Geometric conditions for saturating the data processing inequality

Samuel S. Cree and Jonathan Sorce

The data processing inequality (DPI) is a scalar inequality satisfied by distinguishability measures on density matrices. For some distinguishability measures, saturation of the *scalar* DPI implies an *operator* equation relating the arguments of the measure. These results are typically derived using functional analytic techniques. In a complementary approach, we use geometric techniques to derive a formula that gives an operator equation from DPI saturation for *any* distinguishability measure; moreover, for a broad class of distinguishability measures, the derived operator equation is sufficient to imply saturation as well. Our operator equation coincides with known results for the sandwiched Rényi relative entropies, and gives new results for α-z Rényi relative entropies and a family of of quantum f-divergences, which we compute explicitly. We also comment on connections to Petz recovery, and on how our framework could be used to derive operator equations from approximate DPI saturation.

1 Introduction

The distinguishability of two pure quantum states $|\psi\rangle$ and $|\phi\rangle$ is completely characterized by the inner product $\langle\psi|\phi\rangle$. For mixed states ρ and σ, however, characterizing the "distinguishability" of the states is more complicated. As a result, there is a zoo of distinguishability measures (see e.g. [1–14]), each with their own advantages and disadvantages. To qualify as a distinguishability measure, a function of two density matrices really only needs to satisfy one physical principle: it should not be possible for two states to become *more* distinguishable by applying a quantum channel. This condition is known as the *data processing inequality*, or *DPI*.

Formally, a distinguishability measure $\mathcal{B}$ for states on a Hilbert space $\mathcal{H}$ is a function[1]

$$\mathcal{B} : \mathrm{Pos}(\mathcal{H}) \times \mathrm{Pos}(\mathcal{H}) \to \mathbb{R} \tag{1.1}$$

that satisfies the data processing inequality: for any quantum channel[2] Λ, and any two (generally non-normalized) states $\rho, \sigma \in \mathrm{Pos}(\mathcal{H})$, we have

$$\mathcal{B}(\rho, \sigma) \geq \mathcal{B}(\Lambda(\rho), \Lambda(\sigma)). \tag{1.2}$$

We will also assume, as part of our definition, that $\mathcal{B}$ is differentiable; this is the case for every distinguishability measure we consider in this paper.

We say that Λ is *recoverable* on the states ρ and σ if there exists a channel $\mathcal{R}$ satisfying

$$[\mathcal{R} \circ \Lambda](\rho) = \rho \tag{1.3}$$

and

$$[\mathcal{R} \circ \Lambda](\sigma) = \sigma. \tag{1.4}$$

Samuel S. Cree: scree@stanford.edu
Jonathan Sorce: jsorce@stanford.edu

[1]Here $\mathrm{Pos}(\mathcal{H})$ denotes the set of positive definite operators, i.e., the set of operators P such that $\langle\psi|P|\psi\rangle > 0$ holds for all $|\psi\rangle \in \mathcal{H}$.

[2]Recall that a quantum channel is a completely positive, trace-preserving, linear map on the space of operators.

When such a channel exists, applying the DPI with respect to Λ and then with respect to $\mathcal{R}$ shows that we must have equality in (1.2) for any distinguishability measure $\mathcal{B}$. For certain choices of distinguishability measure, the converse is also true; if the distinguishability of ρ and σ does not change under the application of Λ, then Λ is recoverable on those states. This fact, that saturation of DPI is equivalent to recoverability for certain distinguishability measures, is known as *Petz recovery* in honor of Petz's proof that the relative entropy is one distinguishability measure with this property [15, 16].

Part of the difficulty in proving Petz recovery for a given distinguishability measure is that saturation of the DPI is a scalar equation, while the result to be proved is an equality of operators – the initial state ρ must equal the final state $[\mathcal{R} \circ \Lambda](\rho)$. For this reason, several authors have worked to derive operator equations that are implied by (or equivalent to) saturation of the data-processing inequality [5, 9–11, 14, 17–25]. The techniques for deriving these equations generally fall under the mathematical umbrella of functional analysis; the purpose of this paper is to introduce a complementary, geometric toolkit for deriving operator equations from DPI saturation. As we will see, this approach immediately reproduces known results for the relative entropy and fidelity, and more generally for the full class of sandwiched Rényi relative entropies. It also allows us to derive new conditions for the α-z Rényi relative entropies, and for a general family of f-divergences.

The technique is simple. Because a distinguishability measure $\mathcal{B}$ has as its domain two operator manifolds, its *gradient* with respect to either argument is a tangent vector on the corresponding operator manifold. Tangent vectors on manifolds of operators are themselves operators. For density matrices ρ, σ that saturate DPI for a particular quantum channel Λ, the function

$$f_\Lambda(\rho, \sigma) = \mathcal{B}(\rho, \sigma) - \mathcal{B}(\Lambda(\rho), \Lambda(\sigma)) \tag{1.5}$$

is at an extreme value – its minimum, zero. As such, the gradient of f with respect to either of its arguments must vanish. Since the gradient is an operator, this implies an operator equation. The technical matter of this paper is primarily in (i) explicitly computing this vanishing-gradient equation for specific examples, and (ii) showing that for a broad class of distinguishability measures, this vanishing-gradient equation is also sufficient to imply DPI saturation.

The plan of the paper is as follows.

In section 2, we give a refresher on how to take derivatives of functions defined on manifolds of operators. We give several explicit formulas, the most nontrivial of which is derived in appendix A, that are then used in section 3. Readers who are already familiar with derivatives on matrix manifolds may safely skip section 2 and refer to it only as needed while reading section 3.

In section 3, we detail the "vanishing gradient" argument alluded to above. We give the vanishing-gradient equation a simple form in theorem 3.1, and provide a condition under which the vanishing-gradient equation is *equivalent* to DPI saturation, rather than merely implied by it. We then use theorem 3.1 to derive necessary and sufficient conditions for various distinguishability measures to saturate the data processing inequality. In particular, we replicate a previously known result for the sandwiched Rényi relative entropies [25], though our proof technique is different and, for those less familiar with functional analytic techniques, hopefully more intuitive. We also comment on the case of the α-z Rényi relative entropies, where the vanishing-gradient equation is not superficially identical to either of the DPI saturation conditions derived in [23, 24].

In section 4, we derive operator equations implied by DPI saturation in the case that one or both of the density matrices are positive semidefinite but not strictly positive. The basic idea is that while a general distinguishability measure will not be differentiable in a neighborhood of a density matrix with a vanishing eigenvalue, nor will it necessarily even satisfy DPI in a neighborhood of that density matrix, directional derivatives along the boundary of the space of positive operators are required to vanish.

In section 5, we comment on potential applications to Petz recovery and other directions for future work.

2 User's guide to derivatives on matrix manifolds

Let $\mathcal{H}$ be an n-dimensional Hilbert space. The set of Hermitian operators on $\mathcal{H}$, denoted $\mathrm{Herm}(\mathcal{H})$, is an n^2-dimensional real manifold with coordinates given by the real and complex parts of the independent matrix entries under any choice of basis for $\mathcal{H}$. The space of positive operators, $\mathrm{Pos}(\mathcal{H})$, is an n^2-dimensional real submanifold of $\mathrm{Herm}(\mathcal{H})$.[3] Our goal in this section will be to understand the derivatives of functions $f : \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}$. We develop this machinery so that for any distinguishability measure $\mathcal{B} : \mathrm{Pos}(\mathcal{H}) \times \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}$, and any positive operator σ, we will be able to compute the derivative of the restricted map $\mathcal{B}|_\sigma : \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}$ defined by $\mathcal{B}|_\sigma(\rho) = \mathcal{B}(\rho, \sigma)$. Our pedagogy in this section roughly follows section 5 of [26].

Let $f : M \to N$ be a smooth map between manifolds M and N. The derivative of f at a point $p \in M$, denoted $\mathrm{d}f|_p$, is defined as a linear map from the tangent space T_pM to the tangent space $T_{f(p)}N$. Its action on the tangent space is such that for any curve γ in M passing through p, the derivative $\mathrm{d}f|_p$ maps the tangent vector of γ at the point $p \in M$ to the tangent vector of $f(\gamma)$ at the point $f(p) \in N$.

