Monotone Riemannian metrics and relative entropy on noncommutative probability spaces

Andrew Lesniewski
Paribas Capital Markets, The Equitable Tower,
787 Seventh Avenue, New York, New York 10019

Mary Beth Ruskai[a]
Department of Mathematics, University of Massachusetts-Lowell, Lowell, MA 01854

We use the relative modular operator to define a generalized relative entropy for any convex operator function g on $(0,\infty)$ satisfying $g(1)=0$. We show that these convex operator functions can be partitioned into convex subsets, each of which defines a unique symmetrized relative entropy, a unique family (parametrized by density matrices) of continuous monotone Riemannian metrics, a unique geodesic distance on the space of density matrices, and a unique monotone operator function satisfying certain symmetry and normalization conditions. We describe these objects explicitly in several important special cases, including $g(w) = -\log w$, which yields the familiar logarithmic relative entropy. The relative entropies, Riemannian metrics, and geodesic distances obtained by our procedure all contract under completely positive, trace-preserving maps. We then define and study the maximal contraction associated with these quantities. © *1999 American Institute of Physics.* [S0022-2488(99)01410-3]

I. INTRODUCTION

For quantum systems, a state is described by a density matrix P, i.e., a positive semidefinite operator with trace one. We will let $\bar{\mathcal{D}}$ denote the set of *density matrices*. For classical discrete or commutative systems we can identify the states with the subset of diagonal density matrices, each of which defines a probability vector $p \in \mathbf{R}^n$. For commutative systems the usual logarithmic relative entropy,

$$H_{\log}(p,q) = \sum_k p_k \log(p_k/q_k), \tag{1}$$

can be generalized to

$$h_g(p,q) = \sum_k p_k g(q_k/p_k), \tag{2}$$

where g is a convex function on $(0,\infty)$ with $g(1)=0$. It is well known that any such H_g contracts under stochastic mappings, i.e., $H_g(Ap,Aq) \leq H_g(p,q)$ when A is a column stochastic matrix. Cohen *et al.*[1] defined the entropy contraction coefficient as

$$\eta_g(A) = \sup_{p \neq q} \frac{H_g(Ap,Aq)}{H_g(p,q)}. \tag{3}$$

a)Electronic mail: bruskai@cs.uml.edu

5702

In the pair of papers,[1,2] it was shown that for each fixed A all the contraction coefficients associated with those g that are also *operator* convex are equivalent, more precisely the following.

Theorem I.1: *If g is operator convex, then*

$$\eta_g(A) = \eta_{\log}(A) = \eta_{(w-1)^2}(A) \leqslant \eta_{|w-1|}(A). \tag{4}$$

A summary of these results is given in Ref. 3. It suffices to mention here that the observation,

$$\left. \frac{d^2}{dt^2} H_g(p, p+tv) \right|_{t=0} = g''(0) \sum_k (v_k)^2/p_k = H_{(w-1)^2}(p, p+v), \tag{5}$$

plays a critical role. The quantity $\Sigma_k(v_k)^2/p_k$ can also be written as $M_p(v,v)$, where

$$M_p(u,v) = -\left. \frac{\partial^2}{\partial\alpha\,\partial\beta} H_g(p+\alpha u, p+\beta v) \right|_{\alpha=\beta=0} \tag{6}$$

is the Riemannian metric corresponding to the Fisher information. Čencov[4,5] showed that, for commutative systems, this is the *only* Riemannian metric that satisfies the monotonicity condition $M_{Ap}(Av, Av) \leqslant M_p(v,v)$. Thus, we can regard Theorem I.1 as stating that for operator convex g the maximal contraction of the relative entropy and its associated Riemannian metric are the same. Since there is only one Riemannian metric, all the contraction coefficients must be equal.

For quantum systems, the usual logarithmic relative entropy is given by

$$H_{\log}(P,Q) = \operatorname{Tr} P(\log P - \log Q) \tag{7}$$

$$= \int_0^\infty \operatorname{Tr} P\left[\frac{1}{Q+tI}(P-Q)\frac{1}{P+tI} \right] dt, \tag{8}$$

with P, Q in $\mathcal{D}$, the set of invertible density matrices. The integral representation (8) can be used to show that

$$M_P^{\log}(A,B) \equiv -\left. \frac{\partial^2}{\partial\alpha\,\partial\beta} H_{\log}(P+\alpha A, Q+\beta B) \right|_{\alpha=\beta=0}$$

$$= \int_0^\infty \operatorname{Tr} A\left[\frac{1}{P+tI} B \frac{1}{P+tI} \right] dt. \tag{9}$$

Although $M_P^{\log}(A,B)$ is a monotone Riemannian metric, it is not the only possibility; $M_P(A,B) = \operatorname{Tr} A^* P^{-1} B$ is also monotone under completely positive, trace-preserving maps. The study of monotone Riemannian metrics on noncommutative probability spaces was initiated by Morozova and Čencov,[6] who did not, however, provide any explicit examples. A complete characterization of monotone Riemannian metrics (which includes the examples above) was given recently by Petz.[7–10] The quantum structure is much richer because left and right multiplications by P^{-1} are not equivalent. We will see that $M_P(A,B)$ can always be written in the form $\operatorname{Tr} A^* \Omega_P(B)$, where Ω_P reduces to multiplication by P^{-1} when P and B commute. Thus, for example, (9) above gives $\Omega_P(B) = \int_0^\infty [1/(P+tI)]B[1/(P+tI)] dt$, which becomes $P^{-1}B$ when P and B commute.

Earlier, Ruskai[3] tried to extend the entropy contraction coefficient results of Cohen *et al.* to noncommutative situations, but obtained only a few preliminary results. Although one can formally define $H_g(P,Q) = \operatorname{Tr} Pg(Q/P)$ the expression Q/P is ambiguous in the quantum case. Using the nonstandard definition $Q/P = P^{-1/2}QP^{-1/2}$ [which yields $H_g(P,Q) = \operatorname{Tr} P \log P^{-1/2}QP^{-1/2}$ rather than (7) when $g(w) = -\log w$] Ruskai and Petz[11] were able to prove an analog of Theorem I.1, using the fact that

$$\left.\frac{d^2}{dt^2}H_g(P,P+tA)\right|_{t=0}=g''(0)\mathrm{Tr}\,AP^{-1}A,\tag{10}$$

for all g. In essence, their convention for Q/P always yields the Riemannian metric $M_P(A,B)=\mathrm{Tr}\,A^*P^{-1}B$.

A better alternative is to use the relative modular operator introduced by Araki[12–16,7] to define Q/P. This yields the usual logarithmic entropy (7) and a rich family of generalized relative entropies. Moreover, differentiation then yields the entire family of monotone Riemannian metrics found by Petz.[7–10]

In this paper we use the relative modular operator to study both the relative entropies and Riemannian metrics associated with convex operator functions. For simplicity, we restrict ourselves to the matrix algebras associated with finite dimension systems. Although we do not believe this restriction is essential, it avoids many technical complications. [The most serious arises when the condition $\mathrm{Tr}\,P=1$ is not compatible with the requirement that P be invertible (in the sense of having a bounded inverse in the relevant operator algebra). In that case, one must restrict the domain of $H_g(P,Q)$ to those pairs P,Q that have comparable approximate null spaces in some suitable sense.] We show that each convex operator function defines a convex family of relative entropies, a unique symmetrized relative entropy, a unique family (parametrized by density matrices) of continuous monotone Riemannian metrics, a unique geodesic distance on the space of density matrices, and a unique monotone operator function. We describe these objects explicitly in several important special cases, including $g(w)=-\log w$. We then define and study the contraction coefficient associated with the relative entropy, Riemannian metrics, and metrics. Finally, we present examples showing that these contraction coefficients can have any value in $[0,1]$ for a suitable stochastic map.

The paper is organized as follows. In Sec. II, we give some basic definitions and results for relative entropy and Riemannian metrics. In Sec. III, we define the corresponding geodesic distance, including the Bures metric as a special case. Finally, in Sec. IV we study the contraction of all the quantities under stochastic maps and give bounds on the maximal contraction.

II. RELATIVE ENTROPY AND RIEMANNIAN METRICS

A. Definitions

We begin by describing the relative modular operator that was originally introduced by Araki to generalize the logarithmic relative entropy to type III von Neumann algebras.[12–16,7] Later, Petz[16] used it to generalize relative entropy itself. Let $\mathcal{D}$ denote the subset of invertible operators in $\mathcal{D}$. Let $P,Q\in\mathcal{D}$, i.e., P and Q are positive definite matrices with $\mathrm{Tr}(P)=\mathrm{Tr}(Q)=1$. For matrix algebras, the relative modular operator associated with the pair of states $\rho_P(A)=\mathrm{Tr}(AP)$ and $\rho_Q(A)=\mathrm{Tr}(AQ)$ reduces to

$$\Delta_{Q,P}=L_Q R_P^{-1},\tag{11}$$

where L_Q and R_P are the left and right multiplication operators, respectively. Thus, $\Delta_{Q,P}(A)=QAP^{-1}$. It is easy to verify directly that $\Delta_{Q,P}$ is a positive Hermitian operator with respect to the Hilbert–Schmidt inner product.

Definition II.1: Let g be an operator convex function defined on $(0,\infty)$ such that $g(1)=0$. The relative g entropy of P and Q is

$$H_g(P,Q)=\mathrm{Tr}(P^{1/2}g(\Delta_{Q,P})P^{1/2}).\tag{12}$$

We will let $\mathcal{G}$ denote the set of functions satisfying these conditions. Note, however, that the argument of g, as defined here, is shifted from that (which we here denote g_C) in Refs. 1 and 2 so that $g_C(w)=g(w+1)$. Using standard results from the theory of monotone and convex operator functions, one can show that $\mathcal{G}$ is the class of functions that can be written in the form

$$g(w) = a(w-1) + b(w-1)^2 + c\frac{(w-1)^2}{w} + \int_0^\infty \frac{(w-1)^2}{w+s} d\nu(s), \tag{13}$$

where $b, c > 0$ and ν is a positive measure on $(0, \infty)$ with finite mass $\int_0^\infty d\nu(s)$. The term $(w-1)^2/w$ may seem unfamiliar, as it is usually included implicitly in the integral. However, writing it separately will be convenient later and is necessary to ensure that the measure has finite mass. The function $g(w) = -\log w$ yields the usual logarithmic relative entropy (7), which we continue to denote as $H_{\log}(P, Q)$. The function $g(w) = (w-1)^2$ yields

$$H_{(w-1)^2} = \mathrm{Tr}(P-Q)P^{-1}(P-Q), \tag{14}$$

which we call the "quadratic relative entropy;" it plays an extremely important role in our development. The function $g(w) = (w-1)^2/(w+1)$ yields the equally important, but less familiar, $H_{\mathrm{Bures}}(P, Q) = \mathrm{Tr}(P-Q)[L_Q + R_P]^{-1}(P-Q)$, where we use the subscript Bures because (as will be explained in Sec. III) it eventually leads to a geodesic on $\mathcal{D}$ referred to as the "metric of Bures."

