152}$$

For every $X \in \mathrm{Pd}(\mathcal{X})$, the function g is differentiable at X, and

$$\nabla g(X) = -\frac{1}{2}\big(P \,\#\, X^{-1}\big), \tag{153}$$

where $P \,\#\, X^{-1}$ denotes the operator geometric mean of P and X^{-1}, as discussed in Remark 9. For every operator $X \in \mathrm{Pos}(\mathcal{X})$ that is not positive definite, it is the case that $\partial g(X) = \varnothing$.

Proof. Define a linear map $\Lambda : \mathrm{Herm}(\mathcal{X}) \to \mathrm{Herm}(\mathcal{Y})$ as

$$\Lambda(X) = \sqrt{P} X \sqrt{P} \tag{154}$$

for all $X \in \mathrm{Herm}(\mathcal{X})$. It is the case that $g = f \circ \Lambda$, where f is as defined in (148) above. By the chain rule for differentiation, one has

$$\nabla g(X) = -\frac{1}{2}\Lambda^*(\nabla f(\Lambda(X))) = -\frac{1}{2}\sqrt{P}\Big(\sqrt{P}X\sqrt{P}\Big)^{-\frac{1}{2}}\sqrt{P}, \tag{155}$$

provided that

$$\sqrt{P}X\sqrt{P} \in \mathrm{Pd}(\mathcal{Y}), \tag{156}$$

which is equivalent to $X \in \mathrm{Pd}(\mathcal{X})$. In the case when X is positive definite, one therefore has that

$$\nabla g(X) = -\frac{1}{2}\sqrt{P}\Big(\sqrt{P^{-1}}X^{-1}\sqrt{P^{-1}}\Big)^{\frac{1}{2}}\sqrt{P} = -\frac{1}{2}\big(P \,\#\, X^{-1}\big), \tag{157}$$

as desired.

Now suppose that $X \in \mathrm{Pos}(\mathcal{X})$ is not positive definite, and let Δ be the projection onto the kernel of $\sqrt{P}X\sqrt{P}$, which is nonzero by the assumption that X is not positive definite. Consider the operator

$$Y = X + \lambda P^{-\frac{1}{2}} \Delta P^{-\frac{1}{2}} \tag{158}$$

for an arbitrary choice of $\lambda > 0$. It is the case that

$$g(Y) - g(X) = \mathrm{Tr}\sqrt{\sqrt{P}X\sqrt{P}} - \mathrm{Tr}\sqrt{\sqrt{P}X\sqrt{P} + \lambda\Delta} = -\sqrt{\lambda}\,\mathrm{Tr}(\Delta). \tag{159}$$

Thus, if there were to exist an element $Z \in \partial g(X)$, one would have

$$g(Y) - g(X) \geq \langle Z, Y - X \rangle \tag{160}$$

for all $Y \in \mathrm{dom}(f)$, including the operator (158) for every $\lambda > 0$. It would then follow that

$$\lambda\left\langle Z, P^{-\frac{1}{2}} \Delta P^{-\frac{1}{2}} \right\rangle = \langle Z, Y - X \rangle \leq g(Y) - g(X) = -\sqrt{\lambda}\,\mathrm{Tr}(\Delta), \tag{161}$$

or equivalently

$$\frac{\left\langle Z, P^{-\frac{1}{2}} \Delta P^{-\frac{1}{2}} \right\rangle}{\mathrm{Tr}(\Delta)} \leq -\frac{1}{\sqrt{\lambda}}, \tag{162}$$

for every $\lambda > 0$, which is impossible given that the left-hand side is a finite value independent of λ and the right-hand side approaches $-\infty$ as λ approaches 0. □

Gradients of functions involving the quantum relative entropy

Let $\mathcal{Y}$ be a complex Euclidean space, let $P \in \mathrm{Pd}(\mathcal{Y})$ be a positive definite operator, and consider the function $f : \mathrm{Herm}(\mathcal{Y}) \to \mathbb{R} \cup \{\infty\}$ defined as

$$f(Y) = \begin{cases} \mathrm{D}(P\|Y) & \text{if } Y \in \mathrm{Pd}(\mathcal{Y}) \\ \infty & \text{otherwise.} \end{cases} \tag{163}$$

