Quantum Information and Computation, Vol. 11, No. 11&12 (2011) 1028–1044

LOCAL EXTREMA OF ENTROPY FUNCTIONS UNDER TENSOR PRODUCTS

SHMUEL FRIEDLAND
Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago
851 S. Morgan Street, Chicago, IL 60607-7045, USA
friedlan@uic.edu

GILAD GOUR
Institute for Quantum Information Science and Department of Mathematics and Statistics
University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4
gour@ucalgary.ca

AIDAN ROY
Institute for Quantum Computing, University of Waterloo
200 University Avenue West, Waterloo, Ontario, Canada N2L 3G
aproy@uwaterloo.ca

Received May 24, 2011
Revised September 9, 2011

We show that under a certain condition of local commutativity the minimum von-Neumann entropy output of a quantum channel is locally additive. We also show that local minima of the 2-norm entropy functions are closed under tensor products if one of the subspaces has dimension 2.

Keywords: Quantum channels, entropy functions, local additivity
Communicated by: S Braunstein & A Harrow

Let K be a subspace of the $m \times n$ complex matrices, and let $x \in K$, $\mathrm{Tr}[xx^*] = 1$. Then the von Neumann entropy of x is

$$\mathrm{H}(x) := -\,\mathrm{Tr}[xx^* \ln xx^*],$$

and the minimum entropy output of the subspace K is

$$\mathrm{H}(K) := \min_{x \in K, \mathrm{Tr}[xx^*]=1} \mathrm{H}(xx^*).$$

Recently, Hastings [1] disproved the famous additivity conjecture, which posited that

$$\mathrm{H}(K_1 \otimes K_2) = \mathrm{H}(K_1) + \mathrm{H}(K_2). \tag{1}$$

This conjecture was considered one of the most significant open problems in quantum information theory, spawning a large literature [2]. Its importance was motivated in part by the problem of finding the classical capacity of a quantum channel, and in part by a result of Peter

Shor [3] that showed that a number of apparently distinct additivity conjectures, including the additivity of the minimum entropy output of a quantum channel, the additivity of the entanglement of formation, and the additivity of the Holevo capacity, were all equivalent.

Hastings' counterexample showed that the von Neumann entropy function is not globally additive on subspaces: in other words, if x_1 is a global minimum in K_1 and x_2 is a global minimum in K_2, then $x_1 \otimes x_2$ is not necessarily a global minimum in $K_1 \otimes K_2$. On the other hand, in this paper we show that under certain conditions the von Neumann entropy is locally additive. More precisely, we show that if K_i is a subspace with a local minimum x_i, and $x_i x_i^*$ commutes with $x_i y_i^*$ for every $y_i \in K_i$, then $x_1 \otimes x_2$ is a local minimum of $K_1 \otimes K_2$; we call this condition the local commutativity condition. More generally, we study the behaviour of entropy functions of the eigenvalues of xx^*, and we consider when the tensor product of two local minima is again a local minimum.

The paper is organized as follows. In Section 1 we analyze the local commutativity condition. In Section 2, we consider the first derivative of the entropy function and note that critical points of the von Neumann and Renyi entropies are closed under tensor products. These results are due to a group participating in the American Institute for Mathematics workshop on "Geometry and representation theory"[4]. In Section 3, we consider the second derivative of the von Neumann entropy function, and show that local minima of von Neumann entropy are closed under tensor products, given the previously mentioned commutativity assumption. Finally, in Section 4, we consider the second derivative of the 2-norm entropy function. We show that local minima of the 2-norm are closed under tensor products if one of the subspaces has dimension 2. In the Appendix A we analyze the affine parametrization and use it to derive a necessary condition for local minima. In Appendix B we show that there is a simple counter example for the additivity conjecture over the real numbers.

1 The local commutativity condition

For a given function $f : [0, \infty) \to (-\infty, \infty)$ we define $f(x) = \sum_{i=1}^{m} f(\lambda_i(xx^*))$ for $x \in C^{m \times n}$, and λ_i are the eigenvalues of xx^*. We assume that either f is smooth on $[0, \infty)$, i.e. has two continuous derivatives at every $t \geq 0$, or $f(t) = H(t) \equiv -t \log t$. Let $D_y f(x), D_y^2 f(y)$ denote the first and the second derivative of f in the y direction:

$$D_y f(x) = \frac{d}{d\epsilon} f(x + \epsilon y)\big|_{\epsilon=0}, \quad D_y^2 f(x) = \frac{d^2}{d^2\epsilon} f(x + \epsilon y)\big|_{\epsilon=0}$$

Then x is a critical point if and only if $D_y f(x) = 0$ for each $y \in K$ (in the next section we will discuss in more details this condition).

Here we focus on the function $f(t) = H(t) \equiv -t \log t$. In this case we need to be very careful when dealing with xx^* which have zero eigenvalues. We will see that for any $x, y \in \mathrm{C}^{m \times n}$, $D_y f(x) \in \mathrm{R}$. However it is possible that $D_y^2 f = \infty$, and below we give the exact conditions on y when this happens. Hence if x is a critical point of the von Neumann entropy, $H(x)$, and $D_y^2 H(x) = \infty$ then $H(x + \epsilon y) > H(x)$ for small enough ϵ. Thus when we study in the next sections the local minimum of $H(K_1 \otimes K_2)$ at the critical point $x_1 \otimes x_2$ we need only to consider y_i such that $D_{y_i}^2 f < \infty$ for $i = 1, 2$. This will also give a partial explanation of the local commutativity condition discussed in the introduction.