We will study the properties of relative entropy and related quantities under a class of maps referred to as "stochastic" based on the concept of a completely positive, trace-preserving map $\phi: \mathcal{A}_1 \rightarrow \mathcal{A}_2$ from one operator algebra to another. Such maps are linear and positivity preserving, not only on the original algebra, but when lifted to tensors products. A precise definition and useful representation theorems can be found in Stinespring,[17] Arveson,[18] Choi,[19] Kraus,[20] and Lindblad.[21] In particular, ϕ is a completely positive map if and only if there exist operators $\{V_k\}$ with $V_k: \mathcal{A}_1 \rightarrow \mathcal{A}_2$, such that

$$\phi(A) = \sum_{k=1}^N V_k A V_k^* . \tag{15}$$

It then follows that ϕ is trace preserving if and only if $\Sigma_k V_k^* V_k = I$; whereas ϕ is unital [i.e., $\phi(I_1) = I_2$] if and only if $\Sigma_k V_k V_k^* = I$. For von Neumann algebras with finite trace (as is the case here) one can use the Hilbert–Schmidt inner product $\langle A, B \rangle = \mathrm{Tr}\, A^* B$ to define the adjoint $\hat{\phi}$ of any completely positive map so that $\mathrm{Tr}\, A^* \phi(B) = \mathrm{Tr}\, \hat{\phi}(A)^* B$. It is then easy to see that $\hat{\phi}(A) = \Sigma_k V_k^* A V_k$ and that ϕ is trace preserving if and only if $\hat{\phi}$ is unital. Then, for the finite-dimensional algebras considered here, we can use the following.

Definition II.2: A stochastic map $\phi: \mathcal{A}_1 \rightarrow \mathcal{A}_2$ is a completely positive, trace-preserving map.

(For general von Neumann algebras, a stochastic map should be defined as the dual of a unital completely positive map.) For commutative systems, a stochastic map always corresponds to a column stochastic matrix, as discussed in the Introduction above and in Refs. 1–3. For noncommutative systems, a partial trace (see Sec. IV D or, e.g., Refs. 22, 21) is an example of a stochastic map.

B. Relative entropy

We begin by defining a relative entropy distance as a bilinear function on $\mathcal{D}$ with the properties we expect of the relative g entropy $H_g(P, Q)$. It is sometimes convenient to extend our definition from $\mathcal{D} \times \mathcal{D}$ to the somewhat larger set of pairs P, Q of positive definite matrices with $\mathrm{Tr}\, P = \mathrm{Tr}\, Q$.

Definition II.3: By a relative entropy distance we mean a function $H(P, Q)$ satisfying the following.

(a) $H(P, Q) \geq 0$ *with $H(P, Q) = 0 \Leftrightarrow P = Q$.*
(b) $H(\lambda P, \lambda Q) = \lambda H(P, Q)$ *for $\lambda > 0$.*
(c) $H(P, Q)$ *is jointly convex in P and Q.*
    In addition, we say that the relative entropy is monotone *if*

(d) $H(P,Q)$ *decreases under stochastic maps* ϕ,
       that it is symmetric *if*

(e) $H(P,Q)=H(Q,P)$,
       and that it is differentiable *if*

(f)   *the function* $g(x,y)=H(P+xA,Q+yB)$ *is differentiable.*

Conditions (b), (c), and (d) are not independent. It is well known that by embedding $\mathbf{C}^{n\times n}$ in $\mathbf{C}^{n\times n}\otimes\mathbf{C}^{2\times 2}$ and choosing ϕ to correspond to the partial trace over $\mathbf{C}^2$, one can show that (d) implies the subadditivity relation,

$$H(P_1+P_2,Q_1+Q_2)\leqslant H(P_1,Q_1)+H(P_2,Q_2).\tag{16}$$

But for functions satisfying the homogeneity condition (b) this is equivalent to joint convexity. Because any stochastic map can be represented as a partial trace,[21] it follows that when (a) and (b) hold, then (c)$\Leftrightarrow$(d). Nevertheless, the properties of convexity and monotonicity are each of sufficient importance to justify explicitly stating them separately.

A relative entropy distance (even if symmetric) is not a metric in the usual sense, because it need not satisfy the triangle inequality. Nevertheless, such quantities have been widely used[23–25] to measure the difference between P and Q. Later, we shall show that every relative g entropy defines a relative entropy distance that then defines a Riemannian metric and an associated geodesic distance.

Theorem II.4: *Every relative g entropy of the form given in Definition II.1 is a differentiable monotone relative entropy distance in the sense of Definition II.3.*

Proof: Properties (a), (b), and (f) are straightforward; (d) is due to Petz[16] and implies (c) by the above remarks. A simple new proof of (d) is given in Sec. II F.

Theorem II.5: *For each operator convex function* $g\in\mathcal{G}$,

$$H_g(P,Q)=\mathrm{Tr}(Q-P)[b_g P^{-1}+c_g Q^{-1}](Q-P),\tag{17}$$

$$\int_0^\infty \mathrm{Tr}\left((Q-P)\frac{1}{L_Q+sR_P}(Q-P)\right)d\nu_g(s)$$

$$=\mathrm{Tr}[(Q-P)R_P^{-1}g(\Delta_{QP})(Q-P)],\tag{18}$$

where b_g, c_g, *and* ν_g *are as in (13).*

Proof: We first observe that

$$(\Delta_{Q,P}-I)(P^{1/2})=(Q-P)P^{-1/2}=R_{P^{-1/2}}(Q-P),\tag{19}$$

so that

$$H_{w-1}(P,Q)=\mathrm{Tr}[P^{1/2}(Q-P)P^{-1/2}]=0,\tag{20}$$

and the linear term in (13) does not contribute. We also find using (19) again,

$$H_g(P,Q)=\langle(\Omega_{Q,P}-I)(P^{1/2}),(\Delta_{Q,P}+sI)^{-1}(\Delta_{Q,P}-I)(P^{1/2})\rangle$$

$$=\mathrm{Tr}[(Q-P)(\Delta_{Q,P}+sI)^{-1}R_{P^{-1}}(Q-P)]$$

$$=\mathrm{Tr}(Q-P)\frac{1}{L_Q+sR_p}(Q-P).\tag{21}$$

Letting $s=0$ yields

$$H_{(w-1)^2/w}(P,Q) = \text{Tr}[(Q-P)Q^{-1}(Q-P)] = H_{(w-1)^2}(Q,P), \tag{22}$$

and one easily verifies that

$$H_{(w-1)^2}(P,Q) = \text{Tr}((Q-P)P^{-1}(Q-P)). \tag{23}$$

Using these results in (13) gives the desired result (17).

It is worth pointing out that the cyclicity of the trace implies that

$$\text{Tr}(Q-P)\frac{1}{R_P+sL_Q}(Q-P) = \text{Tr}(Q-P)\frac{1}{L_P+sR_Q}(Q-P), \tag{24}$$

although

$$\text{Tr}(Q-P)\frac{1}{R_P+sL_Q}(Q-P) \neq \text{Tr}(Q-P)\frac{1}{R_Q+sL_P}(Q-P),$$

in general.

One can also use the heat kernel representation,

$$(\Delta_{Q,P}+sI)^{-1} = \int_0^\infty e^{-u(\Delta_{Q,P}+sI)}\,du, \tag{25}$$

to obtain another integral representation of $H_g(P,Q)$.

Theorem II.6: *Let $m_g(u) = \int_0^\infty e^{-us}d\nu(s)$ denote the Laplace transform of the measure ν_g. Then*

$$H_g(P,Q) = b_g H_{(w-1)^2}(P,Q) + c_g H_{(w-1)^2}(Q,P) + \int_0^\infty H_{(w-1)^2 e^{-uw}}(P,Q)m_g(u)du,$$

where we formally extend our definition of $H_g(P,Q)$ to the nonconvex function $g(w) = (w-1)^2 e^{-uw}$.

Proof: We use (25) in (13) to get

$$\int_0^\infty \langle (\Delta_{Q,P}-I)(P^{1/2}),(\Delta_{Q,P}+sI)^{-1}(\Delta_{Q,P}-I)(P^{1/2})\rangle d\nu_g(s)$$

$$= \int_0^\infty \langle (\Delta_{Q,P}-I)(P^{1/2}),e^{-u\Delta_{Q,P}}(\Delta_{Q,P}-I)(P^{1/2})\rangle m_g(u)du$$

$$= \int_0^\infty \text{Tr}((Q-P)(R_{P^{-1}}e^{-u\Delta_{Q,P}})(Q-P))m_g(u)du$$

$$= \int_0^\infty H_{(w-1)^2 e^{-uw}}(P,Q)m_g(u)du,$$

where we have interchanged the order of integration and then used (19) again.

C. Monotone Riemannian metrics

We now consider the relation between relative g entropy and Riemannian metrics. Note that the set of density matrices $\mathcal{D}$ has a natural structure as a smooth manifold, so that we can define a Riemannian metric on its tangent bundle $T_*\mathcal{D}$, whose fibers consist of traceless, self-adjoint matrices or

$$T_P\mathcal{D} = \{A = A^*: \text{Tr}\,A = 0\}. \tag{26}$$

Definition II.7: By a Riemannian metric on $\mathcal{D}$, we mean a positive definite bilinear form $M_P(A,B)$ on $T_P\mathcal{D}$ such that the map $P \rightarrow M_P(A,A)$ is smooth for each fixed $A \in T_*\mathcal{D}$. The metric is monotone if it contracts under stochastic maps in the sense

$$m_{\phi(P)}[\phi(A),\phi(B)] \leq M_P(A,B), \tag{27}$$

when ϕ is a stochastic map.

Note that this definition of monotone requires that the stochastic map ϕ act on the base point (i.e., the indexing density matrix P) as well as the arguments of the bilinear form.

Theorem II.8: For each $g \in \mathcal{G}$ and density matrix $P \in \mathcal{D}$,

$$M_P^g(A,B) = -\frac{\partial^2}{\partial\alpha\,\partial\beta} H_g(P+\alpha A, P+\beta B)\bigg|_{\alpha=\beta=0}, \tag{28}$$

$$= \langle A, \Omega_P^g(B) \rangle = \operatorname{Tr} A_P^g(B), \tag{29}$$

defines a Riemannian metric on $T_P\mathcal{D}$, and a positive linear operator Ω_P^g on $T_P\mathcal{D}$.

The theorem follows easily from the fact that R_P, L_P and their inverses are positive semidefinite operators with respect to the Hilbert–Schmidt inner product, e.g., $\operatorname{Tr} A^* R_P A > 0$, and the integral representation in Theorem II.5. We find

$$\langle A, \Omega_P^g(B) \rangle = (b_g + c_g)\operatorname{Tr}[AL_P^{-1}(B) + BL_P^{-1}(A)]$$

$$+ \int_0^\infty \operatorname{Tr}[A(L_P+sR_P)^{-1}(B) + B(L_P+sR_P)^{-1}(A)]d\nu_g(s)$$

$$= (b_g + c_g)\operatorname{Tr} A[L_P^{-1} + R_P^{-1}](B)$$

$$+ \int_0^\infty \operatorname{Tr} A[(L_P+sR_P)^{-1} + (R_P+sL_P)^{-1}](B)d\nu_g(s)$$

$$= \int_0^\infty \operatorname{Tr} A[(L_P+sR_P)^{-1} + (R_P+sL_P)^{-1}](B)N_g(s)ds$$

$$= \left\langle A, \int_0^\infty [(L_P+sR_P)^{-1} + (R_P+sL_P)^{-1}](B)N_g(s)ds \right\rangle, \tag{30}$$

where, for simplicity, we temporarily subsume the quadratic terms into the integral by defining N_g so that $N_g(s)ds = (b_g + c_g)\delta(s)ds + d\nu_g(s)$. It is critical that A and B are self-adjoint so that we can interchange A and B by replacing L_P by R_P as in

$$\operatorname{Tr} BL_P^{-1}(A) = \operatorname{Tr} BP^{-1}A = \operatorname{Tr} ABP^{-1} = \operatorname{Tr} AR_P^{-1}(B). \tag{31}$$

This result would not hold if we did not require the perturbations of P and Q to be self-adjoint. Given that requirement, the result is necessarily symmetric in the sense that we get the same result from both $H_g(P,Q)$ and $H_g(Q,P)$. This is already evident in the quadratic term, whose coefficient depends only on the sum $b+c$, and will be discussed further below.