This function is differentiable at every positive definite operator $Y \in \mathrm{Pd}(\mathcal{Y})$, with its gradient being

$$\nabla f(Y) = -D\log(Y)(P), \tag{164}$$

where $D\log(Y)$ is the derivative of the logarithm function at Y. If

$$Y = \sum_{k=1}^{m} \lambda_k \Pi_k \tag{165}$$

is the spectral decomposition of Y, then by means of the expression (147) this function can be described explicitly as

$$D\log(Y)(Z) = \sum_{k=1}^{m} \frac{1}{\lambda_k} \Pi_k Z \Pi_k + \sum_{j \neq k} \frac{\log(\lambda_j) - \log(\lambda_k)}{\lambda_j - \lambda_k} \Pi_j Z \Pi_k. \tag{166}$$

B Proof of multiplicativity of fidelity of recovery

In this appendix, we use the result of Theorem 7 to prove that the generalized fidelity of recovery is multiplicative. This fact has already been shown in [BT16], where it was shown by examining the fidelity of recovery as a semidefinite program.

Let $\rho_0 \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{Z}_0)$, $\rho_1 \in \mathrm{D}(\mathcal{X}_1 \otimes \mathcal{Z}_1)$, $\sigma_0 \in \mathrm{D}(\mathcal{Y}_0 \otimes \mathcal{Z}_0)$, and $\sigma_1 \in \mathrm{D}(\mathcal{Y}_1 \otimes \mathcal{Z}_1)$ be density operators for some complex Euclidean spaces $\mathcal{X}_0, \mathcal{X}_1, \mathcal{Y}_0, \mathcal{Y}_1, \mathcal{Z}_0, \mathcal{Z}_1$, and consider the following pair of optimization problems:

$$\begin{array}{ll} \text{maximize} & \mathrm{F}\big(\sigma_0, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_0)}\big)(\rho_0)\big) \\ \text{subject to} & \Phi \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0) \end{array} \qquad \begin{array}{ll} \text{maximize} & \mathrm{F}\big(\sigma_1, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_1)}\big)(\rho_1)\big) \\ \text{subject to} & \Phi \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1). \end{array} \tag{167}$$

Define density operators $\rho \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1 \otimes \mathcal{Z}_0 \otimes \mathcal{Z}_1)$ and $\sigma \in \mathrm{D}(\mathcal{Y}_0 \otimes \mathcal{Y}_1 \otimes \mathcal{Z}_0 \otimes \mathcal{Z}_1)$ as

$$\rho = W(\rho_0 \otimes \rho_1)W^* \qquad \text{and} \qquad \sigma = V(\sigma_0 \otimes \sigma_1)V^* \tag{168}$$

where

$$\begin{aligned} & W \in \mathrm{L}(\mathcal{X}_0 \otimes \mathcal{Z}_0 \otimes \mathcal{X}_1 \otimes \mathcal{Z}_1, \mathcal{X}_0 \otimes \mathcal{X}_1 \otimes \mathcal{Z}_0 \otimes \mathcal{Z}_1) \\ \text{and} \quad & V \in \mathrm{L}(\mathcal{Y}_0 \otimes \mathcal{Z}_0 \otimes \mathcal{Y}_1 \otimes \mathcal{Z}_1, \mathcal{Y}_0 \otimes \mathcal{Y}_1 \otimes \mathcal{Z}_0 \otimes \mathcal{Z}_1) \end{aligned}$$

are the isometries defined as

$$\begin{aligned} & W(x_0 \otimes z_0 \otimes x_1 \otimes z_1) = x_0 \otimes x_1 \otimes z_0 \otimes z_1 \\ \text{and} \quad & V(y_0 \otimes z_0 \otimes y_1 \otimes z_1) = y_0 \otimes y_1 \otimes z_0 \otimes z_1 \end{aligned}$$