Now, we restrict to the case where f is a map from $\mathrm{Pos}(\mathcal{H})$ to the real numbers. The tangent space to $\mathrm{Pos}(\mathcal{H})$ at any point $\rho \in \mathrm{Pos}(\mathcal{H})$ is isomorphic to the set of Hermitian operators $\mathrm{Herm}(\mathcal{H})$.[4] The tangent space to $\mathbb{R}$ at any point $f(\rho) \in \mathbb{R}$ is isomorphic to the real numbers $\mathbb{R}$. So $\mathrm{d}f|_\rho : T_\rho \mathrm{Pos}(\mathcal{H}) \to T_{f(\rho)}\mathbb{R}$ can be thought of as a linear map $\mathrm{d}f|_\rho : \mathrm{Herm}(\mathcal{H}) \to \mathbb{R}$. But using the Hilbert-Schmidt inner product, any linear map from $\mathrm{Herm}(\mathcal{H})$ to $\mathbb{R}$ can be written as a Hermitian operator! More specifically, there will always be a unique Hermitian operator $\nabla f|_\rho \in \mathrm{Herm}(\mathcal{H})$ satisfying[5]

$$\langle \nabla f|_\rho, M \rangle_{\mathrm{HS}} = \mathrm{d}f|_\rho(M) \tag{2.1}$$

for all $M \in \mathrm{Herm}(\mathcal{H})$. Here the Hilbert-Schmidt inner product on Hermitian matrices A, B is defined by

$$\langle A, B \rangle_{\mathrm{HS}} = \mathrm{Tr}(AB). \tag{2.2}$$

As explained in the introduction and detailed in section 3, saturation of the data processing inequality for a distinguishability measure $\mathcal{B}$ implies certain operator equations for derivatives of related functions $f : \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}$. To put these equations in a useful analytic form, we will need to be able to compute the Hermitian operator $\nabla f|_\rho$ appearing in equation (2.1). This is done by computing $\mathrm{d}f|_\rho$, then dualizing with respect to the Hilbert-Schmidt inner product. To compute $\mathrm{d}f$, it is helpful to write f as a composition of simple functions $f = f_1 \circ \cdots \circ f_n$, then to compute the derivative of f using the chain rule.

Every distinguishability measure we study in section 3 can be decomposed into pieces that are either *linear* or *locally analytic*. Denoting by $\mathcal{L}(\mathcal{H})$ the full space of linear operators on $\mathcal{H}$, any linear map f_j between vector subspaces[6] of $\mathcal{L}(\mathcal{H})$ has derivative given by

$$\mathrm{d}f_j|_A(M) = f_j(M), \tag{2.3}$$

which is straightforward to derive from the formula

$$\mathrm{d}f_j|_A(M) = \lim_{\epsilon \to 0} \frac{f_j(A + \epsilon M) - f(A)}{\epsilon}. \tag{2.4}$$

Note that the derivative is independent of the point $A \in \mathcal{L}(\mathcal{H})$ where the derivative is taken; the fact that the derivative of a real linear function $y(x) = ax + b$ is independent of x is a special case of this more general principle.

[3]Because the eigenvalues of a Hermitian matrix are continuous functions of the matrix entries, the eigenvalues of a strictly positive matrix remain strictly positive under a small Hermitian perturbation. So $\mathrm{Pos}(\mathcal{H})$ is an open subset of $\mathrm{Herm}(\mathcal{H})$, and hence a submanifold.

[4]A tangent vector to ρ in $\mathrm{Pos}(\mathcal{H})$ is a matrix M such that $\rho + \epsilon M$ is still in $\mathrm{Pos}(\mathcal{H})$ for sufficiently small ϵ. This is true if and only if the eigenvalues of M are real, i.e., if and only if M is Hermitian.

[5]For the more geometrically inclined: the derivative of f at ρ is in the dual space $T^*_\rho \mathrm{Pos}(\mathcal{H})$. The gradient ∇f is obtained by "raising the index" by mapping $T^*_\rho \mathrm{Pos}(\mathcal{H}) \to T_\rho \mathrm{Pos}(\mathcal{H})$ isomorphically using the Hilbert-Schmidt inner product on $T_\rho \mathrm{Pos}(\mathcal{H})$.

[6]Note that since the f_j are intermediary functions that must *compose* to some function $f : \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}$, we can allow them to have arbitrary domains and codomains within $\mathcal{L}(\mathcal{H})$. In particular, left-multiplication by a fixed operator is a linear map from $\mathcal{L}(\mathcal{H})$ to itself, and the trace is a linear map from $\mathcal{L}(\mathcal{H})$ to $\mathbb{R}$.

We call a function f_j from $\mathrm{Herm}(\mathcal{H})$ to itself *locally analytic* at $A \in \mathrm{Herm}(\mathcal{H})$ if, in a neighborhood of A, f can be written as a Taylor series centered at a multiple of the identity:

$$f(A) = \sum_{m=0}^{\infty} c_m (A - \alpha I)^m. \tag{2.5}$$

Examples include $f(A) = e^A$ for A Hermitian, and $f(A) = A^\alpha$ or $f(A) = \log(A)$ for A positive and α real. The derivative of a locally analytic function is given by the formula

$$\mathrm{d}f|_A(M) = \sum_j f'(\lambda_j)\Pi_j M \Pi_j + \sum_{j \neq k} \frac{f(\lambda_j) - f(\lambda_k)}{\lambda_j - \lambda_k} \Pi_j M \Pi_k, \tag{2.6}$$

where $A = \sum_j \lambda_j \Pi_j$ is the spectral decomposition of A. We give a derivation of this formula in appendix A.

Finally, we note that derivatives of matrix functions satisfy sum, product, and chain rules analogous with those of single-variable calculus. The sum and product rules are given by

$$\mathrm{d}(f+g)|_A(M) = \mathrm{d}f|_A(M) + \mathrm{d}g|_A(M) \tag{2.7}$$

and

$$\mathrm{d}(f \times g)|_A(M) = \mathrm{d}f|_A(M) g(A) + f(A)\mathrm{d}g|_A(M). \tag{2.8}$$

The chain rule says that if $f_2 : Q \to R$ and $f_1 : R \to S$ are maps of matrix manifolds, then for a point $q \in Q$ and a tangent vector $v \in T_q Q$, the composition $(f_1 \circ f_2)$ satisfies

$$\mathrm{d}(f_1 \circ f_2)|_q(v) = \mathrm{d}f_1|_{f_2(q)}[\mathrm{d}f_2|_q(v)]. \tag{2.9}$$

All three of these rules can be derived from equation (2.4) using the same proof techniques as are used to derive their analogues in single-variable calculus.

Note that even when $f_1 \circ f_2$ is a map from $\mathrm{Pos}(H)$ to $\mathbb{R}$, there is not a general chain rule for the matrix gradient $\nabla(f_1 \circ f_2)|_\rho$ (cf. equation (2.1)). This is because the gradient is only defined for matrix functions whose codomain is $\mathbb{R}$; even though $f_1 \circ f_2$ and f_1 both have codomain $\mathbb{R}$, the codomain of f_2 will generally be a matrix manifold. When we compute the matrix gradients of quantum distinguishability measures in section 3, we will always compute $\mathrm{d}f$ directly using the chain rule, then dualize to find ∇f.

3 Saturation of DPI for positive definite matrices

In this section, we show that several known "DPI saturation $\Leftrightarrow$ operator equation" results for particular distinguishability measures $\mathcal{B}$ are special cases of the universal equation

$$\mathrm{d}_1(\mathcal{B} - \mathcal{B} \circ (\Lambda \times \Lambda))|_{\rho,\sigma} = 0, \tag{3.1}$$

where the triple (ρ, σ, Λ) saturates the data processing inequality. We also give some examples of new conditions for equality arising from this equation.

In the first subsection, we explain this equation; we also put it in a more obviously computable form in theorem 3.1. In the second subsection, we apply theorem 3.1 to several specific distinguishability measures. We reproduce known results for the sandwiched Rényi relative entropies, and new results for a family of quantum f-divergences. We also calculate equation (3.1) for the general α-z Rényi relative entropies studied in [13], and compare with previous results from [23, 24].

3.1 Main theorem

As defined in the introduction, a distinguishability measure $\mathcal{B}$ on quantum states is a map $\mathcal{B}$: $\mathrm{Pos}(\mathcal{H}) \times \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}$ such that (i) $\mathcal{B}$ is differentiable as a matrix function in either of its arguments, and (ii) $\mathcal{B}$ satisfies the data processing inequality

$$\mathcal{B}(\rho, \sigma) \geq \mathcal{B}(\Lambda(\rho), \Lambda(\sigma)) \tag{3.2}$$

for any quantum channel Λ.

Let us define a new function $f_\Lambda : \mathrm{Pos}(\mathcal{H}) \times \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}$ given by

$$f_\Lambda(\rho, \sigma) = \mathcal{B}(\rho, \sigma) - \mathcal{B}(\Lambda(\rho), \Lambda(\sigma)). \tag{3.3}$$

The function f_Λ is differentiable, since both of its terms are differentiable, and is bounded below by zero. So when f_Λ reaches its minimum – i.e. when the data processing inequality is saturated – we must have $\mathrm{d}_{(1)} f_\Lambda|_{\rho,\sigma} = \mathrm{d}_{(2)} f_\Lambda|_{\rho,\sigma} = 0$, where $\mathrm{d}_{(1)}$ and $\mathrm{d}_{(2)}$ signify the manifold derivatives of f with respect to its first or second argument, respectively. As explained in section 2 in the discussion surrounding equation (2.1), vanishing of $\mathrm{d}_{(j)} f$ is equivalent to the vanishing of $\nabla_{(j)} f$, which is the unique operator in $\mathrm{Herm}(\mathcal{H})$ satisfying $\mathrm{Tr}[(\nabla_{(j)} f)M] = \mathrm{d}_{(j)} f(M)$ for all Hermitian M.