We can now use (30) to obtain several explicit formulas for Ω_P^g:

$$\Omega_P^g = \int_0^\infty \left(\frac{1}{sR_P+L_P} + \frac{1}{sL_P+R_P}\right)N_g(s)ds \tag{32}$$

$$= \int_0^\infty \frac{1}{sR_P+L_P}(N_g(s) + s^{-1}N_g(s^{-1}))ds \tag{33}$$

$$= R_P^{-1} \int_0^\infty \frac{1}{s + \Delta_{P,P}} \sigma_g(s) \, ds \tag{34}$$

$$= \int_0^1 \left(\frac{1}{sR_P + L_P} + \frac{1}{sL_P + R_P} \right) \sigma_g(s) \, ds, \tag{35}$$

where we have used the change of variable $s \to s^{-1}$ and

$$\sigma_g(s) = N_g(s) + s^{-1} N_g(s^{-1}).$$

Note that $\sigma_g(s^{-1}) = s \sigma_g(s)$. Then, if we define

$$k(\lambda) = \int_0^\infty \frac{1}{s + \lambda} \sigma_g(s) \, ds = \int_0^1 \left[\frac{1}{s + \lambda} + \frac{1}{s\lambda + 1} \right] \sigma_g(s) \, ds, \tag{36}$$

we find that $k(\lambda^{-1}) = \lambda k(\lambda)$, $\Omega_P^g = R_P^{-1} k(\Delta_{P,P})$, and that k can be expressed in terms of g as

$$k(w) = \frac{g(w) + w g(w^{-1})}{(w-1)^2}. \tag{37}$$

We will let $\mathcal{K}$ denote this set of functions, i.e.,

$$\mathcal{K} = \{k : -k \text{ is operator monotone}, \quad k(w^{-1}) = w k(w), \quad \text{and} \quad k(1) = 1\}. \tag{38}$$

We have recovered half of Petz's result[7–10] that there is a one-to-one correspondence between symmetric Riemannian metrics and functions of the form (36) that satisfy the normalization condition $k(1) = 1$. (But note that our k corresponds to $1/f$ in Petz's notation.) Our approach also easily yields an explicit expression for both Ω_P^g and its inverse.

Theorem II.9: *For each $g \in \mathcal{G}$ and $P \in \mathcal{D}$, the operator Ω_P^g as defined in Theorem II.8 satisfies $\Omega_P^g = R_P^{-1} k(L_P R_P^{-1})$ and $[\Omega_P^g]^{-1} = R_P f(L_P R_P^{-1})$, where $k(w)$ is given by (37) and $f(w) = 1/k(w)$.*

Although Ω_P^g is initially defined only on $T_* \mathcal{D}$, it can easily be extended to all traceless matrices using the natural complexification $\operatorname{Tr} A = 0 \Rightarrow A = A_1 + i A_2$ with $A_1, A_2 \in T_P \mathcal{D}$ and then to all of $\mathbf{C}^{n \times n}$ using linearity and $\Omega_P^g(I) = P^{-1} I$. The result is equivalent to using any of the formulas for Ω_P^g above together with the obvious extension of L_P and R_P to all of $\mathbf{C}^{n \times n}$. We can summarize this discussion as follows.

Theorem II.10: *For each $g \in \mathcal{G}$ and $P \in \mathcal{D}$, the operator Ω_P^g as defined in Theorem II.8 can be extended to a positive linear operator on $\mathbf{C}^{n \times n}$ so that $M_P^g(A, B) = \operatorname{Tr} A^* \Omega_P^g(B)$ defines an inner product on $\mathbf{C}^{n \times n}$. On the other hand, for each $g \in \mathcal{G}$ and $P \in \mathcal{D}$, Eq. (34) defines a positive linear operator Ω_P^g on all of $\mathbf{C}^{n \times n}$, and the bilinear form $M_P^g(A, B) = \operatorname{Tr} A^* \Omega_P^g(B)$ extends to a monotone Riemannian metric satisfying the symmetry condition $M_P^g(A, B) = M_P^g(B^*, A^*)$.*

This result is essentially due to Petz,[7–10] who also showed the converse result that every symmetric monotone Riemannian metric is of this form. We give an independent proof of monotonicity at the end of this section. That the metric is symmetric is a consequence of the cyclicity of the trace.

The following result is essentially due to Kubo and Ando,[26] who developed a theory of operator means.

Theorem II.11: *If k given by (36) satisfies $k(1) = 1$, then for all $P, Q \in \mathcal{D}$,*

$$R_P^{-1} + L_Q^{-1} \geq R_P^{-1} k(\Delta_{Q,P}) \geq (R_P + L_Q)^{-1}. \tag{39}$$

Proof: This follows easily from (36), the elementary inequality,

$$\frac{w+1}{2w} \geq \frac{1+t}{2}\left[\frac{1}{t+w}+\frac{1}{tw+1}\right] \geq \frac{2}{w+1}, \tag{40}$$

and the fact that the normalization $k(1)=1$ implies that $2\sigma_g(t)/(t+1)$ is a probability measure on $[0,1]$.

As immediate corollaries, we find

$$\Omega_P^{(w-1)^2}=L_P^{-1}+R_P^{-1} \geq \Omega_P^\sigma \geq (R_P+L_P)^{-1}=\Omega_p^{\text{Bures}}, \tag{41}$$

$$M_P^{(w-1)^2}(A,A) \geq M_P^g(A,A) \geq M_P^{\text{Bures}}(A,A), \tag{42}$$

$$H_{(w-1)^2}^{\text{sym}}(P,Q) \geq H_g^{\text{sym}}(P,Q) \geq H_{\text{Bures}}(P,Q), \tag{43}$$

where the superscript indicates the symmetric relative entropy associated with g. Thus $k(w) = 2/(w+1)$ corresponds to the minimum symmetric relative entropy and minimum Riemannian metric among the class studied here. By contrast, we will see that $g(w)=(w-1)^2$ corresponds to $k(w)=(w+1)/(2w)$ so that the quadratic relative entropy is maximal.

The operators Ω_P^g and $[\Omega_P^g]^{-1}$ are noncommutative versions of multiplication by P^{-1} and P, respectively. Hence, in view of the cyclicity of the trace, the following result is not surprising.

Theorem II.12: *The operator* Ω_P^g *given by (34) satisfies* $\text{Tr}\,\Omega_P^g(A)=\text{Tr}\,AP^{-1}$ *and* $\text{Tr}[\Omega_P^g]^{-1}(A)=\text{Tr}\,AP$.

Proof: We first observe that in a basis in which P is diagonal with eigenvalues p_k,

$$\left[R_P^{-1}\frac{1}{s+\Delta_{P,P}}(A)\right]_{jk}=\left[\frac{1}{sR_P+L_P}(A)\right]_{jk}=\frac{1}{sp_k+p_j}a_{jk}, \tag{44}$$

so that

$$[\Omega_P^g(A)]_{jk}=\int_0^\infty \frac{a_{jk}}{sp_k+p_j}\sigma_g(s)ds. \tag{45}$$

Then for every $g \in \mathcal{G}$, $P \in \mathcal{D}$, and $A \in T_P\mathcal{D}$,

$$\text{Tr}\,\Omega_P^g(A)=\sum_j \int_0^\infty \frac{a_{jj}}{sp_j+p_j}\sigma_g(s)ds$$

$$=\sum_j p_j^{-1}a_{jj}\int_0^\infty \frac{1}{s+1}\sigma_g(s)ds=k(1)\text{Tr}\,P^{-1}A=\text{Tr}\,P^{-1}A.$$

The proof for the inverse is similar. Since $1/k$ is also operator monotone, we can use Theorem II.9 to conclude that $[\Omega_P^g]^{-1}$ can be written in the form

$$[\Omega_P^g]^{-1}=aR_P+bL_P-\int_0^\infty \frac{R_P^2}{sR_P+L_P}d_\mu(s),$$

for some positive measure μ.

D. Correspondence between defining functions

We now make some remarks on the relation between $g(w)$, $wg(w^{-1})$, and $k(w)$. It should be clear from the development above that every function $g \in \mathcal{G}$ defines a Riemannian metric and a function k as in (36) or (37). If we now consider $\hat{g}(w)=wg(w^{-1})$, it is easy to verify that $\hat{g}(w) \in \mathcal{G}$ as well and that $H_{\hat{g}}(P,Q)=H_g(Q,P)$. Thus, the map $g(w) \rightarrow wg(w^{-1})$ has the effect of

switching the arguments of the relative entropy and the function $g(w)+wg(w^{-1})$ yields the symmetrized relative entropy $H_g(P,Q)+H_g(Q,P)$. Now, if we begin with a function g and relative entropy $H_g(P,Q)$, the differentiation in (28) automatically yields a symmetric result. Thus, all convex combinations $ag(w)+(1-a)\hat{g}(w)$ of g and $\hat{g}(w)$ yield the same Riemannian metric and the same function $k\in\mathcal{K}$.

Conversely, every $k\in\mathcal{K}$ defines a unique *symmetric* relative entropy via the function $g^{\mathrm{sym}}(w)=(w-1)^2 k(w)$. It follows immediately from the integral representation (37) and (13) that $g^{\mathrm{sym}}(w)$ is also in $\mathcal{G}$ and that $wg^{\mathrm{sym}}(w^{-1})=g^{\mathrm{sym}}(w)$. Thus, k selects from the convex set of relative entropies associated with a given $g\in\mathcal{G}$, the symmetric one. If we observe that the integral representation (36) is equivalent to $-k$ being an operator monotone function, we can summarize the discussion above as follows.

Theorem II.13: *There is a one-to-one correspondence between each of the following.*

(a) *Monotone Riemannian metrics extended to bilinear forms via the symmetry condition* $M_P^g(A,B)=M_P^g(B^*,A^*)$.
(b) *Monotone (decreasing) operator functions satisfying* $k(w^{-1})=wk(w)$ *with the normalization* $k(1)=1$.
(c) *Convex operator functions in $\mathcal{G}$ that satisfy the symmetry relation* $wg(w^{-1})=g(w)$.

The relations between these are given by (34), (36), and (37). In view of this theorem, it would be appropriate to identify a given operator Ω_P^g by using the (unique) symmetric function g^{sym}. However, we will continue to use the asymmetric g for such familiar cases as the logarithm. One might expect the one-to-one correspondence to extend to twice-differentiable symmetric monotone relative entropies. However, Petz and Ruskai[11] consider relative entropies of the form $\tilde{H}_g(P,Q)=\mathrm{Tr}\,Pg(P^{-1/2}QP^{-1/2})$. This class of monotone relative entropies can be symmetrized, however, differentiation of H_g yields the Riemannian metric $M_P^{(w-1)^2}(A,B)=\mathrm{Tr}\,A^*[P^{-1}B+BP^{-1}]$ for all $g\in\mathcal{G}$. Thus, in particular $\tilde{H}_{\log}(P,Q)=\mathrm{Tr}\,P\log(P^{-1/2}QP^{-1/2})$ is an example of a relative entropy distance (in the sense of Definition II.3) that is *not* a relative g entropy. Another class of distinct relative entropy distances is given by squares of the geodesic distances introduced in Sec. III. Thus, the properties in Definition 1.3 are *not* sufficient to completely characterize the relative g entropy and allow us to extend the one-to-one correspondence in Theorem 2.4 to a class of relative entropies. Although we believe that such an additional condition must exist, we have not found it.