for all choices of vectors $x_0 \in \mathcal{X}_0$, $x_1 \in \mathcal{X}_1$, $y_0 \in \mathcal{Y}_0$, $y_1 \in \mathcal{Y}_1$, $z_0 \in \mathcal{Z}_0$, and $z_1 \in \mathcal{Z}_1$. Consider now the following optimization problem:

$$\begin{array}{ll} \text{maximize} & \mathrm{F}\big(\sigma, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_0 \otimes \mathcal{Z}_1)}\big)(\rho)\big) \\ \text{subject to} & \Phi \in \mathrm{C}(\mathcal{X}_0 \otimes \mathcal{X}_1, \mathcal{Y}_0 \otimes \mathcal{Y}_1). \end{array} \tag{169}$$

The fact that the generalized fidelity of recovery is multiplicative can be stated as follows. Let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels and suppose that this pair of channels is optimal for the pair of optimization problems in (167). Then the channel $\Phi_0 \otimes \Phi_1 \in \mathrm{C}(\mathcal{X}_0 \otimes \mathcal{X}_1, \mathcal{Y}_0 \otimes \mathcal{Y}_1)$ is optimal for the optimization problem in (169). To prove this fact, we may define operators $H_0 \in \mathrm{Herm}(\mathcal{Y}_0 \otimes \mathcal{X}_0)$ and $H_1 \in \mathrm{Herm}(\mathcal{Y}_1 \otimes \mathcal{X}_1)$ as

$$\begin{aligned} & H_0 = -\frac{1}{2}\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y}_0)} \otimes \Psi^*_{\rho_0}\big)\Big(\sqrt{\sigma_0}\big(\sqrt{\sigma_0}\,(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_0)})(\rho_0)\sqrt{\sigma_0}\big)^{-\frac{1}{2}}\sqrt{\sigma_0}\Big) \\ \text{and} \quad & H_1 = -\frac{1}{2}\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y}_1)} \otimes \Psi^*_{\rho_1}\big)\Big(\sqrt{\sigma_1}\big(\sqrt{\sigma_1}\,(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_1)})(\rho_1)\sqrt{\sigma_1}\big)^{-\frac{1}{2}}\sqrt{\sigma_1}\Big). \end{aligned}$$

From the assumption that both Φ_0 and Φ_1 are optimal, it follows from Theorem 7 that the following conditions hold:

1. $\mathrm{im}(\sigma_0) \subseteq \mathrm{im}\big((\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_0)})(\rho_0)\big)$ and $\mathrm{im}(\sigma_1) \subseteq \mathrm{im}\big((\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_1)})(\rho_1)\big)$
2. $\mathrm{Tr}_{\mathcal{Y}_0}(H_0 J(\Phi_0)) \in \mathrm{Herm}(\mathcal{X}_0)$ and $\mathrm{Tr}_{\mathcal{Y}_1}(H_1 J(\Phi_1)) \in \mathrm{Herm}(\mathcal{X}_1)$
3. $H_0 \geq \mathbb{1}_{\mathcal{Y}_0} \otimes \mathrm{Tr}_{\mathcal{Y}_0}(H_0 J(\Phi_0))$ and $H_1 \geq \mathbb{1}_{\mathcal{Y}_1} \otimes \mathrm{Tr}_{\mathcal{Y}_1}(H_1 J(\Phi_1))$.

It is evident that $\mathrm{im}(\sigma_0 \otimes \sigma_1) \subseteq \mathrm{im}\big((\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_0)} \otimes \Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_1)})(\rho_0 \otimes \rho_1)\big)$ by taking tensor products, and it follows that

$$\mathrm{im}(\sigma) \subseteq \mathrm{im}\big((\Phi_0 \otimes \Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_0 \otimes \mathcal{Z}_1)})(\rho)\big) \tag{170}$$

by rearranging the spaces. Define the operator

$$H = -\frac{1}{2}(\mathbb{1}_{\mathrm{L}(\mathcal{Y}_0 \otimes \mathcal{Y}_1)} \otimes \Psi_\rho^*)\Big(\sqrt{\sigma}\big(\sqrt{\sigma}\,(\Phi_0 \otimes \Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_0 \otimes \mathcal{Z}_1)})(\rho)\sqrt{\sigma}\big)^{-\frac{1}{2}}\sqrt{\sigma}\Big), \tag{171}$$