It is the vanishing of the Hermitian operator $\nabla_{(j)} f_\Lambda|_{\rho,\sigma}$ that we claim is equivalent to some existing DPI saturation conditions in the literature. In the following theorem, we provide a more convenient formula for the equation $\nabla_{(j)} f_\Lambda|_{\rho,\sigma} = 0$ and give a condition for when this equation is equivalent to DPI saturation, rather than only *implied by* DPI saturation. We prove this theorem in a very general setting, with the function $\mathcal{B}$ only required to satisfy DPI and differentiability in a neighborhood of two fixed Hermitian operators. However, our primary application is to distinguishability measures, which are smooth and satisfy the DPI for any positive operators.

Theorem 3.1. *Let $\mathcal{H}$ be a finite-dimensional Hilbert space, and $\Lambda : \mathrm{Herm}(\mathcal{H}) \to \mathrm{Herm}(\mathcal{H})$ a linear map on Hermitian operators. (In particular, Λ can be a quantum channel.) Suppose that $\mathcal{B}$ is a map from $\mathrm{Herm}(\mathcal{H}) \times \mathrm{Herm}(\mathcal{H})$ to $\mathbb{R}$, and ρ, σ are operators in $\mathrm{Herm}(\mathcal{H})$ such that $\mathcal{B}$ satisfies the data processing inequality with respect to Λ in a neighborhood of (ρ, σ). I.e., for any Hermitian operators M_1, M_2 and ϵ sufficiently small, $\mathcal{B}$ must satisfy*

$$\mathcal{B}(\rho + \epsilon M_1, \sigma + \epsilon M_2) \geq \mathcal{B}(\Lambda(\rho + \epsilon M_1), \Lambda(\sigma + \epsilon M_2)). \tag{3.4}$$

Furthermore, suppose that $\mathcal{B}$ is differentiable with respect to either of its arguments in a neighborhood of (ρ, σ), and that the data processing inequality is saturated at (ρ, σ):

$$\mathcal{B}(\rho, \sigma) = \mathcal{B}(\Lambda(\rho), \Lambda(\sigma)). \tag{3.5}$$

Then:

1. *The operator equations*

$$\nabla_{(1)} \mathcal{B}|_{\rho,\sigma} = \Lambda^* \left(\nabla_{(1)} \mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)} \right), \tag{3.6}$$

$$\nabla_{(2)} \mathcal{B}|_{\rho,\sigma} = \Lambda^* \left(\nabla_{(2)} \mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)} \right) \tag{3.7}$$

   are satisfied, where $\nabla_{(j)}$ denotes the matrix gradient with respect to the j-th argument of $\mathcal{B}$, and Λ^* is the adjoint of Λ with respect to the Hilbert-Schmidt inner product.

   (If $\mathcal{B}$ is only differentiable with respect to its j-th argument, then the corresponding equation still holds.)

2. *If $\mathrm{d}_{(1)}\mathcal{B}$ satisfies*

$$\mathrm{d}_{(1)} \mathcal{B}|_{\rho,\sigma}(\rho) = \mathcal{E}\left[\mathcal{B}(\rho, \sigma)\right], \tag{3.8}$$

   and

$$\mathrm{d}_{(1)} \mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}[\Lambda(\rho)] = \mathcal{E}\left(\mathcal{B}[\Lambda(\rho), \Lambda(\sigma)]\right), \tag{3.9}$$

   for some invertible function $\mathcal{E}$, then equation (3.6) implies that $\mathcal{B}$ satisfies the data processing inequality with respect to Λ at the point (ρ, σ). An analogous condition holds for $\mathrm{d}_{(2)}\mathcal{B}$ and equation (3.7).

Remark 3.2. The conditions (3.8) and (3.9) may seem odd, but they are quite natural features for a distinguishability measure. The derivative $d_{(1)}\mathcal{B}|_{\rho,\sigma}$ evaluated on ρ measures how the value of the distinguishability measure would change under a perturbation of the form $\rho \to (1 + \varepsilon)\rho$. Many functions have some well-defined transformation under scalar multiplication, such as $\mathcal{B}(k\rho, \sigma) = f(k)\mathcal{B}(\rho, \sigma) + g(k)$ for a positive constant k and some functions $f(k)$ and $g(k)$. If f is nonzero, then this is an affine, invertible function of $\mathcal{B}(\rho, \sigma)$. More generally, the condition just means that the distinguishability changes in some straightforward, predictible manner when one of its arguments is scaled by a constant.

Proof. 1. We have already explained in the preamble to this subsection why the equation $\mathrm{d}_{(1)} f_\Lambda(\rho,\sigma) = \mathrm{d}_{(2)} f_\Lambda(\rho,\sigma) = 0$ is implied by saturation of DPI for f_Λ defined as in equation (3.3). We need only show that this is equivalent to equations (3.6), (3.7).

Without loss of generality, we will restrict our attention to the first-argument derivative $d_{(1)}$, and for simplicity of notation we will also temporarily drop the (1) subscript. We rewrite f_Λ from equation (3.3) as

$$f_\Lambda|_{\rho,\sigma} = \mathcal{B}|_{\rho,\sigma} - [\mathcal{B}\circ(\Lambda\times \mathrm{id})]\,|_{\rho,\Lambda(\sigma)}. \tag{3.10}$$

The action of $\mathrm{d}f_\Lambda$ on a Hermitian operator M can be written using the chain rule (equation (2.9)) as

$$\mathrm{d}f_\Lambda|_{\rho,\sigma}(M) = \mathrm{d}\mathcal{B}|_{\rho,\sigma}(M) - \mathrm{d}\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}\left[d(\Lambda\times\mathrm{id})|_{\rho,\Lambda(\sigma)}(M)\right]. \tag{3.11}$$

The map $\Lambda\times\mathrm{id}$ is linear, so thanks to equation (2.3) we know that its derivative satisfies $\mathrm{d}(\Lambda\times\mathrm{id}) = \Lambda$. So we have

$$\mathrm{d}f_\Lambda|_{\rho,\sigma}(M) = \mathrm{d}\mathcal{B}|_{\rho,\sigma}(M) - \mathrm{d}\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}\left[\Lambda(M)\right]. \tag{3.12}$$

To define the gradient ∇, we dualize with respect to the Hilbert-Schmidt inner product using equation (2.1). The gradient $\nabla\mathcal{B}|_{\rho,\sigma}$ is the unique Hermitian operator satisfying

$$\mathrm{Tr}(\nabla\mathcal{B}|_{\rho,\sigma}M) = \mathrm{d}\mathcal{B}|_{\rho,\sigma}(M) \tag{3.13}$$

for all M. Using this definition, we rewrite equation (3.12) as

$$\mathrm{d}f_\Lambda|_{\rho,\sigma}(M) = \mathrm{Tr}(\nabla\mathcal{B}|_{\rho,\sigma}M) - \mathrm{Tr}[\nabla\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}\Lambda(M)]. \tag{3.14}$$

In terms of the adjoint of Λ, defined by $\mathrm{Tr}[\Lambda^*(A)B] = \mathrm{Tr}[A\Lambda(B)]$ for all Hermitian A and B, we have

$$\mathrm{d}f_\Lambda|_{\rho,\sigma}(M) = \mathrm{Tr}(\nabla\mathcal{B}|_{\rho,\sigma}M) - \mathrm{Tr}[\Lambda^*\left(\nabla\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}\right)M]. \tag{3.15}$$

Since saturation of the DPI implies that this equation vanishes for all M, we have (reinserting the subscript (1) to denote the derivative with respect to the first argument)

$$\nabla_{(1)}\mathcal{B}|_{\rho,\sigma} = \Lambda^*\left(\nabla_{(1)}\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}\right), \tag{3.16}$$

as desired.

The same argument holds for the gradient with respect to the second argument of $\mathcal{B}$, so long as $\mathcal{B}$ is differentiable in that argument in a neighborhood of ρ,σ.

2. In the previous part of this proof, we showed that equation (3.6) is equivalent to

$$\mathrm{d}_{(1)}\mathcal{B}|_{\rho,\sigma}(M) = \mathrm{d}_{(1)}\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}\left[\Lambda(M)\right]. \tag{3.17}$$

If we set $M=\rho$, this gives

$$\mathrm{d}_{(1)}\mathcal{B}|_{\rho,\sigma}(\rho) = \mathrm{d}_{(1)}\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}\left[\Lambda(\rho)\right]. \tag{3.18}$$

Applying equations (3.8) and (3.9) yields

$$\mathcal{E}\left[\mathcal{B}(\rho,\sigma)\right] = \mathcal{E}\left[\mathcal{B}\left[\Lambda(\rho),\Lambda(\sigma)\right]\right], \tag{3.19}$$

which gives $\mathcal{B}(\rho,\sigma) = \mathcal{B}\left[\Lambda(\rho),\Lambda(\sigma)\right]$ by the invertibility of $\mathcal{E}$, as desired.