E. Examples

We now give explicit expressions for the relative entropy, Ω_P^g and related quantities in several important special cases. These examples will also illustrate the relation between the functions g, $\hat{g}$, g^{sym}, and k discussed above.

Example 1: Take $g(w)=-\log w$. Then $\hat{g}(w)=w\log w$, $g^{\mathrm{sym}}=(w-1)\log w$, $k(w)=(w-1)^{-1}\log w$, $N_g(s)=(s+1)^{-2}$ and $\sigma_g(s)=1/(s+1)$. Then, $H_{\log}(P,Q)$ is given by (7), and

$$H_{\log}^{\mathrm{sym}}=H_{(w-1)\log w}=\mathrm{Tr}(P-Q)[\log P-\log Q] \tag{46}$$

$$=\int_0^\infty \mathrm{Tr}(P-Q)\frac{1}{Q+xI}(P-Q)\frac{1}{P+xI}dx \tag{47}$$

and

$$\Omega_P^{\log}=\int_0^\infty\left(\frac{1}{sR_P+L_P}+\frac{1}{sL_P+R_P}\right)\frac{1}{(s+1)^2}ds=\int_0^\infty\frac{1}{s+1}\frac{1}{L_P+sR_P}ds.$$

Making the change of variables $s\to sR_P$ in the last integral yields

$$\Omega_P^{\log} = \int_0^\infty \frac{1}{s+L_P} \frac{1}{s+R_P} \, ds, \tag{48}$$

so that

$$\langle A, \Omega_P^{\log}(B) \rangle = \mathrm{Tr} \int_0^\infty A^* \frac{1}{sI+P} B \frac{1}{sI+P} \, ds, \tag{49}$$

a result that we obtained earlier (9) using the integral representation (8), or

$$\log P - \log Q = \int_0^\infty \left[\frac{1}{Q+xI} - \frac{1}{P+xI} \right] dx. \tag{50}$$

In this case, it is also well known[27,15] that the inverse operator can be written as

$$[\Omega_P^{\log}]^{-1}(B) = \int_0^1 P^t B P^{1-t} \, dt. \tag{51}$$

Example 2: Take $g(w) = (w-1)^2$. Then $\hat{g}(w) = (w-1)^2/w$, $g^{\mathrm{sym}}(w) = (w-1)^2(w+1)/w$ and $k(w) = (w+1)/(2w)$. Then $H_{(w-1)^2}(P,Q)$ is given by (14),

$$H_{(w-1)^2}^{\mathrm{sym}}(P,Q) = \mathrm{Tr}(Q-P)[P^{-1}+Q^{-1}](Q-P),$$

$$\Omega_P^{(w-1)^2} = R_P^{-1} + L_P^{-1},$$

and

$$\langle A, \Omega_P^{(w-1)^2}(A) \rangle = H_{(w-1)^2}(P, P+A) = \mathrm{Tr}\, A P^{-1} A.$$

The associated function is the maximal function satisfying the prescribed conditions. The operator $\Omega_P^{(w-1)^2}(B) = P^{-1}B + BP^{-1}$, so that

$$\Omega_P^{(w-1)^2} = R_P^{-1} + L_P^{-1} = R_P^{-1}[R_P + L_P]L_P^{-1}. \tag{52}$$

Example 3: For $s_0 > 0$ take $g_{s_0}(w) = (w-1)^2/(w+s_0)$. Then $\hat{g}_{s_0}(w) = (w-1)^2/(1+ws_0)$, $g^{\mathrm{sym}}(w) = (w-1)^2(w+1)(1+s_0)/(1+ws_0)(w+s_0)$, $k(w) = (w+1)(1+s_0)/(1+ws_0)(w+s_0)$, and $N_g(s) = \delta(s-s_0)$. Thus

$$\Omega_P^{g_{s_0}} = \frac{1}{s_0 R_P + L_P} + \frac{1}{s_0 L_P + R_P} = (s_0+1)[s_0 R_P + L_P]^{-1}[R_P + L_P][R_P + s_0 L_P]^{-1}. \tag{53}$$

When no confusion will result, it will be convenient to employ a slight abuse of notation and write $\Omega_P^{s_0}$ for $\Omega_P^{g_{s_0}}$. The case $s_0 = 1$ is particularly important; we have already seen that it yields the minimal $k \in \mathcal{K}$. Then $k(w) = 2/(1+w)$, $g(w) = g^{\mathrm{sym}}(w) = (w-1)^2/(w+2)$, and $\Omega_P^{g_{s_0=1}} \equiv \Omega_P^{\mathrm{Bures}} = [R_P + L_P]^{-1}$. The corresponding Riemannian metric is $\langle A, \Omega_P^{\mathrm{Bures}}(B) \rangle = \mathrm{Tr}\, A^* [R_P + L_P]^{-1}(B)$ and the corresponding relative entropy,

$$H_{\mathrm{Bures}}(P,Q) = \mathrm{Tr}(Q-P)[R_P + L_Q]^{-1}(Q-P) = \mathrm{Tr}\, QXPX, \tag{54}$$

where $X = [R_P + L_Q]^{-1}(Q-P)$. Because of the cyclicity of the trace, $H_{\mathrm{Bures}}(P,Q)$ is already symmetric and $[R_Q + L_P]^{-1}$ would have given the same result.

Example 4: Take $g(w) = 1 - w^\alpha$. Then $k(w) = (1-w^\alpha)(1-w^{1-\alpha})/\alpha(1-\alpha)(1-w)^2$ and $N_g(s) = (\sin \pi s/\pi)(1+s)^{\alpha-2}$. Thus

$$H_{1-w^{\alpha}}(P,Q)=1-\operatorname{Tr} Q^{\alpha}P^{1-\alpha},$$

$$\Omega_P^g=R_P^{-1}\int_0^{\infty}\frac{1}{sI+\Delta_{P,P}}\frac{\sin\pi s}{\pi}(1+s)^{\alpha-2}\,ds.$$

After the change of variables $s\rightarrow sR_P$, this becomes

$$\Omega_P^g=\frac{\sin\pi s}{\pi}\int_0^{\infty}\frac{1}{L_P+s}\frac{R_P^{1-\alpha}+s^{1-\alpha}}{(R_P+s)^{2-\alpha}}\,ds. \tag{55}$$

F. Monotonicity proof

We now present a new proof of the monotonicity of the relative entropies and Riemannian metrics associated with convex operator functions.

Theorem II.14: *For every convex operator function g of the type considered here, both the relative entropy $H_g(P,Q)$ and the corresponding Riemannian metric are monotone, i.e.,*

$$H_g(P,Q)\geq H_g[\phi(P),\phi(Q)], \tag{56}$$

$$\langle A\Omega_P^g A\rangle\geq\langle\phi(A)\Omega_{\phi(P)}^g\phi(A)\rangle. \tag{57}$$

This result is essentially due to Petz.[16] We given an independent proof as an immediate corollary of the following theorem and the integral representations (17) and (34).

Theorem II.15: *If ϕ is stochastic,*

$$\operatorname{Tr}A^*\frac{1}{R_P+sL_Q}A=\operatorname{Tr}\phi\left(A^*\frac{1}{R_P+sL_Q}A\right)\geq\operatorname{Tr}\phi(A^*)\frac{1}{R_{\phi(P)}+sL_{\phi(Q)}}\phi(A). \tag{58}$$

Proof: If $P>0$, then $\operatorname{Tr}A^*PA\geq0$, and $\operatorname{Tr}A^*AP\geq0$, so that both L_P and R_P are positive as operators on the Hilbert–Schmidt space. Thus, for $Q>0$, the operator R_P+sL_Q is also positive. Let $X=[R_P+sL_Q]^{-1/2}(A)-[R_P+sL_Q]^{1/2}\hat{\phi}(B)$ with $B=[R_{\phi(P)}+sL_{\phi(Q)}]^{-1}\phi(A)$. Then $\operatorname{Tr}X^*X\geq0$, so that

$$\operatorname{Tr}A^*\frac{1}{R_P+sL_Q}A-\operatorname{Tr}A^*\hat{\phi}(B)-\operatorname{Tr}\hat{\phi}(B^*)A+\operatorname{Tr}\hat{\phi}(B^*)[R_P+sL_Q]\hat{\phi}(B)\geq0. \tag{59}$$

Since it is easy to see that

$$-\operatorname{Tr}A^*\hat{\phi}(B)-\operatorname{Tr}\phi(B^*)A=-2\operatorname{Tr}\phi(A^*)\frac{1}{R_{\phi(P)}+sL_{\phi(Q)}}\phi(A),$$

the desired result will follow if we can show that the last term in (59) is bounded above by the right side of (58). We find

$$\operatorname{Tr}\hat{\phi}(B^*)[R_P+sL_Q]\hat{\phi}(B)=\operatorname{Tr}\hat{\phi}(B^*)\hat{\phi}(B)P+\hat{\phi}(B^*)sQ\hat{\phi}(B)$$

$$=\operatorname{Tr}\hat{\phi}(B^*)\hat{\phi}(B)P+\hat{\phi}(B)\hat{\phi}(B^*)sQ$$

$$\leq\operatorname{Tr}\hat{\phi}(B^*B)P+\hat{\phi}(BB^*)sQ,$$

where the inequality follows from the positivity of P and Q and the operator inequality:

$$\hat{\phi}(B^*)\hat{\phi}(B)\leq\hat{\phi}(B^*B), \tag{60}$$

which holds for any B because the trace-preserving condition on ϕ gives $\hat{\phi}(I_2)=I_1$. Then using, e.g., $\mathrm{Tr}\,\hat{\phi}(B^*B)P=\mathrm{Tr}\,B^*B\phi(P)$, we find

$$\mathrm{Tr}\,\hat{\phi}(B^*)[R_P+sL_Q]\hat{\phi}(B)\leq \mathrm{Tr}\,B^*B\phi(P)+BB^*s\phi(Q)$$
$$=\mathrm{Tr}\,B^*[B\phi(P)+s\phi(Q)B]$$
$$=\mathrm{Tr}\,B^*[R_{\phi(P)}+sL_{\phi(Q)}]B=\mathrm{Tr}\,B^*\phi(A)$$
$$=\mathrm{Tr}\,\phi(A^*)\frac{1}{R_{\phi(P)}+sL_{\phi(Q)}}\phi(A).$$

It is interesting to observe that the strategy used here is very similar to that used by Lieb and Ruskai[22] to prove a Schwarz inequality for completely positive mappings and, as a special case, the monotonicity of the quadratic relative entropy. At that time, Lieb and Ruskai could use these Schwarz inequalities to prove many special cases of the strong subadditivity of the logarithmic relative entropy, but not the general case. A complete proof of strong subadditivity[28] (see also Refs. 3, 29) seemed to require one of the convex trace function theorems of Lieb.[27] It is therefore curious that now, some 25 years later, we have finally found a way to recover strong subadditivity directly from the Schwarz strategy of Lieb and Ruskai.[22]

It should also be noted that Uhlmann had earlier[30] used a very different approach (based on interpolation theory) to show the logarithmic relative entropy was monotone under a related class of mappings that are Schwarz in the sense $\phi(A^*A)\geq \phi(A^*)\phi(A)$ and Petz[16] extended this to other relative entropies.

III. GEODESIC DISTANCE

We now wish to consider the contraction of the relative entropy and corresponding Riemannian metric under stochastic mappings. Before doing so, it will be useful to consider the geodesic distance that arises from the Riemannian metrics considered here.