and define the isometry $U \in \mathrm{L}(\mathcal{Y}_0 \otimes \mathcal{X}_0 \otimes \mathcal{Y}_1 \otimes \mathcal{X}_1, \mathcal{Y}_0 \otimes \mathcal{Y}_1 \otimes \mathcal{X}_0 \otimes \mathcal{X}_1)$ as the operator satisfying

$$U(y_0 \otimes x_0 \otimes y_1 \otimes x_1) = y_0 \otimes y_1 \otimes x_0 \otimes x_1 \tag{172}$$

for every choice of vectors $x_0 \in \mathcal{X}_0$, $x_1 \in \mathcal{X}_1$, $y_0 \in \mathcal{Y}_0$, and $y_1 \in \mathcal{Y}_1$. With these choices of operators, it is evident that $H = U(H_0 \otimes H_1)U^*$ and it follows that

$$\mathrm{Tr}_{\mathcal{Y}_0 \otimes \mathcal{Y}_1}(HJ(\Phi_0 \otimes \Phi_1)) = \mathrm{Tr}_{\mathcal{Y}_0}(H_0 J(\Phi_0)) \otimes \mathrm{Tr}_{\mathcal{Y}_1}(H_1 J(\Phi_1)) \in \mathrm{Herm}(\mathcal{X}_0 \otimes \mathcal{X}_1), \tag{173}$$

as the tensor product of Hermitian operators is Hermitian. Moreover it holds that

$$\begin{aligned} H &= U(H_0 \otimes H_1)U^* \\ &\geq U\Big(\big(\mathbb{1}_{\mathcal{Y}_0} \otimes \mathrm{Tr}_{\mathcal{Y}_0}(H_0 J(\Phi_0))\big) \otimes \big(\mathbb{1}_{\mathcal{Y}_1} \otimes \mathrm{Tr}_{\mathcal{Y}_1}(H_1 J(\Phi_1))\big)\Big)U^* \\ &= \mathbb{1}_{\mathcal{Y}_0 \otimes \mathcal{Y}_1} \otimes \mathrm{Tr}_{\mathcal{Y}_0 \otimes \mathcal{Y}_1}(HJ(\Phi_0 \otimes \Phi_1)), \end{aligned}$$

which follows from the fact that $P_0 \otimes P_1 \geq Q_0 \otimes Q_1$ holds for every choice of positive semidefinite operators P_0, P_1, Q_0, Q_1 satisfying $P_0 \geq Q_0$ and $P_1 \geq Q_1$. As the conditions of Theorem 7 are satisfied for the optimization problem in (169), one has that $\Phi_0 \otimes \Phi_1$ is optimal for this problem.

References

[AFG+12] Scott Aaronson, Edward Farhi, David Gosset, Avinatan Hassidim, Jonathan Kelner, and Andrew Lutomirski. Quantum money. *Communications of the ACM*, 55(8):84–92, 2012.

[BC09] Stephen Barnett and Sarah Croke. Quantum state discrimination. *Advances in Optics and Photonics*, 1(2):238–278, 2009.

[BCvD05] Dave Bacon, Andrew Childs, and Wim van Dam. From optimal measurement to efficient quantum algorithms for the hidden subgroup problem over semidirect product groups. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 469–478, 2005.

[BDE+98] Dagmar Bruß, David DiVincenzo, Artur Ekert, Christopher Fuchs, Chiara Macchiavello, and John Smolin. Optimal universal and state-dependent quantum cloning. *Physical Review A*, 57:2368–2378, Apr 1998.

[BFT17] Mario Berta, Omar Fawzi, and Marco Tomamichel. On variational expressions for quantum relative entropies. *Letters in Mathematical Physics*, 107(12):2239–2265, 2017.

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, 1997.

[Bha07] Rajendra Bhatia. *Positive Definite Matrices*, volume 24 of *Princeton Series in Applied Mathematics*. Princeton University Press, 2007.