An analogous argument holds for derivatives with respect to the second argument of $\mathcal{B}$. □

3.2 Examples

We now compute equations (3.6) and (3.7) for particular distinguishability measures to derive the corresponding DPI saturation conditions. We derive our first two DPI saturation conditions (the relative entropy and fidelity) in some detail to help the reader get a feel for calculating matrix derivatives of distinguishability measures. We then give a quicker computation of the saturation condition for the full family of sandwiched Rényi relative entropies, an even quicker calculation for the α-z Rényi relative entropies, and speed on through the quantum f-divergences.

3.2.1 Relative entropy

The relative entropy, defined by $D(\rho||\sigma) = \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\rho\log\sigma)$, satisfies the data processing inequality so long as ρ is positive semidefinite and σ is positive. The case when ρ is not strictly positive will be dealt with in section 4; for the moment, we assume both ρ and σ are in $\mathrm{Pos}(\mathcal{H})$.

As a function of its first argument, D can be written

$$D|_{\rho,\sigma} = \left[\mathrm{Tr}\circ\left(\mathrm{id}\times\log - R_{\log(\sigma)}\right)\right](\rho), \tag{3.20}$$

where id is the identity superoperator and $R_{\log(\sigma)}$ denotes right multiplication by $\log(\sigma)$. By successively applying the chain rule (2.9), the product rule (2.8), the linearity of the derivative (2.7), the formula for the derivative of a linear function (2.3), and the derivative of the matrix logarithm via equation (2.6), we may compute the $\mathrm{d}_{(1)}$ derivative of D as

$$\mathrm{d}_{(1)}D|_{\rho,\sigma}(M) = \mathrm{Tr}\left[\left(\log(\rho) - \log(\sigma) + \mathbb{1}\right)M\right]. \tag{3.21}$$

So the gradient of D with respect to its first argument is

$$\nabla_{(1)}D|_{\rho,\sigma} = \log(\rho) - \log(\sigma) + \mathbb{1}. \tag{3.22}$$

By theorem 3.1, DPI saturation implies the operator equation

$$\log(\rho) - \log(\sigma) = \Lambda^*\left[\log[\Lambda(\rho)] - \log[\Lambda(\sigma)]\right], \tag{3.23}$$

where we have also used that when Λ is a quantum channel, Λ^* is unital.

To see that this equation is *equivalent* to DPI saturation, we can simply left-multiply by ρ and take the trace. The fact that left-multiplying by ρ and taking the trace reproduces the relative entropy is a special case of condition (3.8) in theorem 3.1.

For completeness, we also give the vanishing-gradient equation (3.7) for the $\mathrm{d}_{(2)}$ derivative of D. As a function of its second argument, D can be written

$$D|_{\rho,\sigma} = \left[\left[-\mathrm{id} + \mathrm{Tr}(\rho\log\rho)\right]\circ\mathrm{Tr}\circ L_\rho\circ\log\right](\sigma), \tag{3.24}$$

where L_ρ denotes left-multiplication by ρ. It is straightforward to compute the $\mathrm{d}_{(2)}$ derivative of D using the derivative rules from section 2. The final result of this calculation gives the $\nabla_{(2)}$ gradient of D as

$$\nabla_{(2)}D|_{\rho,\sigma} = -\sum_{j=k}\frac{1}{\lambda_j}\Pi_j\rho\Pi_j - \sum_{j\neq k}\frac{\log(\lambda_j) - \log(\lambda_k)}{\lambda_j - \lambda_k}\Pi_k\rho\Pi_j, \tag{3.25}$$

where $\sum_j \lambda_j\Pi_j$ is the spectral decomposition of σ. The corresponding operator equation (3.7) must vanish when DPI is saturated, but this equation is much less elegant than the $\nabla_{(1)}$-gradient equation (3.23). Furthermore, it is not obviously *equivalent* to DPI saturation, since the $\mathrm{d}_{(2)}$ derivative of D does not satisfy condition (3.9) of theorem 3.1.

3.2.2 Fidelity

The fidelity [2], which we define according to the convention

$$F(\rho,\sigma) = \mathrm{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}, \tag{3.26}$$

does not satisfy the data processing inequality with the sign given in equation (1.2); however, the proof of theorem 3.1 does not actually depend on the sign of the data processing inequality, so the theorem still applies.[7]

The fidelity is symmetric, so we may compute the derivative with respect to one argument, and then obtain the other by interchanging ρ with σ. We write the fidelity as a function of its first argument as

$$F|_{\rho,\sigma} = \left(\mathrm{Tr}\circ g_{1/2}\circ\chi_{\sigma^{1/2}}\right)(\rho), \tag{3.27}$$

[7] The function f_Λ defined in the preamble to theorem 3.1 is at either a minimum or a maximum when the DPI is saturated, depending on the sign of the inequality. In either case, its derivative vanishes and the theorem holds.

where $\chi_{\sigma^{1/2}}$ denotes the conjugation $\chi_{\sigma^{1/2}}(A) = \sigma^{1/2} A \sigma^{1/2}$, and $g_{1/2}$ denotes the square root function $g_{1/2}(A) = A^{1/2}$. The trace and conjugation maps are linear, so their derivatives can be computed using equation (2.3). The square-root map $g_{1/2}$ is locally analytic, so its derivative can be computed using equation (2.6). Taking the derivative of equation (3.27) using the chain rule, simplifying with these formulas, and exploiting the cyclicity of the trace, we obtain the formula

$$\mathrm{d}_{(1)} F|_{\rho,\sigma}(M) = \frac{1}{2} \operatorname{Tr}\left(\sigma^{1/2}(\sigma^{1/2}\rho\sigma^{1/2})^{-1/2}\sigma^{1/2} M\right). \tag{3.28}$$

Dualizing according to equation (2.1) gives the gradient of F as

$$\nabla_{(1)} F|_{\rho,\sigma}(M) = \frac{1}{2}\sigma^{1/2}(\sigma^{1/2}\rho\sigma^{1/2})^{-1/2}\sigma^{1/2}, \tag{3.29}$$

a result previously shown in [26]. So by theorem 3.1, saturation of the data processing inequality for the fidelity implies the equation

$$\sigma^{1/2}(\sigma^{1/2}\rho\sigma^{1/2})^{-1/2}\sigma^{1/2} = \Lambda^* \left[\Lambda(\sigma)^{1/2}\left(\Lambda(\sigma)^{1/2}\Lambda(\rho)\Lambda(\sigma)^{1/2}\right)^{-1/2}\Lambda(\sigma)^{1/2}\right]. \tag{3.30}$$

The "second gradient" version of this equation is the same, but with ρ and σ interchanged. Each equation individually implies DPI saturation, which can be seen by a similar calculation to that given for the relative entropy in the previous example.

3.2.3 Sandwiched Rényi Relative Entropy

The sandwiched Rényi relative entropies [3, 4] are a family of distinguishability measures defined by

$$\widetilde{D}_\alpha(\rho,\sigma) = \frac{1}{\alpha-1}\log \operatorname{Tr}\left[(\sigma^\gamma\rho\sigma^\gamma)^\alpha\right] \tag{3.31}$$

with $\gamma := \frac{1-\alpha}{2\alpha}$. They are known to satisfy the data-processing inequality for $\alpha \in [1/2, \infty)$ [4]. Following the same process as in the previous two examples, we compute the gradient with respect to the first argument as

$$\boldsymbol{\nabla}_{(1)}\widetilde{D}_\alpha|_{\rho,\sigma} = \frac{\alpha}{\alpha-1}\frac{1}{\operatorname{Tr}\left[(\sigma^\gamma\rho\sigma^\gamma)^\alpha\right]}\sigma^\gamma\left(\sigma^\gamma\rho\sigma^\gamma\right)^{\alpha-1}\sigma^\gamma. \tag{3.32}$$

The gradient with respect to the second argument is more complicated; we delay the formula to the following subsection where we present it as a special case of the α-z Rényi relative entropies.

Applying theorem 3.1 to the gradient from equation (3.32) gives

$$\sigma^\gamma\left(\sigma^\gamma\rho\sigma^\gamma\right)^{\alpha-1}\sigma^\gamma = \Lambda(\sigma)^\gamma\left(\Lambda(\sigma)^\gamma\Lambda(\rho)\Lambda(\sigma)^\gamma\right)^{\alpha-1}\Lambda(\sigma)^\gamma, \tag{3.33}$$

where we have also made use of $\widetilde{D}_\alpha(\rho,\sigma) = \widetilde{D}_\alpha(\Lambda(\rho),\Lambda(\sigma))$ directly to deal with the trace term in the denominator of (3.32). The same equation was originally derived in [25], and recently rederived in [22] using different techniques.