Definition III.1: Associated with every Riemannian metric $\langle A,\Omega_P^g(B)\rangle$ of the form (28) is a geodesic distance $D_g(P,Q)$, which is defined as

$$D_g(P,Q)\equiv \inf\int_0^1\sqrt{\langle \dot{S}(t),\Omega_{S(t)}^g\dot{S}(t)\rangle}dt,$$

where the infimum is taken over all smooth paths $S(t)$ with $S(0)=P$ and $S(1)=Q$.

Theorem III.2: *The square $[D_g(P,Q)]^2$ of every geodesic distance of the form given in Definition III.1 is a differentiable monotone relative entropy distance in the sense of Definition II.3. In addition, $D_g(P,Q)$ satisfies the triangle inequality $D_g(P,R)\leq D_g(P,Q)+D_g(Q,R)$.*

Proof: Properties (a), (b), and (e) of Definition II.3 are readily verified. Property (d), i.e., the monotonicity $D_g[\phi(P),\phi(Q)[\leq D_g(P,Q)$, can be proven directly, but also follows easily as a corollary to Theorem IV.2 below. The triangle inequality is standard. That $D_g(P+xA,Q+yB)$ is differentiable in the sense of Definition II.3(f) follows from standard results (see, e.g., Theorem 3.6, part (2) of Ref. 31.   Q.E.D.

It is well known (see, e.g., Refs. 32–36) that the metric associated with the minimal function $k(w)=2/(1+w)$, discussed in Example 3, is (except for normalization) the metric of Bures, i.e., $D_{2(w-1)^2/(1+w)}(P,Q)=4D^{\mathrm{Bures}}(P,Q)$, where

$$[D^{\mathrm{Bures}}(P,Q)]^2=\inf\{\mathrm{Tr}(W-X)(W-X)^*:WW^*=P,XX^*=Q\}$$
$$=2[1-\mathrm{Tr}(\sqrt{P}Q\sqrt{P})^{1/2}] \tag{61}$$
$$\leq \mathrm{Tr}[\sqrt{P}-\sqrt{Q}]^2=2[1-\mathrm{Tr}\sqrt{P}\sqrt{Q}]=H_{1-\sqrt{w}}(P,Q). \tag{62}$$

It follows immediately from (41) that

$$D^{\text{sym}}_{(w-1)^2}(P,Q) \geq D_g(P,Q) \geq 4D^{\text{Bures}}(P,Q), \tag{63}$$

so that $4D^{\text{Bures}}(P,Q)$ gives the minimal geodesic distance of this type.

IV. CONTRACTION UNDER STOCHASTIC MAPS

A. Contraction coefficients

Because the relative entropies, Riemannian metrics, and geodesic distances all contract under stochastic maps, their maximal contraction is a well-defined quantity in the following sense.

Definition IV.1: For each fixed convex operator function g of the form given in Def. II.1 and stochastic map ϕ, we define three entropy contraction coefficients:

$$\eta_g^{\text{RelEnt}}(\phi) = \sup_{P \neq Q} \frac{H_g[\phi(P),\phi(Q)]}{H_g[P,Q]}, \tag{64}$$

$$\eta_g^{\text{Riem}}(\phi) = \sup_P \sup_{A \in T_P \mathcal{D}} \frac{\langle \phi(A), \Omega^g_{\phi(P)}[\phi(A)] \rangle}{\langle A, \Omega^g_P[A] \rangle}, \tag{65}$$

$$\eta_g^{\text{geod}}(\phi) = \sup_{P \neq Q} \frac{[D_g(\phi(P),\phi(Q))]^2}{[D_g(P,Q)]^2}. \tag{66}$$

In Refs. 1, 2 it was shown that for commutative systems, $\eta_g^{\text{RelEnt}}(\phi) = \eta_g^{\text{Riem}}(\phi)$ $= \eta_{(w-1)^2}(\phi)$. Here, we will prove some relations between these various η.

Theorem IV.2: *The three contraction coefficients defined above satisfy*

$$1 \geq \eta_g^{\text{RelEnt}}(\phi) \geq \eta_g^{\text{Riem}}(\phi) \geq \eta_g^{\text{geod}}(\phi). \tag{67}$$

The intuition behind the second inequality can be seen by letting $A = B = Q - P$ in the integral representations of Theorems 2.2 and 2.3. Then the only difference between the ratios in (64) and (65) is that the modular operator in the former is $\Delta_{Q,P}$ while that in the latter is $\Delta_{P,P}$. This would seem to indicate that the first supremum is taken over a larger set. However, the two are not directly comparable because the condition $P \neq Q$ in the first case precludes the choice $\Delta_{P,P}$. Hence, we consider $Q = P + \epsilon A$.

Proof: The upper bound of 1 follows immediately from Theorem II.14. To prove the second inequality $\eta_g^{\text{RelEnt}}(\phi) \geq \eta_g^{\text{Riem}}(\phi)$, we consider, as suggested above, $H_g(P, P + \epsilon A)$ $= \text{Tr } P^{1/2}g(\Delta_{P,P+\epsilon A})(P^{1/2})$. Proceeding as in the proof of Theorem II.5, but with the shorthand $dN_g(s) = (b_g + c_g)\delta(s)ds + d\nu_g(s)$, we obtain

$$H_g(P, P + \epsilon A) = \epsilon^2 \int_0^\infty \text{Tr}\left[A \frac{1}{L_{P+\epsilon A} + sR_P}(A)\right] dN_g(s)$$

$$= \epsilon^2 \int_0^\infty \text{Tr}\left[A \frac{1}{L_P + sR_P}(A)\right] dN_g(s) + O(\epsilon^2)$$

$$= \epsilon^2 \langle \phi(A), \Omega^g_{\phi(P)}(A)) \rangle + O(\epsilon^3).$$

Thus

$$\eta_g^{\text{RelEnt}}(\phi) = \sup_{P \neq Q} \frac{H_g[\phi(P),\phi(Q)]}{H_g[P,Q]} \geq \sup_P \sup_{A \in T_* \mathcal{D}} \frac{H_g[\phi(P),\phi(P+\epsilon A)]}{H_g(P,P+\epsilon A)}.$$

However,

$$\frac{H_g[\phi(P),\phi(P+\epsilon A)]}{H_g(P,P+\epsilon A)} = \frac{\langle \phi(A),\Omega_{\phi(P)}^g[\phi(A)]\rangle + O(\epsilon)}{\langle A,\Omega_P^g[A]\rangle + O(\epsilon)}.$$

Since the quantity on the right can be made arbitrarily close to $\eta_g^{\mathrm{Riem}}(\phi)$, we conclude that $\eta_g^{\mathrm{RelEnt}}(\phi) \geq \eta_g^{\mathrm{Riem}}(\phi)$. Finally, to prove the third inequality we first choose $S_0(t)$ to be a minimizing path for $D_g(P,Q)$, i.e.,

$$D_g(P,Q) = \int_0^1 \sqrt{\langle \dot{S}_0(t),\Omega_{S_0(t)}^g \dot{S}_0(t)\rangle}\, dt.$$

Then, $\phi \circ S_0$ is a smooth path from $\phi(P)$ to $\phi(Q)$. Moreover, the linearity of ϕ implies that $(d/dt)\phi \circ S_0(t) = \phi \circ \dot{S}_0(t)$. Thus

$$D_g[\phi(Q),\phi(Q)] \leq \int_0^1 \sqrt{\langle \phi \circ \dot{S}_0(t),\Omega_{\phi \circ S_0(t)}^g \phi \circ \dot{S}_0(t)\rangle}\, dt$$

$$\leq [\eta_g^{\mathrm{Riem}}(\phi)]^{1/2} \int_0^1 \sqrt{\langle \dot{S}_0(t),\Omega_{S_0(t)}^g \dot{S}_0(t)\rangle}\, dt$$

$$= [\eta_g^{\mathrm{Riem}}(\phi)]^{1/2} D_g(P,Q).$$

Dividing both sides by $D_g(P,Q)$ and taking the supremum of the left-hand side, gives the desired result.   Q.E.D.

In this case of the first inequality $\eta_g^{\mathrm{RelEnt}}(\phi) \geq \eta_g^{\mathrm{Riem}}(\phi)$, we proved slightly more, namely, that either equality holds or the supremum in $\eta_g^{\mathrm{RelEnt}}(\phi)$ is actually attained for some non-negative (but not necessarily strictly positive) density matrices P,Q, i.e., strict inequality implies that there exists $P \neq Q \in \overline{\mathcal{D}}$ such that

$$H_g[\phi(P),\phi(Q)] = \eta_g^{\mathrm{RelEnt}}(\phi) H_g(P,Q). \tag{68}$$

This follows from the fact that we can always find a maximizing sequence (P_k,Q_k) such that

$$\lim_{k\to\infty} \frac{H_g[\phi(P_k),\phi(Q)_k]}{H_g(P_k,Q_k)} = \eta_g^{\mathrm{RelEnt}}(\phi).$$

Since we are in a finite-dimensional space, the space of non-negative density matrices is compact so that we can find a convergent subsequence $(P_{k_j},Q_{k_j}) \to (P,Q)$. Then either $P = Q$, in which case we necessarily have $\eta_g^{\mathrm{RelEnt}}(\phi) = \eta_g^{\mathrm{Riem}}(\phi)$, or (68) holds. (Strictly speaking, we must also exclude the possibility that both $H_g(P_k,Q_k)$ and $H_g[\phi(P_k),\phi(Q_k)]$ diverge to ∞.) We expect that for most choices of g, Eq. (68) holds only in very special cases [see, e.g., the partial trace example in Sec. IV D, which yield $\eta_g^{\mathrm{RelEnt}}(\phi) = 1 = \eta_g^{\mathrm{Riem}}(\phi)$.] Indeed, even for commutative systems, early proofs[37,1] that equality holds for $\eta_{\log}(A) = \eta_{(w-1)^2}(A)$ depended on a demonstration that (68) could not hold in general.

Another special situation occurs for the minimal g, which yields the Bures metric. If P,Q commute, then

$$H_{\mathrm{Bures}}(P,Q) \equiv \mathrm{Tr}(P-Q)([L_P+R_Q]^{-1} + [L_Q+R_P]^{-1})(P-Q)$$

$$= 2\,\mathrm{Tr}(P-Q)(P+Q)^{-1}(P-Q) = 2\langle (P-Q),\Omega_{P+Q}^{\mathrm{Bures}}[(P-Q)]\rangle.$$

Thus, if the supremum for $\eta_{\mathrm{Bures}}^{\mathrm{Riem}}(\phi)$ happens to be attained for a commuting pair R,A (with $R \in \mathcal{D}$ and $A \in T_P\mathcal{D}$) whose images $\phi(R),\phi(A)$ also commute, then

$$H_{\text{Bures}}[\phi(R+A),\phi(R-A)]=\eta_{\text{Bures}}^{\text{Riem}}(\phi)H_{\text{Bures}}(R+A,R-A). \tag{69}$$

If $\eta_{\text{Bures}}^{\text{RelEnt}}(\phi)=\eta_{\text{Bures}}^{\text{Riem}}(\phi)$, then this also yields equality in (68); however, it does *not* give a strict inequality for $\eta_{\text{Bures}}^{\text{RelEnt}}(\phi)\geqslant\eta_{\text{Bures}}^{\text{Riem}}(\phi)$. On the contrary, it seems to offer some heuristic support for equality.

We expect that in those exceptional situations in which the supremum $\eta_g^{\text{RelEnt}}(\phi)$ is attained the result is equal to $\eta_g^{\text{Riem}}(\phi)$ so that equality always holds, at least for the first inequality in Theorem IV.2.