[BHOS15] Fernando Brandão, Aram Harrow, Jonathan Oppenheim, and Sergii Strelchuk. Quantum conditional mutual information, reconstructed states, and state redistribution. *Physical Review Letters*, 115(5):050501, 2015.

[BL06] Jonathan Borwein and Adrian Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. Springer, 2006.

[BT16] Mario Berta and Marco Tomamichel. The fidelity of recovery is multiplicative. *IEEE Transactions on Information Theory*, 62(4):1758–1763, 2016.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[CHM+16] Tom Cooney, Christoph Hirche, Ciara Morgan, Jonathan Olson, Kaushik Seshadreesan, John Watrous, and Mark Wilde. Operational meaning of quantum measures of recovery. *Physical Review A*, 94(2):022310, 2016.

[EMV03] Yonina Eldar, Alexandre Megretski, and George Verghese. Designing optimal quantum detectors via semidefinite programming. *IEEE Transactions on Information Theory*, 49(4):1007–1012, 2003.

[FF18] Hamza Fawzi and Omar Fawzi. Efficient optimization of the quantum relative entropy. *Journal of Physics A: Mathematical and Theoretical*, 51(15):154003, 2018.

[FR15] Omar Fawzi and Renato Renner. Quantum conditional mutual information and approximate Markov chains. *Communications in Mathematical Physics*, 340(2):575–611, 2015.

[FSP18] Hamza Fawzi, James Saunderson, and Pablo Parrilo. Semidefinite approximations of the matrix logarithm. *Foundations of Computational Mathematics*, 2018.

[Hol73a] Alexander Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3:337–394, 1973.

[Hol73b] Alexander Holevo. Statistical problems in quantum physics. In *Proceedings of the Second Japan–USSR Symposium on Probability Theory*, volume 330 of *Lecture Notes in Mathematics*, pages 104–119. Springer, 1973.

[HP14] Fumio Hiai and Dénes Petz. *Introduction to Matrix Analysis and Applications*. Universitext. Springer International Publishing, 2014.

[Ip03] Lawrence Ip. Shor's algorithm is optimal. Unpublished manuscript, available from http://www.lawrenceip.com/papers/hspsdp.pdf, 2003.

[JUW09] Rahul Jain, Sarvagya Upadhyay, and John Watrous. Two-message quantum interactive proofs are in PSPACE. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 534–543, 2009.

[JvF02] Miroslav Ježek, Jaroslav Řeháček, and Jaromír Fiurášek. Finding optimal strategies for minimum-error quantum-state discrimination. *Physical Review A*, 65(6):060301, 2002.

[Lan13] Kenneth Lange. *Optimization*. Springer Texts in Statistics. Springer New York, 2013.

[MN13] Boris Mordukhovich and Nguyen Mau Nam. *An Easy Path to Convex Analysis and Applications*. Morgan & Claypool Publishers, 2013.

[MN17] Boris Mordukhovich and Nguyen Mau Nam. Geometric approach to convex subdifferential calculus. *Optimization*, 66(6):839–873, 2017.

[NC00] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[Roc70] R. Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[SIGA05] Valerio Scarani, Sofyan Iblisdir, Nicolas Gisin, and Antonio Acín. Quantum cloning. *Reviews of Modern Physics*, 77(4):1225–1256, 2005.

[Sut18] David Sutter. *Approximate Quantum Markov Chains*, volume 28 of *SpringerBriefs in Mathematical Physics*. Springer, 2018.

[SW15] Kaushik Seshadreesan and Mark Wilde. Fidelity of recovery, squashed entanglement, and measurement recoverability. *Physical Review A*, 92(4):042321, 2015.

[Wat18] John Watrous. *Theory of Quantum Information*. Cambridge University Press, 2018.

[Wil17] Mark Wilde. *Quantum Information Theory*. Cambridge University Press, second edition, 2017.

[YKL70] Horace Yuen, Robert Kennedy, and Melvin Lax. On optimal quantum receivers for digital signal detection. *Proceedings of the IEEE*, 58(10):1770–1773, 1970.

[YKL75] Horace Yuen, Robert Kennedy, and Melvin Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Transactions on Information Theory*, 21(2):125–134, 1975.