3.2.4 α-z Rényi relative entropy

The α-z Rényi relative entropies [13] generalize the sandwiched Rényi relative entropies. They are defined by the equation

$$D_{\alpha,z}(\rho,\sigma) = \frac{1}{\alpha-1}\log\operatorname{Tr}\left[\left(\sigma^\gamma\rho^{\frac{\alpha}{z}}\sigma^\gamma\right)^z\right] \tag{3.34}$$

with $\gamma := \frac{1-\alpha}{2z}$. The α-z Rényi relative entropies satisfy the DPI for the following ranges of parameters [27]:

- $0 < \alpha < 1$ and $z \geq \max\{\alpha, 1-\alpha\}$;
- $1 < \alpha \leq 2$ and $\frac{\alpha}{2} \leq z \leq \alpha$; and

- $2 \leq \alpha < \infty$ and $\alpha - 1 \leq z \leq \alpha$.

For $\alpha = z$, this reduces to the sandwiched Rényi relative entropy, and the DPI range agrees with the one from section 3.2.3. The first and second gradients of $D_{\alpha,z}$ can be computed using the exact same techniques as for the other distinguishability measures, and are given by

$$\boldsymbol{\nabla}_{(1)} D_{\alpha,z}|_{\rho,\sigma} = \frac{z}{\alpha-1} \frac{1}{\mathrm{Tr}\left[\left(\sigma^\gamma \rho^{\frac{\alpha}{z}} \sigma^\gamma\right)^z\right]} \left[\sum_j \frac{\alpha}{z} \lambda_j^{\frac{\alpha-z}{z}} \Pi_j \sigma^\gamma \left(\sigma^\gamma \rho^{\frac{\alpha}{z}} \sigma^\gamma\right)^{z-1} \sigma^\gamma \Pi_j \right. \tag{3.35}$$

$$\left. + \sum_{j \neq k} \frac{\lambda_j^{\frac{\alpha}{z}} - \lambda_k^{\frac{\alpha}{z}}}{\lambda_j - \lambda_k} \Pi_j \sigma^\gamma \left(\sigma^\gamma \rho^{\frac{\alpha}{z}} \sigma^\gamma\right)^{z-1} \sigma^\gamma \Pi_k \right], \tag{3.36}$$

$$\boldsymbol{\nabla}_{(2)} D_{\alpha,z}|_{\rho,\sigma} = \frac{z}{\alpha-1} \frac{1}{\mathrm{Tr}\left[\left(\sigma^\gamma \rho^{\frac{\alpha}{z}} \sigma^\gamma\right)^z\right]} \left[\sum_j \gamma \mu_j^{\gamma-1} \Phi_j \left\{ \left(\sigma^\gamma \rho^{\frac{\alpha}{z}} \sigma^\gamma\right)^z, \sigma^{-\gamma} \right\} \Phi_j \right. \tag{3.37}$$

$$\left. + \sum_{j \neq k} \frac{\mu_j^\gamma - \mu_k^\gamma}{\mu_j - \mu_k} \Phi_j \left\{ \left(\sigma^\gamma \rho^{\frac{\alpha}{z}} \sigma^\gamma\right)^z, \sigma^{-\gamma} \right\} \Phi_k \right] \tag{3.38}$$

Here the spectral decompositions of ρ and σ are $\rho = \sum_i \lambda_i \Pi_i$ and $\sigma = \sum_i \mu_i \Phi_i$, and $\{A, B\}$ denotes the anticommutator $AB + BA$. Substituting these equations into equations (3.6) and (3.7) from theorem 3.1 yields an operator equality; admittedly, this equality is rather complicated in the general case $\alpha \neq z$. However, for $\alpha = z$, it is straightforward to check that the first-gradient condition coincides with equation (3.33).

Both operator equations are sufficient to imply DPI saturation, which follows from the remark after theorem 3.1, because the α-z entropy varies under scalar multiplication as

$$D_{\alpha,z}(k\rho, k'\sigma) = D_{\alpha,z}(\rho, \sigma) + \alpha \log k + (1 - \alpha) \log k'. \tag{3.39}$$

This is clearly an invertible function of $D_{\alpha,z}(\rho, \sigma)$.

This result seems to be distinct from known operator equalities from two previous works. The first of these equations was derived in [23], and is given by

$$f(\rho, \sigma) = \Lambda^\dagger(f(\Lambda(\rho), \Lambda(\sigma))), \tag{3.40}$$

$$\text{with } f(\rho, \sigma) = \sigma^{\frac{1-z}{2z}} \left(\sigma^{\frac{1-\alpha}{2z}} \rho^{\frac{\alpha}{z}} \sigma^{\frac{1-\alpha}{2z}}\right)^{z-1} \sigma^{\frac{1-z}{2z}}. \tag{3.41}$$

This result is necessary and sufficient for DPI saturation, meaning it must be equivalent to our equations; however, this equivalence is not obvious by inspection. Secondly, [24] shows a similar result where instead f is defined to be

$$f(\rho, \sigma) = \sigma^{\frac{1-\alpha}{2z}} \left(\sigma^{\frac{1-\alpha}{2z}} \rho^{\frac{\alpha}{z}} \sigma^{\frac{1-\alpha}{2z}}\right)^{\alpha-1} \sigma^{\frac{1-\alpha}{2z}}. \tag{3.42}$$

3.2.5 Quantum f-divergences

For a real-valued continuous function f on $[0, \infty)$, the quantum f-divergence is defined by [14]

$$D_f(\rho\|\sigma) := \mathrm{Tr}\left(\sqrt{\sigma} f(L_\rho R_\sigma^{-1})(\sqrt{\sigma})\right), \tag{3.43}$$

where the superoperator L_ρ acts on matrices via left multiplication by ρ and similarly R_σ acts via right multiplication by σ. This recovers the relative entropy for $f(x) = x \log x$, and the Petz-Rényi relative entropy for $f(x) = x^\alpha$. From [10] (Theorem 5.1), it is known that DPI is satisfied for any operator convex function f on $[0, \infty)$ that is sufficiently regular.

With spectral decompositions $\rho = \sum_i \lambda_i \Pi_i$ and $\sigma = \sum_j \mu_j \Phi_j$, and assuming both matrices are invertible, we have

$$D_f(\rho\|\sigma) = \sum_{i,j} \mu_j f(\lambda_i/\mu_j) \langle \Pi_i, \Phi_j \rangle, \tag{3.44}$$

from which it can be shown that gradient of D_f with respect to the first argument is

$$\nabla_{(1)} D_f(\rho\|\sigma) = \sum_{i,j} f'(\lambda_i/\mu_j)\Pi_i\Phi_j\Pi_i + \sum_j \sum_{i\neq k} \mu_j \frac{f(\lambda_i/\mu_j) - f(\lambda_k/\mu_j)}{\lambda_i - \lambda_k}\Pi_i\Phi_j\Pi_k, \tag{3.45}$$

giving us a new operator equation again according to theorem 3.1. The gradient with respect to the second argument is more complicated so we omit the expression, but it can be derived using the same technique.

In addition to the examples given here, similar expressions can be derived for the maximal f-divergences [5], the optimized f-divergences [28], the Jencova-Ruskai function [19], and any other smooth distinguishability measures.

4 Saturation of DPI for positive semidefinite matrices

In this section, we generalize the results of section 3 to distinguishability measures $\mathcal{B}$ that are defined not just on the space of positive matrices, but on the space of positive semidefinite matrices. The most general result is less aesthetically simple than the one given in theorem 3.1, but is still broadly applicable. The basic idea is that while equation (3.6) need not hold as a general operator equation when one of the matrices ρ or σ has a vanishing eigenvalue, it does hold in the orthogonal complement of the space of operators that act on the zero-eigenspace of ρ.

In the first subsection we introduce the geometry of the space of positive semidefinite operators $\mathrm{PSD}(\mathcal{H})$. In the second, we prove a generalization of theorem 3.1 that holds for matrices with vanishing eigenvalues. In the third, we work out an explicit example of this theorem: an operator equation implied by DPI saturation for the relative entropy between two density matrices when one has a nontrivial zero-eigenspace.

4.1 Geometry of the positive semidefinite space

As discussed in section 2, the Hermitian operators on an n-dimensional Hilbert space $\mathcal{H}$ form an n^2-dimensional real manifold $\mathrm{Herm}(\mathcal{H})$. The set of positive operators, $\mathrm{Pos}(\mathcal{H})$, is an n^2-dimensional real submanifold. Since $\mathrm{Pos}(\mathcal{H})$ consists of all Hermitian operators whose eigenvalues are strictly positive, its topological closure consists of all Hermitian operators whose eigenvalues are nonnegative – we denote this space of positive semidefinite operators by $\mathrm{PSD}(\mathcal{H})$. The *boundary* $\partial\,\mathrm{PSD}(\mathcal{H})$ consists of all operators that are in $\mathrm{PSD}(\mathcal{H})$ but not in $\mathrm{Pos}(\mathcal{H})$; i.e., those with at least one vanishing eigenvalue.