Recall that many common choices for g [e.g., $g(w)=(w-1)^2$ or $g(w)=-\log w$] do not yield a symmetric relative entropy, i.e., $H_g(P,Q)\neq H_g(Q,P)$. This raises the question of whether or not the entropy contraction coefficient [which we denote $\eta_g^{\text{sym}}(\phi)\equiv\eta_{g(w)+wg(w^{-1})}^{\text{RelEnt}}(\phi)$] for the symmetrized relative entropy,

$$H_g^{\text{sym}}(P,Q)=H_g(P,Q)+H_g(Q,P)=H_{g(w)+wg(w^{-1})}(P,Q), \tag{70}$$

is the same as $\eta_g^{\text{RelEnt}}(\phi)$. Although we believe equality holds, we can only prove that

$$\eta_g^{\text{sym}}(\phi)\leqslant\eta_g^{\text{RelEnt}}(\phi). \tag{71}$$

Nevertheless, Theorem IV.2 holds for any g. In fact, since there is a unique Riemannian metric associated with all g, which yield the same symmetrized relative entropy, we have $\eta_g^{\text{RelEnt}}(\phi)\geqslant\eta_g^{\text{sym}}(\phi)\geqslant\eta_g^{\text{Riem}}(\phi)$. To prove (71) it suffices to observe that

$$H_{g(w)+wg(w^{-1})}(P,Q)=H_g^{\text{sym}}(P,Q)=H_g(P,Q)+H_g(Q,P),$$

so that

$$\begin{aligned}
H_g^{\text{sym}}[\phi(P),\phi(Q)]&=H_g[\phi(P),\phi(Q)]+H_g[\phi(Q),\phi(P)]\\
&\leqslant\eta_g^{\text{RelEnt}}(\phi)H_g(P,Q)+\eta_g^{\text{RelEnt}}(\phi)H_g(Q,P)\\
&=\eta_g^{\text{RelEnt}}(\phi)H_g^{\text{sym}}(P,Q).
\end{aligned}$$

In the case of the quadratic entropy, it easily follows that $\eta_{(w-1)^2}^{\text{Riem}}(\phi)=\eta_{(w-1)^2}^{\text{RelEnt}}(\phi)=\eta_{(w-1)^2}^{\text{sym}}(\phi)$.

Finally, we note that the joint convexity of relative entropy, Riemannian metrics, and $[D_g(P,Q)]^2$ imply that the corresponding contraction coefficients are convex in ϕ. [Although we did not explicitly state the joint convexity for $M_P(A,A)$ it is an easy consequence of homogeneity and contraction under partial traces.]

Theorem IV.3: *For each fixed $g\in\mathcal{G}$, each of the contraction coefficients $\eta_g^{\text{RelEnt}}(\phi)$, $\eta_g^{\text{Riem}}(\phi)$, and $\eta_g^{\text{geod}(\phi)}$ is convex in ϕ.*

Proof: Since the argument is straightforward, we give details only for the relative entropy. Let $\phi=x\phi_1+(1-x)\phi_2$:

$$\begin{aligned}
H_g[\phi(P),\phi(Q)]&=H_g[x\phi_1(P)+(1-x)\phi_2(P),x\phi_1(Q)+(1-x)\phi_2(Q)]\\
&\leqslant xH_g[\phi_1(P),\phi_1(Q)]+(1-x)H_g[\phi_2(P),\phi_2(Q)]\\
&\leqslant x\eta_g^{\text{RelEnt}}(\phi_1)H_g(P,Q)+(1-x)\eta_g^{\text{RelEnt}}(\phi_2)H_g(P,Q)\\
&=[x\eta_g^{\text{RelEnt}}(\phi_1)+(1-x)\eta_g^{\text{RelEnt}}(\phi_2)]H_g(P,Q).
\end{aligned}$$

Dividing both sides by $H_g(P,Q)$ implies

$$\eta_g^{\text{RelEnt}}(\phi)\leqslant x\eta_g^{\text{RelEnt}}(\phi_1)+(1-x)\eta_g^{\text{RelEnt}}(\phi_2). \qquad\text{Q.E.D.}$$

B. Eigenvalue formulation of $\eta_g^{\text{Riem}}(\phi)$

We now show how $\eta_g^{\text{Riem}}(\phi)$ is related to the following set of eigenvalue problems:

$$[\hat{\phi}\circ\Omega_{\phi(P)}^g\circ\phi](A)=\lambda\,\Omega_P^g(A). \tag{72}$$

In view of Theorem II.10, this is a well-defined linear eigenvalue problem on $\mathbf{C}^{n\times n}$ for each fixed pair ϕ and P. The following remarks are easily verified.

(a) The eigenvalue problem (72) can be rewritten as $\Phi_P^g\circ\phi(B)=\lambda B$, where

$$\Phi_P^g\equiv(\Omega_P^g)^{-1}\circ\hat{\phi}\circ\Omega_{\phi(P)}^g. \tag{73}$$

Furthermore, Φ_P^g is trace preserving. This follows from Theorem II.12 and

$$\begin{aligned}
\operatorname{Tr}\Phi_P^g(B)=\operatorname{Tr}P\,\hat{\phi}\circ\Omega_{\phi(P)}^g(B) &=\langle P,\hat{\phi}\circ\Omega_{\phi(P)}^g(B)\rangle \\
&=\langle\phi(P),\Omega_{\phi(P)}^g(B)\rangle=\langle\Omega_{\phi(P)}^g[\phi(P)],B\rangle \\
&=\langle I,B\rangle=\operatorname{Tr}B.
\end{aligned}$$

(b) We can assume without loss of generality that the matrices that are eigenvectors in (72) are self-adjoint, i.e., that $A=A^*$. Indeed, it is easy to check that the operator $\Omega_P^{g_{s_0}}(A)=(sR_P+L_P)$ $\times[R_P+L_P]^{-1}(R_P+sL_P)(A)$ satisfies $[\Omega_P^{g_{s_0}}(A)]^*=\Omega_P^{g_{s_0}}(A^*)$. Therefore, the operators Ω_P^g, $\Omega_{\phi(P)}^g$, ϕ, $\hat{\phi}$ and Φ_P^g all map adjoints to adjoints.

(c) For each fixed P, the eigenvalue equation is satisfied with $A=P$ and eigenvalue $\lambda=1$, which is the largest eigenvalue. The operators on both sides of (72) are self-adjoint (in fact, positive definite) with respect to the Hilbert–Schmidt inner product and the corresponding orthogonality condition for the other eigenvectors reduces to $\operatorname{Tr}A=0$.

In view of these observations, it is easy to conclude from the max–min principle that the second-largest eigenvalue $\lambda_2^g(\phi,P)$ satisfies

$$\lambda_2^g(\phi,P)=\sup_{A\in T_P\mathcal{D}}\frac{\langle\phi(A),\Omega_{\phi(P)}^g[\phi(A)]\rangle}{\langle A,\Omega_P^g[A]\rangle}, \tag{74}$$

for each fixed P. Then taking the supremum over $\mathcal{D}$ yields

Theorem IV.4: *For each $g\in\mathcal{G}$ and stochastic map ϕ,*

$$\eta_g^{\text{Riem}}(\phi)=\sup_{P\in\mathcal{D}}\lambda_2^g(\phi,P). \tag{75}$$

We have already observed that every Ω_P^g can be regarded as a noncommutative variant of multiplication by P^{-1}. Indeed, if both pairs of operators P,A and $\phi(P),\phi(A)$ associated with a particular eigenvalue commute for some g, then $\Omega_P^g(A)=R_{P^{-1}}(A)=L_{P^{-1}}(A)$ for *all* g and the corresponding eigenvalue equations are the same. It may be tempting to conjecture that the eigenvalue equations for different g are related by a similarity transform, which would then imply that all $\lambda_2(\phi,P)$ are equal so that all $\eta_g^{\text{Riem}}(\phi)$ are identical. However, for a given fixed P, R_P, and L_P commute, which implies that Ω_P^g and Ω_P^h commute for any pair of functions g and h. Since commuting operators are simultaneously diagonalizable and similar operators have the same eigenvalues, this would imply that all of the eigenvalue operators $B\rightarrow[(\Omega_P^g)^{-1}\circ\hat{\phi}\circ\Omega_{\phi(P)}^g\circ\phi](B)$ are identical. This is easily seen to be false in specific examples. Moreover, as discussed at the end of Sec. IV D, one can find examples of nonunital ϕ for which different $\eta_g^{Riem}(\phi)$ are *not* identical.

Theorem IV.5: *We can rewrite the eigenvalue problem (72) so that*

$$\lambda_2^g(\phi,P) = \sup_\alpha \frac{\langle \hat{\phi}(\alpha), (\Omega_P^g)^{-1}[\hat{\phi}(\alpha)]\rangle}{\langle \alpha, (\Omega_{\phi(P)}^g)^{-1}[\alpha]\rangle},$$

where the supremum is now taken over $\{\alpha \in \mathrm{Range}(\phi) : \mathrm{Tr}[\Omega_{\phi(P)}^g]^{-1}(\alpha)=0\}$.

Proof:

$$\lambda_2^g(\phi,P) = \sup_{A:\mathrm{Tr}(A)=0} \frac{\langle \phi(A), \Omega_{\phi(P)}^g[\phi(A)]\rangle}{\langle A, \Omega_P^g[A]\rangle}$$

$$= \sup_{B:\mathrm{Tr}[\Omega_P^g]^{-1/2}(B)=0} \frac{\langle B[\Omega_P^g]^{-1/2}\circ\hat{\phi}\circ\Omega_{\phi(P)}^g\circ\phi\circ[\Omega_P^g]^{-1/2}B]\rangle}{\langle B,B\rangle}.$$

If we now write $\Gamma = [\Omega_{\phi(P)}^g]^{1/2}\circ\phi\circ[\Omega_P^g]^{-1/2}$, we see that $\lambda_2^g(\phi,P)$ is the largest eigenvalue of $\Gamma^*\Gamma$, where Γ maps

$$\{B:\mathrm{Tr}[\Omega_P^g]^{-1/2}(B)=0\} \to \{\beta \in \mathrm{Range}(\phi):\mathrm{Tr}[\Omega_{\phi(P)}^g]^{-1/2}(\beta)=0\}.$$

Since $\Gamma\Gamma^*$ and $\Gamma^*\Gamma$ have the same nonzero eigenvalues,

$$\lambda_2^g(\phi,P) = \sup_{\beta:\mathrm{Tr}[\Omega_{\phi(P)}^g]^{-1/2}(\beta)=0} \frac{\langle \beta[\Omega_{\phi(P)}^g]^{1/2}\circ\phi\circ[\Omega_P^g]^{-1}\circ\hat{\phi}\circ[\Omega_{\phi(P)}^g]^{1/2}]\beta\rangle}{\langle \beta,\beta\rangle}$$

$$= \sup_{\alpha:\mathrm{Tr}[\Omega_{\phi(P)}^g]^{-1}(\alpha)=0} \frac{\langle \hat{\phi}(\alpha)[\Omega_P^g]^{-1}\hat{\phi}(\alpha)\rangle}{\langle \alpha,[\Omega_{\phi(P)}^g]^{-1}\alpha\rangle}.$$

If we apply this result with $\Omega_P^{\mathrm{Bures}}=[R_P+L_p]^{-1}$, it follows easily from the theorem above that

$$\lambda_2^{\mathrm{Bures}}(\phi,P) = \sup_{\alpha:\mathrm{Tr}\,\phi(P)\alpha=0} \frac{\mathrm{Tr}\,\phi(\alpha)P\hat{\phi}(\alpha)}{\mathrm{Tr}\,\alpha\phi(P)\alpha}. \tag{76}$$

It is tempting to write $\phi(P)\alpha=\beta=\phi(B)$ and replace the constraint $\mathrm{Tr}\,\phi(P)\alpha=0$ by $\mathrm{Tr}\,B=0$. The denominator would then become $\langle \phi(B)[\phi(P)]^{-1}\phi(B)\rangle$, which has the same form as the numerator in (65) when $k(w)=(w+1)/2w$ [corresponding to $g=(w-1)^2$]. However, there is no guarantee that $\phi(\alpha)=\hat{\phi}([\phi(P)]^{-1}]B)$. On the contrary, this cannot possibly hold because we would then have that the λ (and hence η) for the two extremal functions $k(w)=2/(1+w)$ and $k(w)=(w+1)/2w$ are inverses, which is inconsistent with $\lambda^g(\phi,P) \leqslant \eta_g^{\mathrm{Riem}}(\phi) \leqslant 1$ (except in the case $\lambda=1$, which is not generic). There is, however, a sense in which the operators associated with these two extremal functions are inverses since $\Omega_P^{(w-1)^2}=R_P^{-1}+L_P^{-1}=R_P^{-1}[R_P+L_P]L_P^{-1}$ $=R_P^{-1}[\Omega_P^{\mathrm{Bures}}]^{-1}L_P^{-1}$. It seems likely that if the η_g^{Riem} for these two extremal functions are equal, then all of them are.