For an operator ρ in $\mathrm{PSD}(\mathcal{H})$, on the boundary or otherwise, the tangent space $T_\rho\,\mathrm{PSD}(\mathcal{H})$ consists of the Hermitian operators that are orthogonal to all operators acting on the zero-eigenspace of ρ, with respect to the Hilbert-Schmidt inner product. In other words, if we think of $\mathrm{Herm}(\mathcal{H})$ as a Hilbert space with respect to the Hilbert-Schmidt inner product, and denote the zero-eigenspace of ρ by $E_0(\rho)$, then $T_\rho\,\mathrm{PSD}(\mathcal{H})$ is the orthogonal complement of $\mathrm{Herm}(E_0(\rho))$ within $\mathrm{Herm}(\mathcal{H})$. We provide a simple proof of this fact in appendix B. If k is the dimension of $E_0(\rho)$, then $T_\rho\,\mathrm{PSD}(\mathcal{H})$ has real dimension $n^2 - k^2$.

4.2 Data processing equality for positive semidefinite matrices

Let $\mathcal{B}$ be a distinguishability measure such that at least one argument can be taken to be positive *semi*definite, i.e.,

$$\mathcal{B} : \mathrm{PSD}(\mathcal{H}) \times \mathrm{Pos}(\mathcal{H}) \to \mathbb{R}. \tag{4.1}$$

If we assume that $\mathcal{B}$ is continuous in the limit as its first argument is taken to the boundary $\partial\,\mathrm{PSD}(\mathcal{H})$, then the data processing inequality holds in that limit by continuity as well.

Now, let ρ and σ be density matrices, and Λ a linear map of Hermitian operators, such that the data processing inequality

$$\mathcal{B}(\rho, \sigma) \geq \mathcal{B}(\Lambda(\rho), \Lambda(\sigma)) \tag{4.2}$$

holds in a neighborhood of ρ in $\mathrm{PSD}(\mathcal{H})$. (Cf. equation (3.4).) Then, by the same arguments given in section 3, the function

$$f_{\sigma,\Lambda}(\rho) = \mathcal{B}(\rho, \sigma) - \mathcal{B}(\Lambda(\rho), \Lambda(\sigma)) \tag{4.3}$$

must be locally minimal *in any direction tangent to* $\mathrm{PSD}(\mathcal{H})$. It need not be locally minimal in directions non-tangent to $\mathrm{PSD}(\mathcal{H})$, since the DPI need not be satisfied outside of $\mathrm{PSD}(\mathcal{H})$ and $f_{\sigma,\Lambda}$ may go negative. So the equation $\mathrm{d}f_{\sigma,\Lambda}|_\rho = 0$, which was essential to our analysis in section 3, need not be satisfied for all vectors tangent to ρ.

What we have instead is the following condition: for every $M \in T_\rho \,\mathrm{PSD}(\mathcal{H})$, the derivative of $f_{\sigma,\Lambda}$ must satisfy

$$\mathrm{d}f_{\sigma,\Lambda}|_\rho(M) = 0. \tag{4.4}$$

For Hermitian operators M that are not in this tangent space, $\mathrm{d}f$ need not vanish, and in fact may not even be defined. This derivative cannot be dualized with respect to the Hilbert-Schmidt inner product on all of $\mathrm{Herm}(\mathcal{H})$, since the condition

$$\mathrm{Tr}(\nabla f_{\sigma,\Lambda}|_\rho M) = \mathrm{d}f_{\sigma,\Lambda}|_\rho(M) \qquad \forall M \in T_\rho \,\mathrm{PSD}(\mathcal{H}) \tag{4.5}$$

does not uniquely specify an operator $\nabla f_{\sigma,\Lambda}|_\rho$ on $\mathrm{Herm}(\mathcal{H})$. However, it *does* uniquely specify an operator within the tangent space $T_\rho \,\mathrm{PSD}(\mathcal{H})$.

The final, most general theorem is as follows: for a distinguishability measure $\mathcal{B}$ saturating DPI for the channel Λ at the density matrices (ρ, σ), the derivative of $f_{\sigma,\Lambda}$ as defined by equation (4.3) must vanish on $T_\rho \,\mathrm{PSD}(\mathcal{H})$. Dualizing $df_{\sigma,\Lambda}|_\rho$ with respect to the Hilbert-Schmidt inner product on $T_\rho \,\mathrm{PSD}(\mathcal{H})$ gives an operator $\nabla f_{\sigma,\Lambda}|_\rho$ in $T_\rho \,\mathrm{PSD}(\mathcal{H})$ which must vanish. For $\rho \in \mathrm{Pos}(\mathcal{H})$, this gives theorem 3.1 as a special case with $T_\rho \,\mathrm{PSD}(\mathcal{H}) \cong \mathrm{Herm}(\mathcal{H})$.

By the same argument given in the proof of theorem 3.1, we can show that the vanishing of this derivative is equivalent to the equation

$$\nabla_{(1)}\mathcal{B}|_{\rho,\sigma} = \Lambda^*(\nabla_{(1)}\mathcal{B}|_{\Lambda(\rho),\Lambda(\sigma)}), \tag{4.6}$$

keeping in mind that $\nabla_{(1)}\mathcal{B}|_{\rho,\sigma}$ is now an element of $T_\rho \,\mathrm{PSD}(\mathcal{H})$, i.e., the subspace of operators that are Hilbert-Schmidt orthogonal to all linear operators acting on the zero-eigenspace of ρ.

Before proceeding to an explicit example, we emphasize that if the zero-eigenspace of ρ is k-dimensional, then based on the considerations from the previous subsection, our operator equation restricted to $T_\rho \,\mathrm{PSD}(\mathcal{H})$ still gives $n^2 - k^2$ scalar equations governing the matrix entries of ρ and σ. While the strongest operator equation implied by DPI saturation is in the case $k = 0$, we still get many constraints on ρ and σ when n is sufficiently large and k sufficiently small.

4.3 Example: relative entropy

Let $\mathcal{H}$ be an n-dimensional Hilbert space, and let ρ be a density matrix on $\mathcal{H}$ with eigendecomposition

$$\rho = \sum_{j=1}^{m} \lambda_j \, |\psi_j\rangle\langle\psi_j| \tag{4.7}$$

with $m < n$, $\lambda_j > 0$, and $\{|\psi_j\rangle\}$ orthonormal. If we choose any orthonormal basis of $\mathcal{H}$ containing $\{|\psi_j\rangle\}$, then ρ can be written in this basis as the block matrix

$$\rho = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}, \tag{4.8}$$

where Δ is a diagonal, $m \times m$ matrix whose entries are the eigenvalues $\{\lambda_j\}$.

Now, let σ be an arbitrary positive operator on $\mathcal{H}$. If ρ were positive, then we could apply the result of section 3.2.1 to show that the equation $D(\rho||\sigma) = D(\Lambda(\rho)||\Lambda(\sigma))$ implies

$$\log(\rho) - \log(\sigma) = \Lambda^*[\log(\Lambda(\rho)) - \log(\Lambda(\sigma))]. \tag{4.9}$$

Since ρ is positive semidefinite, this equation doesn't even make sense – the logarithm of ρ is ill-defined. However, thanks to the previous subsection, we know that it *does* make sense within the subspace of operators orthogonal to those operators acting on the zero-eigenspace of ρ. In the basis of equation (4.8), this set of operators is given by block matrices of the form

$$M = \begin{pmatrix} A & B \\ B^\dagger & 0 \end{pmatrix} \tag{4.10}$$

with A Hermitian. So we conclude that (4.9) holds as a matrix equation in the $\{|\psi_j\rangle\}$ basis outside of the bottom-right block.

Note that equation (4.9) is sufficient to imply DPI saturation for ρ and σ even though it only holds away from the bottom right block in the $\{|\psi_j\rangle\}$ basis. If we left-multiply by ρ and take the trace, the result is independent of whatever we might put in the bottom right block, which is just the familiar statement that $\mathrm{Tr}(\rho\log\rho)$ is a well defined for positive semidefinite operators even when $\log\rho$ is not.

5 Discussion

We now comment on some potential directions for future work.

5.1 Approximate DPI saturation

Our method gives, for any distinguishability measure $\mathcal{B}$ and any quantum channel Λ, an operator equation satisfied by ρ and σ whenever the data processing inequality is *exactly* saturated. As a reminder to the reader, the essential feature of this argument is that when the DPI is saturated, the function

$$f_\Lambda(\rho,\sigma) = \mathcal{B}(\rho,\sigma) - \mathcal{B}(\Lambda(\rho),\Lambda(\sigma)) \tag{5.1}$$

is at its minimum value, and so its gradient with respect to either argument must vanish.

When the DPI is *approximately* saturated, f_Λ is *close* to its minimum value. We believe that for the distinguishability measures commonly studied in the literature, f_Λ being close to its minimum value may also imply that its gradient is small in some appropriate sense. If this is the case, then approximate DPI saturation still implies an operator equation (smallness of the gradient). We are currently investigating this line of reasoning.

5.2 Connection to Petz recovery

Saturation of the data processing inequality is famously linked to the existence recovery channels. In [15, 16], Petz showed that saturation of the data processing inequality for the relative entropy for density matrices ρ and σ and a quantum channel Λ implies the existence of a ρ-independent recovery channel $\mathcal{R}_\sigma$ satisfying

$$[\mathcal{R}_\sigma\circ\Lambda](\rho) = \rho, \tag{5.2}$$

$$[\mathcal{R}_\sigma\circ\Lambda](\sigma) = \sigma. \tag{5.3}$$

Petz's original proof technique relied heavily on functional analysis. Since in this paper we have used geometric arguments to circumvent functional analytic proofs of certain consequences of DPI saturation, we wonder whether it might be possible to use similar ideas to provide an alternate proof of Petz's theorem without using any functional analysis.

For the moment, we have no concrete suggestions for how this might be done. However, we recall from section 3.2.3 that for the sandwiched Rényi relative entropies, the operator equation implied by DPI saturation is

$$\sigma^\gamma\left(\sigma^\gamma\rho\sigma^\gamma\right)^{\alpha-1}\sigma^\gamma = \Lambda^*[\Lambda(\sigma)^\gamma\left(\Lambda(\sigma)^\gamma\Lambda(\rho)\Lambda(\sigma)^\gamma\right)^{\alpha-1}\Lambda(\sigma)^\gamma]. \tag{5.4}$$

For $\alpha = 2$ (i.e., $\gamma = -1/4$), this simplifies to

$$\sigma^{-1/2}\rho\sigma^{-1/2} = \Lambda^*\left[\Lambda(\sigma)^{-1/2}\Lambda(\rho)\Lambda(\sigma)^{-1/2}\right]. \tag{5.5}$$

This is exactly the condition Petz originally produced in [15, 16]![8] It tells us that the Petz map

$$\mathcal{R}_\sigma(\bullet) = \sigma^{1/2}\Lambda^*\left[\Lambda(\sigma)^{-1/2}\Lambda(\bullet)\Lambda(\sigma)^{-1/2}\right]\sigma^{1/2} \tag{5.6}$$

[8]This observation was made previously in [25]. We include it for the purpose of discussion, rather than claiming it as an original result.

perfectly recovers ρ.

We find it suggestive that the vanishing gradient equation for the sandwiched 2-Rényi relative entropy is exactly equivalent to Petz recovery. While this is enough to prove that DPI saturation for the sandwiched 2-Rényi relative entropy implies recoverability, it is not enough to prove that DPI saturation for the relative entropy implies recoverability. For that, it seems you would still need some functional analytic input telling you that DPI saturation for relative entropy is equivalent to DPI saturation for the sandwiched 2-Rényi relative entropy. However, we remain optimistic that geometric techniques can provide an alternate path to Petz recovery.

Finally, we comment that in [11], it was shown that approximate DPI saturation implies approximate recoverability for a family of distinguishability measures that includes the sandwiched Rényi relative entropies for $\alpha \in [1/2, 1) \cup (1, \infty)$. If it is possible to use our methods to derive operator equations implied by approximate DPI saturation, as mentioned in the previous subsection, it is possible that these techniques could give a new perspective on approximate Petz recovery as well.

5.3 Nice equations for α-z Rényi relative entropies

We find it unsatisfying that while our method produces the exact same equations derived in [25] for the sandwiched Rényi relative entropies, it does not manifestly produce the same equations for the α-z Rényi relative entropies that were derived in [23, 24]. In fact, the equations produced by our method, (3.35) and (3.37), are significantly more complicated than the ones derived in [23, 24]. However, as mentioned in section 3.2.4, our equation must be equivalent at least to the equation from [23], since both are equivalent to DPI saturation. We remain hopeful that some extra geometric input may make it possible to simplify our equations (3.35) and (3.37) and to understand how it is related to the equations from [23, 24].

Acknowledgments

We thank Patrick Hayden and Mark Wilde for insightful conversations. The authors are supported by the Simons Foundation It from Qubit collaboration, by AFOSR (FA9550-16-1-0082), and by DOE Award No. DE-SC0019380.

A Derivatives of locally analytic functions

As in section 2, we will call $f : \mathrm{Herm}(\mathcal{H}) \to \mathrm{Herm}(\mathcal{H})$ *locally analytic* at $A \in \mathrm{Herm}(\mathcal{H})$ if, in a neighborhood of A, f can be written as a Taylor series centered at a multiple of the identity:

$$f(A) = \sum_{m=0}^{\infty} c_m (A - \alpha I)^m. \tag{A.1}$$

To linear order in ϵ, we have

$$f(A + \epsilon M) = f(A) + \epsilon \sum_{m=0}^{\infty} c_m \sum_{n=0}^{m-1} (A - \alpha I)^n M (A - \alpha I)^{m-1-n}. \tag{A.2}$$

Let $\{\lambda_j\}$ be the eigenvalues of A and $\{\Pi_j\}$ the orthogonal projectors onto its eigenspaces. Substituting in the spectral decomposition

$$A - \alpha I = \sum_j (\lambda_j - \alpha) \Pi_j, \tag{A.3}$$

we find

$$\mathrm{d}f|_A(M) = \lim_{\epsilon \to 0} \frac{f(A + \epsilon M) - f(A)}{\epsilon} = \sum_{j,k} \Pi_j M \Pi_k \left[\sum_{m=0}^{\infty} c_m \sum_{n=0}^{m-1} (\lambda_j - \alpha)^n (\lambda_k - \alpha)^{m-1-n} \right]. \tag{A.4}$$

For $j \neq k$, we have $\lambda_j \neq \lambda_k$, and in this case the sum over n in (A.4) simplifies to[9]

$$\sum_{n=0}^{m-1} (\lambda_j - \alpha)^n (\lambda_k - \alpha)^{m-1-n} = \frac{(\lambda_j - \alpha)^m - (\lambda_k - \alpha)^m}{\lambda_j - \lambda_k}. \tag{A.5}$$

Splitting the j, k sum in (A.4) into terms with $j = k$ and terms with $j \neq k$, we have

$$\begin{aligned} \mathrm{d}f|_A(M) = &\sum_j \Pi_j M \Pi_j \sum_{m=0}^{\infty} c_m m (\lambda_j - \alpha)^{m-1} \\ &+ \sum_{j \neq k} \Pi_j M \Pi_k \sum_{m=0}^{\infty} c_m \frac{(\lambda_j - \alpha)^m - (\lambda_k - \alpha)^m}{\lambda_j - \lambda_k}. \end{aligned} \tag{A.6}$$

Using the Taylor series expansions for $f(\lambda_j)$ and $f(\lambda_k)$, this simplifies to

$$\mathrm{d}f|_A(M) = \sum_j f'(\lambda_j) \Pi_j M \Pi_j + \sum_{j \neq k} \frac{f(\lambda_j) - f(\lambda_k)}{\lambda_j - \lambda_k} \Pi_j M \Pi_k. \tag{A.7}$$

This is equation (2.6) from the main text.

B Tangent spaces to positive semidefinite operators

Let $\mathcal{H}$ be an n-dimensional Hilbert space, and $\mathrm{PSD}(\mathcal{H})$ the space of positive semidefinite operators. A Hermitian operator M is tangent to $\mathrm{PSD}(\mathcal{H})$ at ρ if the eigenvalues of $\rho + \epsilon M$ are all nonnegative at linear order in ϵ. For M to be in the tangent *space* of $\mathrm{PSD}(\mathcal{H})$ at ρ, this same property must hold for all real multiples of M. The only time we can have M tangent to $\mathrm{PSD}(\mathcal{H})$ at ρ but not in the tangent space of ρ is when ρ is on the boundary $\partial \, \mathrm{PSD}(\mathcal{H})$ – if one of the zero-eigenvalues of ρ becomes positive at linear order when perturbing by M, then M is tangent to $\mathrm{PSD}(\mathcal{H})$, but a perturbation by $-M$ will cause the zero-eigenvalue to become negative, so $-M$ is non-tangent to $\mathrm{PSD}(\mathcal{H})$.