Unlike the case of $\eta_g^{\mathrm{RelEnt}}(\phi)$, we do expect that the supremum for $\eta_g^{\mathrm{Riem}}(\phi)$ is actually attained. Indeed, we know that for each fixed P the supremum in (74) is attained for some $A \neq 0$, which satisfies the eigenvalue problem (72). As before, we can find a maximizing sequence of density matrices P_k for (75) so that $\eta_g^{\mathrm{Riem}}(\phi)=\lim_{k\to\infty}\lambda_2^g(\phi,P_k)$. For each P_k, let A_k be the solution to the eigenvalue problem (72) for $\lambda_2^g(\phi,P_k)$ normalized so that $\mathrm{Tr}|A_k|=1$. Then we can find a convergent subsequence for which $P_k\to P\in\mathcal{D}$ and $A_k\to A\neq 0$, since $\mathrm{Tr}|A|\neq 0$. It then follows that (72) holds for this P,A with $\lambda=\eta_g^{\mathrm{Riem}}(\phi)$ (although P is only non-negative), which implies that

$$\langle \phi(A), \Omega^g_{\phi(P)}[\phi(A)] \rangle = \eta^{\text{Riem}}_g(\phi) \langle A, \Omega^g_P(A) \rangle.$$

C. Bounds on contraction coefficients

We first give an upper bound for $\eta^{\text{Riem}}_{\log}$ using

$$\eta^{\text{Dobrushin}}(\phi) \equiv \sup_{A \in T_*\mathcal{D}} \frac{\text{Tr}|\phi(A)|}{\text{Tr}|A|}. \tag{77}$$

This can be interpreted as the norm of ϕ regarded as an operator on the Banach space of traceless matrices with norm $\text{Tr}|A|$. Although the function $g(w) = |w-1|$ is *not* operator convex, $\eta^{\text{Dobrushin}}(\phi)$ is analogous to the contraction coefficient of the (nondifferentiable) symmetric relative g entropy $H_{|w-1|}(P,Q) = \text{Tr}|P-Q|$, which, however, is *not* the relative g entropy obtained by using $g(w) = |w-1|$ in Definition II.1. Nevertheless, $\eta^{\text{Dobrushin}}(\phi)$ is a natural and useful object to consider. It was shown in Ref. 3 (see Theorem 2) that

$$\eta^{\text{Dobrushin}}(\phi) = \tfrac{1}{2} \sup\{\text{Tr}|\phi(E-F)| : E,F \ 1\text{-dim proj s}; EF=0\}, \tag{78}$$

where ''1-dim projs'' means that E,F are one-dimensional projections in $\bar{\mathcal{D}}$. The expression on the right in (78) shows that we are justified in interpreting $\eta^{\text{Dobrushin}}(\phi)$ as a noncommutative analog of Dobrushin's coefficient of ergodicity.

Theorem IV.6: *If ϕ is stochastic,*

$$\eta^{\text{Riem}}_{\log}(\phi) \leq \eta^{\text{Dobrushin}}(\phi) \equiv \sup_{A \in T_*\mathcal{D}} \frac{\text{Tr}|\phi(A)|}{\text{Tr}|A|}. \tag{79}$$

Proof: The map $B \to (\Omega^{\log}_P)^{-1} \circ \hat{\phi} \circ \Omega^{\log}_{\phi(P)}(B) \equiv \Phi_{\log}(B)$ is positivity preserving, as well as trace preserving. The former follows from the integral representations (49) and (51) for $\Omega^{\log}_P$ and its inverse, together with the fact that the composition of positivity-preserving maps is positive preserving. Then, taking the trace of the absolute value of both sides of the eigenvalue problem $\Phi[\phi(A)] = \lambda A$ and using Theorem 1 of Ref. 3 yields

$$\lambda \, \text{Tr}|A| = \text{Tr}|\Phi[\phi(A)]| \leq \text{Tr}|\phi(A)|. \quad \text{Q.E.D.} \tag{80}$$

Although we believe that this result holds for any g, we do not have a proof except for the log. Our proof depended on the observation that in the case of the log the map $\Phi_g(B) = (\Omega^g_P)^{-1} \circ \hat{\phi} \circ \Omega^g_{\phi(P)}(B)$ is positivity preserving. However, explicit examples can be found to show that Φ_g is *not* positivity preserving in general. Indeed, although both $\Omega^{\text{Bures}}_P = [R_P + L_P]^{-1}$ and $\Omega^{(w-1)^2}_P = R_P^{-1} + L_P^{-1}$ are positive semidefinite with respect to the Hilbert–Schmidt inner product, they are not positivity preserving in the sense of mapping positive operators to positive operators. The difference is analogous to the difference between an ordinary matrix being positive semidefinite and having positive elements.

We now consider lower bounds on $\eta^{\text{Riem}}_g(\phi)$. In Ref. 3 it was shown that

$$\eta^{\text{Dobrushin}}(\phi) \leq \sqrt{\eta^{\text{Riem}}_{(w-1)^2}(\phi)}. \tag{81}$$

We now give a lower bound that holds for all $\eta^{\text{Riem}}_g(\phi)$ when the map ϕ is unital, i.e., $\phi(I) = I$.

Theorem IV.7: *If ϕ is unital,*

$$\eta^{\text{Riem}}_g(\phi) \geq \sup_{\text{Tr}\,A=0} \frac{\text{Tr}|\phi(A)|^2}{\text{Tr}|A|^2}. \tag{82}$$

This is an immediate consequence of the definition (65); it also follows from Theorem IV.4 and the fact that the right side of (82) is just $\lambda_2(\phi,I)$ when ϕ is unital. The right side of (82) can also be interpreted as the square of the norm of ϕ regarded as an operator on the Banach space of traceless matrices with Hilbert–Schmidt norm $\sqrt{\operatorname{Tr}A^*A}$. When ϕ is self-adjoint in the sense $\hat{\phi} = \phi$, every trace-preserving map is unital.

If ϕ maps $\mathbf{C}^{n\times n}$ to itself, then the results of this section can be restated in terms of the eigenvalues and singular values of ϕ. Since ϕ is trace preserving, $\phi(B)=\Lambda B$ implies that either $\Lambda = 1$ or $\operatorname{Tr}B=0$. If we restrict ϕ to the matrices with trace zero, then $\eta^{\text{Dobrushin}}(\phi)$ is the largest magnitude of an eigenvalue and for unital ϕ, $\lambda_2(\phi,I)$ is the largest eigenvalue of $\hat{\phi}\phi$. Thus, for unital stochastic maps, $\lambda_2(\phi,I)=\Lambda_2(\hat{\phi}\phi)$, where we have continued our convention of using the subscript 2 for eigenvalues of maps restricted to $T_*\bar{\mathcal{D}}$. If ϕ is self-adjoint, the two lower bounds (81) and (82) coincide and $\lambda_2(\phi,I)=\Lambda_2(\hat{\phi}\phi)=[\Lambda_2(\phi)]^2$ in the usual sense of second largest eigenvalue of. For general unital ϕ, (82) is stronger since

$$\eta^{\text{Riem}}_{(w-1)^2}(\phi)\geqslant\lambda_2(\phi,I)=\Lambda_2(\hat{\phi}\phi)\geqslant[\eta^{\text{Dobrushin}}(\phi)]^2. \tag{83}$$

We now explicitly state some conjectures that have already been discussed.

Conjecture IV.8: For each fixed $g\in\mathcal{G}$,

$$\eta^{\text{RelEnt}}_g(\phi)=\eta^{\text{Riem}}_g(\phi)=\eta^{\text{geod}}_g(\phi)\leqslant\eta^{\text{Dobrushin}}_g(\phi). \tag{84}$$

Conjecture IV.9: If ϕ is unital, then

$$\eta^{\text{Riem}}_g=\Lambda_2(\hat{\phi}\phi)\equiv\sup_{\operatorname{Tr}A=0}\frac{\operatorname{Tr}|\phi(A)|^2}{\operatorname{Tr}|A|^2}, \tag{85}$$

for all $g\in\mathcal{G}$.

If this conjecture holds, then for unital ϕ the contraction coefficient η^{Riem}_g is independent of g. Theorem IV.13 at the end of the next section contains an explicit example of a nonunital stochastic map for which η^{Riem}_g depends nontrivially on g; therefore, the hypothesis that ϕ be unital is essential. In view of (82) it would suffice to show that $\eta^{\text{Riem}}_g\leqslant\Lambda_2(\hat{\phi}\phi)$.

D. Examples

We now consider some special classes of stochastic maps $\phi:\mathcal{A}_1\to\mathcal{A}_2$. We begin by looking at some maps for which all contraction coefficients are easily seen to be zero or one. We then consider maps from $\mathbf{C}^{2\times2}$ to $\mathbf{C}^{2\times2}$ that provide support for the conjectures above.

We first consider the case in which $\mathcal{A}_2$ is one dimensional, e.g., ϕ projects onto a one-dimensional subalgebra (which need not have an identity) of $\mathcal{A}_1$. Then, since ϕ is trace preserving and maps density matrices to density matrices, we must have $\phi(P)=\phi(Q)$, $\forall P,Q$ with $\operatorname{Tr}\varphi(P)=1$ so that $\phi(P)\neq0$. Thus, $H_g[\phi(P),\phi(Q)]=D_g[\phi(P),\phi(Q)]=0$, $\forall P,Q$, which implies $\eta^{\text{RelEnt}}_g(\phi)=\eta^{\text{geod}}_g(\phi)=0$. If $\operatorname{Tr}B=0$, then $\phi(B)=0$. [To see this note that one can find a,b such that $P=(aI+bB)$ is a density matrix.] Thus, $\langle\phi(B)\Omega^g_{\phi(P)}\phi(B)\rangle=0$ and $\operatorname{Tr}|\phi(B)|=0$ for all B in $T_*\bar{\mathcal{D}}$, which implies that $\eta^{\text{Riem}}_g(\phi)=\eta^{\text{Dobrushin}}(\phi)=0$. We can summarize this as the following.

Theorem IV.10: *If the image of the stochastic map ϕ is one dimensional, then $\eta^{\text{RelEnt}}_g(\phi)=\eta^{\text{Riem}}_g(\phi)=\eta^{\text{geod}}_g(\phi)=\eta^{\text{Dobrushin}}(\phi)=0$ for all $g\in\mathcal{G}$.*

We next consider the important special case in which ϕ is a partial trace τ. In the simplest case, let $\tau:\mathbf{C}^{2n\times2n}\to\mathbf{C}^{n\times n}$ be the map that takes

$$\mathbf{M}=\begin{pmatrix}A & B\\ C & D\end{pmatrix}\to\tau(\mathbf{M})=A+D, \tag{86}$$

where $\mathbf{M} \in \mathbf{C}^{2n \times 2n}$ has been written in block form and $A, B, C, D \in \mathbf{C}^{n \times n}$. Then the homogeneity of relative entropy (see Definition II.3b) implies that for

$$\mathbf{P} = \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix} \quad \text{and} \quad \mathbf{Q} = \begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix},$$

$$H_g(\mathbf{P}, \mathbf{Q}) = H_g(2P, 2Q) = H_g(\tau(\mathbf{P}), \tau(\mathbf{Q})),$$

for any g, and, similarly,

$$\langle \mathbf{A}, \Omega_P^g(\mathbf{A}) \rangle = \langle 2A, \Omega_{2P}^g(2A) \rangle = \langle \tau(\mathbf{A}), \Omega_{\tau(P)}^g(\tau(\mathbf{A})) \rangle,$$

when

$$\mathbf{A} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}.$$

From this, we easily see that

$$\eta_g^{\mathrm{RelEnt}}(\phi) = \eta_g^{\mathrm{Riem}}(\phi) = \eta_g^{\mathrm{geod}}(\phi) = \eta^{\mathrm{Dobrushin}}(\phi) = 1, \tag{87}$$

where we have assumed implicitly that τ acts on the full algebra of all $2n \times 2n$ matrices.

The partial trace described above is similar to a conditional expectation, i.e., a map for which $\mathcal{A}_2$ is a subalgebra (with identity) of $\mathcal{A}_1$ and $\phi(A) = A$, $\forall A \in \mathcal{A}_2$. Both partial traces and conditional expectations are included in the following.

Theorem IV.11: *If the stochastic map ϕ is also an isomorphism from a nontrivial subalgebra (with identity) of $\mathcal{A}_1$ to $\mathcal{A}_2$, then $\eta_g^{\mathrm{RelEnt}}(\phi) = \eta_g^{\mathrm{Riem}}(\phi) = \eta_g^{\mathrm{geod}}(\phi) = \eta^{\mathrm{Dobrushin}}(\phi) = 1$ for all $g \in \mathcal{G}$.*

Since every completely positive map can be represented as a partial trace,[21] this might seem to suggest that $\eta = 1$ always holds. However, these representations involve multiple copies of the algebra, so that the partial trace is not acting on the full algebra in the higher-dimensional space. Thus, the representation of $\mathcal{A}_1$ need necessarily not contain a subalgebra with the desired isomorphism property. Examples of maps with $\eta < 1$ were already found in Ref. 1 for commutative algebras, and two different noncommutative examples are given below.

We now state two results for maps $\phi : \mathbf{C}^{2 \times 2} \to \mathbf{C}^{2 \times 2}$. The proofs are postponed to a subsequent paper.[38] Recall that any density matrix in $\mathbf{C}^{2 \times 2}$ can be written in the form $\frac{1}{2}[I + \mathbf{w} \cdot \sigma]$, where $\mathbf{w} \in \mathbf{R}^3$ and σ denotes the vector of Pauli matrices. The first theorem provides evidence for the two conjectures at the end of the previous section.

Theorem IV.12: *For the unital map $\phi_T : I + \mathbf{w} \cdot \sigma \to I + T\mathbf{w} \cdot \sigma$,*

$$\eta_g^{\mathrm{RelEnt}}(\phi_T) = \eta_g^{\mathrm{Riem}}(\phi_T) = \eta_g^{\mathrm{geod}}(\phi_T) = \|T\|^2, \quad \forall g \in \mathcal{G},$$

and $\eta^{\mathrm{Dobrushin}}(\phi_T) = \|T\|$.

The next example gives a nonunital stochastic map for which $\eta_g^{\mathrm{Riem}}(\phi)$ varies with g. For $\alpha, \tau > 0$ with $\alpha + \tau \leq 1$, define

$$\phi_{\alpha, \tau}[I + \mathbf{w} \cdot \sigma] = I + \alpha w_1 \sigma_1 + \tau \sigma_2. \tag{88}$$

It is easily seen to be stochastic because the condition $\alpha + \tau \leq 1$ ensures that it is a convex combination of stochastic maps. For $g_{s_0}(w) = (w - 1)^2 / (w + s_0)$ as in Example 3 of Sec. II E,

$$\eta_{g_{s_0}}^{\text{Riem}}(\phi) = \sup_{0 \leq \omega \leq 1} \frac{[(1-\tau^2+(\rho-\alpha^2)\omega^2][1-\omega^2]}{[1-\tau^2-\alpha^2\omega^2][1-\tau^2(1-\rho)-(1-\rho)\alpha^2\omega^2]}$$

$$\geq \frac{\alpha^2}{1-\left(\frac{1-s_0}{1+s_0}\right)^2 \tau^2},$$

where $1-\rho=(1-s_0)/(1+s_0)$ and equality holds for $s_0 \approx 0$. In particular, we can conclude the following.

Theorem IV.13: *For the nonunital stochastic map ϕ given by (88), there is an $S>0$ such that for $s_0 \in [0,S)$,*

$$\eta_{s_0}^{\text{Riem}}(\phi) = \frac{\alpha^2}{1-\left(\frac{1-s_0}{1+s_0}\right)^2 \tau^2}.$$

Furthermore,

$$\eta_{(w-1)^2}^{\text{Riem}}(\phi) = \frac{\alpha^2}{1-\tau^2} < \alpha = \eta^{\text{Dobrushin}}(\phi).$$

If $s_1 \in (0,S)$, we have $\eta_{s_1}^{\text{Riem}}(\phi) > \eta_{s_0}^{\text{Riem}}(\phi) = \alpha^2/1-\tau^2$.

ACKNOWLEDGMENTS

[1] J. E. Cohen, Y. Iwasa, Gh. Rautu, M. B. Ruskai, E. Seneta, and Gh. Zbaganu, ''Relative entropy under mappings by stochastic matrices,'' Linear Algebr. Appl. **179**, 211–235 (1993).

[2] M. D. Choi, M. B. Ruskai, and E. Seneta, ''Equivalence of certain entropy contraction coefficients,'' Linear Algebr. Appl. **208/209**, 29–36 (1993).

[3] M. B. Ruskai, ''Beyond strong subadditivity? Improved bounds on the contraction of generalized relative entropy,'' Rev. Math. Phys. **6**, 1147–1161 (1994).

[4] L. L. Campbell, ''An extended Čencov characterization of the information metric,'' Proc. Am. Math. Soc. **98**, 135–141 (1986).

[5] N. N. Čencov, *Statistical Decision Rules and Optimal Inferences*, Transl. Math. Monogr. **53** (American Mathematical Society, Providence, 1982).

[6] E. A. Morozova and N. N. Čencov, ''Markov invariant geometry on state manifolds,'' Itogi Nauk Tekhn **36**, 69–102 (1990) (in Russian).

[7] D. Petz, ''Information geometry on quantum states,'' Hungarian Academy of Sciences preprint No. 39, 1995, in *Quantum Probability Communications*, edited by R. L. Hudson *et al.* (World Scientific, Singapore, 1998.), Vol. 10, pp. 135–158.

[8] D. Petz, ''Monotone metrics on matrix spaces,'' Linear Algebr. Appl. **244**, 81–96 (1996).

[9] D. Petz and Cs. Sudar, ''Geometries of quantum states,'' J. Math. Phys. **37**, 2662–2673 (1996).

[10] D. Petz and Cs. Sudar, ''Extending the Fisher metric to density matrices,'' preprint.

[11] D. Petz and M. B. Ruskai, ''Contraction of generalized relative entropy under stochastic mappings,'' Quantum Probab. Infinite Dimensional Anal. **1**, 83–89 (1998).

[12] H. Araki, ''Relative entropy of state of von Neumann algebras,'' Publ. RIMS Kyoto Univ. **9**, 809–833 (1976).

[13] H. Araki, ''Recent progress in entropy and relative entropy,'' *Proceedings of the VIIIth International Congress on Mathematical Physics*, edited by R. Seneor and M. Mebkhout (World Scientific, Singapore, 1987), pp. 354–365.

[14] O. Bratteli and D. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer-Verlag, Berlin, 1981).

[15] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer-Verlag, Berlin, 1993).

[16] D. Petz, ''Quasi-entropies for finite quantum systems'' Rep. Math. Phys. **23**, 57–65 (1986).

[17] W. F. Stinespring, ''Positive functions on C^*-algebras,'' Proc. Am. Math. Soc. **6**, 211–216 (1955).

[18] W. B. Arveson, ''Subalgebras of C^*-algebras,'' Acta Math. **123**, 141–224 (1969).

[19] M.-D. Choi, ''Completely positive linear maps on complex matrices,'' Linear Algebr. Appl. **10**, 285–290 (1975).

[20] K. Kraus, *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Springer-Verlag, Berlin, 1983).

[21] G. Lindblad, ''Completely positive maps and entropy inequalities,'' Commun. Math. Phys. **40**, 147–151 (1975).

[22] E. Lieb and M. B. Ruskai, ''Some operator inequalities of the Schwarz type,'' Adv. Math. **12**, 269–273 (1974).

[23] J. E. Cohen and Gh. Zbaganu, *Comparison of Stochastic Matrices* (Birkhauser, Basel, 1998).

[24] L. K. Jones and C. L. Byrne, ''General entropy criteria for inverse problems, with applications to data compression, pattern classification and cluster analysis,'' IEEE Trans. Inf. Theory **36**, 23–30 (1990).

[25] V. Vedral and M. B. Plenio, ''Entanglement measures and purification procedures,'' Phys. Rev. A **57**, 1619–1633 (1998).

[26] F. Kubo and T. Ando, ''Means of positive linear operators,'' Math. Ann. **246**, 205–224 (1980).

[27] E. Lieb, ''Convex trace functions and the Wigner–Yanase–Dyson conjecture,'' Adv. Math. **11**, 267–288 (1973).

[28] E. Lieb and M. B. Ruskai, ''Proof of the strong subadditivity of quantum mechanical entropy,'' J. Math. Phys. **14**, 1938–1941 (1973).

[29] A. Wehrl, ''General properties of entropy,'' Rev. Mod. Phys. **50**, 221–260 (1978).

[30] A. Uhlmann, ''Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory,'' Rev. Math. Phys. **54**, 21–32 (1977).

[31] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 1.

[32] J. Dittman and A. Uhlmann, ''Connections and matrics respecting standard purification'' preprint.

[33] A. Uhlmann, ''The metric of Bures and the geometric phase,'' in *Groups and Related Topics*, edited by R. Gielerak *et al.* (Kluwer Academic, City, 1992).

[34] A. Uhlmann, ''Density operators as an arena for differential geometry,'' Rep. Math. Phys. **33**, 253–263 (1993).

[35] A. Uhlmann, ''Spheres and hemispheres as quantum state spaces,'' Erwin Schrödinger Institute preprint No. 102, Vienna, 1994.

[36] A. Uhlmann, ''Geometric phases and related structures,'' Rep. Math. Phys. **36**, 461–481 (1995).

[37] R. Ahlswede and P. Gacs, ''Spreading of sets in product spaces and hypercontraction of the Markov operator,'' Ann. Prob. **4**, 925–939 (1976).

[38] M. B. Ruskai, ''Contraction of Riemannian metrics and relative entropy on the algebra of 2×2 matrices,'' preprint.