We conclude that for $\rho \in \mathrm{PSD}(\mathcal{H})$, a Hermitian operator M is in the tangent space $T_\rho \, \mathrm{PSD}(\mathcal{H})$ if and only if the zero-eigenvalues of ρ remain zero at linear order in the perturbation $\rho + \epsilon M$. For general Hermitian M, the zero-eigenspace of ρ will split into several eigenspaces at linear order in $\rho + \epsilon M$. What this means is that the projector Π_0 onto the zero-eigenspace of ρ can be decomposed into projectors

$$\Pi_0 = \sum_j \hat{\Pi}_j, \tag{B.1}$$

where each is continuously connected to an eigenspace projector of $\rho + \epsilon M$ by the formula

$$\Pi_j^{(\rho + \epsilon M)} = \hat{\Pi}_j + \epsilon \delta \hat{\Pi}_j. \tag{B.2}$$

The eigenvalue equation $(\rho + \epsilon M) \Pi_j^{(\rho+\epsilon M)} = \lambda_j^{(\rho+\epsilon M)} \Pi_j^{(\rho+\epsilon M)}$, evaluated at linear order in ϵ, gives

$$\rho \delta \hat{\Pi}_j + M \hat{\Pi}_j = \delta \lambda_j \hat{\Pi}_j. \tag{B.3}$$

Left-multiplying by $\hat{\Pi}_k$ and using $\hat{\Pi}_k \rho = 0$ gives

$$\hat{\Pi}_k M \hat{\Pi}_j = \delta_{jk} \delta \lambda_j \hat{\Pi}_j. \tag{B.4}$$

If $\mathcal{O}_0$ is a Hermitian operator that acts entirely within the zero-eigenspace of ρ, then we have

$$\mathrm{Tr}(M \mathcal{O}_0) = \mathrm{Tr}(M \Pi_0 \mathcal{O}_0 \Pi_0) = \sum_{j,k} \mathrm{Tr}(M \Pi_j \mathcal{O}_0 \Pi_k) = \sum_j \delta \lambda_j \, \mathrm{Tr}(\hat{\Pi}_j \mathcal{O}_0). \tag{B.5}$$

[9]You can check this formula by multiplying both sides by $(\lambda_j - \alpha) - (\lambda_k - \alpha)$. The sum on the left-hand side of the equation becomes a telescoping series; after the series telescopes, the only remaining terms are the ones in the numerator on the right-hand side.

So we see that when M is in $T_\rho\,\mathrm{PSD}(\mathcal{H})$, meaning none of the zero-eigenvalues of ρ change at linear order in $\rho + \epsilon M$, we must have $\delta\lambda_j = 0$ and therefore $\langle M, \mathcal{O}_0\rangle_{\mathrm{HS}} = 0$.

The argument of the preceding paragraph shows that whenever M is tangent to $\mathrm{PSD}(\mathcal{H})$ at ρ, it is orthogonal (in the sense of the Hilbert-Schmidt inner product) to every Hermitian operator acting on the zero-eigenspace of ρ. Conversely, if M is orthogonal to every operator in the zero-eigenspace of ρ, then it is in particular orthogonal to each projection operator $\hat{\Pi}_j$, which gives $\delta\lambda_j = 0$ for the corresponding eigenvalue by equation (B.5). We conclude that the tangent space to $\mathrm{PSD}(\mathcal{H})$ at ρ is exactly the set of Hermitian operators that are orthogonal to all operators acting on the zero-eigenspace of ρ.

Another convenient way of conceptualizing this tangent space is that it is the space of all operators that can be written

$$M = \Delta\rho + \rho\Delta^\dagger \tag{B.6}$$

for a general linear operator $\Delta \in \mathcal{L}(\mathcal{H})$.[10] It is straightforward to verify that every operator M of the form given in equation (B.6) is orthogonal to every $\mathcal{O}_0 \in \mathrm{Herm}(E_0(\rho))$, so every operator of the form (B.6) must lie within $T_\rho\,\mathrm{PSD}(\mathcal{H})$. The map

$$\Delta \mapsto \Delta\rho + \rho\Delta^\dagger \tag{B.7}$$

is an $\mathbb{R}$-linear map from $\mathcal{L}(\mathcal{H})$ to $T_\rho\,\mathrm{PSD}(\mathcal{H})$. A calculation we omit here shows that its kernel is $(n^2 + k^2)$-dimensional, which tells us that its image must be $(2n^2 - (n^2 + k^2))$-dimensional. This is the same dimension as $T_\rho\,\mathrm{PSD}(\mathcal{H})$, so we conclude that *every* M in $T_\rho\,\mathrm{PSD}(\mathcal{H})$ can be written in the form (B.6).

References

[1] A. Uhlmann, *The "transition probability" in the state space of a *-algebra*, *Reports on Mathematical Physics* **9** (1976), no. 2 273–279.

[2] A. Uhlmann, *The Transition Probability for States of *-Algebras*, *Annalen der Physik* **497** (1985), no. 4-6 524–532.

[3] M. M. Wilde, A. Winter, and D. Yang, *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Renyi relative entropy*, *Communications in Mathematical Physics* **331** (2014), no. 2 [arXiv:1306.1586].

[4] M. M uller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, *On quantum Renyi entropies: a new generalization and some properties*, *Journal of Mathematical Physics* **54** (2013), no. 12 [arXiv:1306.3142].

[5] K. Matsumoto, *A new quantum version of f-divergence*, arXiv:1311.5722.

[6] K. Matsumoto, *Reverse test and quantum analogue of classical fidelity and generalized fidelity*, arXiv:1006.0302.

[7] S. S. Cree and J. Sikora, *A fidelity measure for quantum states based on the matrix geometric mean*, arXiv:2006.06918.

[8] A. S. Kholevo, *On quasiequivalence of locally normal states*, *Theoretical and Mathematical Physics* **13** (1972), no. 2 1071–1082.

[9] F. Hiai and M. Mosonyi, *Different quantum f-divergences and the reversibility of quantum operations*, *Reviews in Mathematical Physics* **29** (2017), no. 07 [arXiv:1604.03089].

[10] F. Hiai, M. Mosonyi, D. Petz, and C. Beny, *Quantum f-divergences and error correction*, *Reviews in Mathematical Physics* **23** (2011), no. 07 [arXiv:1008.2529].

[11] L. Gao and M. M. Wilde, *Recoverability for optimized quantum f-divergences*, arXiv:2008.01668.

[12] H. Umegaki, *Conditional expectation in an operator algebra. IV. Entropy and information*, *Kodai Mathematical Seminar Reports* **14** (1962), no. 2 59–85.

[13] K. M. R. Audenaert and N. Datta, *alpha-z-relative Renyi entropies*, *Journal of Mathematical Physics* **56** (2015), no. 2 022202, [arXiv:1310.7178].

[10] We have not cited any sources in this appendix because the geometry of $T_\rho\,\mathrm{PSD}(\mathcal{H})$ is farly simple and we did not consult any sources in deriving its properties; however, we credit chapter 5 of [29] for introducing us to the analogue of equation (B.6) for the space of positive semidefinite *real* matrices.

[14] D. Petz, *Quasi-entropies for States of a von Neumann Algebra*, *Publications of the Research Institute for Mathematical Sciences, Kyoto University* **21** (1985) 787.

[15] D. Petz, *Sufficient subalgebras and the relative entropy of states of a von neumann algebra*, *Communications in mathematical physics* **105** (1986), no. 1 123–131.

[16] D. Petz, *Sufficiency of channels over von neumann algebras*, *The Quarterly Journal of Mathematics* **39** (1988), no. 1 97–108.

[17] A. Jenčová, *Rényi relative entropies and noncommutative L_p-spaces*, *Annales Henri Poincaré* **19** (2018), no. 8 2513–2542, [arXiv:1609.08462].

[18] A. Jenčová, *Rényi relative entropies and noncommutative L_p-spaces II*, *Annales Henri Poincaré* **19** (2018), no. 8 [arXiv:1707.00047].

[19] A. Jenčová and M. B. Ruskai, *A Unified Treatment of Convexity of Relative Entropy and Related Trace Functions, with Conditions for Equality*, *Reviews in Mathematical Physics* **22** (2010), no. 09 [arXiv:0903.2895].

[20] A. Jenčová, *Preservation of a quantum Renyi relative entropy implies existence of a recovery map*, *Journal of Physics A: Mathematical and Theoretical* **50** (2017), no. 8 [arXiv:1604.02831].

[21] H. F. Jia and M. Rangamani, *Petz reconstruction in random tensor networks*, arXiv:2006.12601.

[22] J. Wang and H. Wilming, *Revisiting the equality conditions of the data processing inequality for the sandwiched Rényi divergence*, arXiv:2009.14197.

[23] S. Chehade and A. Vershynina, *Saturating the Data Processing Inequality for α-z Renyi Relative Entropy*, 2006.0772.

[24] H. Zhang, *Equality conditions of Data Processing Inequality for α-z Rényi relative entropies*, arXiv:2007.06644.

[25] F. Leditzky, C. Rouzé, and N. Datta, *Data processing for the sandwiched Rényi divergence: a condition for equality*, arXiv:1604.02119.

[26] B. Coutts, M. Girard, and J. Watrous, *Certifying optimality for convex quantum channel optimization problems*, arXiv:1810.13295.

[27] H. Zhang, *From Wigner-Yanase-Dyson conjecture to Carlen-Frank-Lieb conjecture (New title)*, arXiv:1811.01205.

[28] M. M. Wilde, *Optimized quantum f-divergences and data processing*, *Journal of Physics A: Mathematical and Theoretical* **51** (Sept., 2018) 374002, [arXiv:1710.10252].

[29] U. Helmke and J. B. Moore, *Optimization and dynamical systems.* Springer Science & Business Media, 2012.