Lemma 1 *Let $x, y \in \mathrm{C}^{m \times n}$, $\mathrm{Tr}\, xx^* > 0$ and $H(t) = -t \log t$. Then $D_y H(x) \in \mathrm{R}$. Change* *standard orthonormal bases in* $\mathrm{C}^m, \mathrm{C}^n$ *to new orthonormal bases such that* x, y *have the forms*

$$x = \begin{bmatrix} x_{11} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} \quad and \quad y = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}, \tag{2}$$

with $0_{i,j} \in \mathrm{C}^{i\times j}$ *and* $x_{11}, y_{11} \in \mathrm{C}^{r\times r}$. *Then* $D_y^2 f(x) = \infty$ *if and only if* $y_{22} \neq 0$.

Proof. By considering UKV, where U, V unitary we may assume that x, y in the form (2). Furthermore $x_{11} = D \equiv \mathrm{diag}(d_1, \ldots, d_r)$, where $d_1 \geq d_2 \geq \ldots \geq d_r > 0$ and r is the rank of x. So d_i is the i-th singular value, $\sigma_i(x_{11})$ for $i = 1, \ldots, r$. Observe next that

$$\mathrm{Tr}((x + \epsilon y)(x + \epsilon y)^*) = \sum_{i=1}^{m} \lambda_i((x + \epsilon y)(x + \epsilon y)^*). \tag{3}$$

We assume here that the eigenvalues of a hermitian matrix are arranged in a nonincreasing order. Note that

$$(x + \epsilon y)(x + \epsilon y)^* = xx^* + \epsilon(xy^* + yx^*) + \epsilon^2(yy^*)$$

Hence

$$\lambda_i((x + \epsilon y)(x + \epsilon y)^*) = \lambda_i(xx^* + \epsilon(xy^* + yx^*)) + O(\epsilon^2).$$

Observe next that

$$xx^* + \epsilon(xy^* + yx^*) = \begin{bmatrix} D + \epsilon(Dy_{11}^* + y_{11}D) & \epsilon D y_{21}^* \\ \epsilon y_{21} D & 0 \end{bmatrix}.$$

For small ϵ, the first variation formula (see [5]) yields

$$\lambda_i((x + \epsilon y)(x + \epsilon y)^*) = d_i + d_i'\epsilon + O(\epsilon^2) \text{ for } i = 1, \ldots, r,$$
$$\lambda_i((x + \epsilon y)(x + \epsilon y)^*) = O(\epsilon^2) \text{ for } i > r.$$

Hence $\lambda_i((x + \epsilon y)(x + \epsilon y)^*) = d_i''\epsilon^2 + O(\epsilon^3)$ for $i > r$, with $d_{r+1}'' \geq \ldots \geq d_m'' \geq 0$. These calculations show that

$$H(x + \epsilon y) = H(D + \epsilon y_{11}) - \sum_{i=r+1}^{m} d_i''\epsilon^2 \log(d_i''\epsilon^2) + O(\epsilon^2).$$

Hence $D_y H(x) \in \mathrm{R}$ and $D_y^2 H(x) = \infty$ if and only if $d_{r+1}'' > 0$. It is left to show that $d_{r+1}'' > 0$ if and only if $y_{22} \neq 0$. Consider $\wedge^{r+1}(x + \epsilon y)$. the $r+1$ compound matrix of $x + \epsilon y$. (Recall that $\wedge^{r+1}(x + \epsilon y)$ is the $\binom{m}{r+1} \times \binom{n}{r+1}$ matrix whose entries are $(r+1) \times (r+1)$ minors of $x + \epsilon y$.) Note that $\wedge^{r+1}(x + \epsilon y)$ is a polynomial matrix in ϵ. Since x has rank r it follows that $\wedge^{r+1}(x) = 0$. Hence $\wedge^{r+1}(x + \epsilon y) = \epsilon z_1 + \epsilon^2 z_2(\epsilon)$, where z_1 is a constant matrix and $z_2(\epsilon)$ is a polynomial matrix in ϵ. We claim that $z_1 = 0$ if and only if $y_{22} = 0$. Indeed since D is diagonal then a minor of order $r+1$ that can have a nonzero derivative at $\epsilon = 0$ is the minor based on the rows $\alpha = \{1, \ldots, r, p\}$ and columns $\beta = \{1, \ldots, r, q\}$. Denote this minor by $\det(x+\epsilon y)[\alpha, \beta]$. Clearly $\det(x+\epsilon y)[\alpha, \beta] = \epsilon(d_1 \ldots d_r y_{p,q}) + O(\epsilon^2)$, where $y_{p,q}$ is the p, q entry of y. So if $y_{22} = 0$ we obtain that $z_1 = 0$. Hence $\| \wedge^{r+1} (x + \epsilon y)\|_2 = \sigma_1(\wedge^{r+1}(x + \epsilon y)) \leq \epsilon^2 a$ for some positive a. Recall that

$$(\sigma_1(\wedge^{r+1}(\xi + \epsilon y)))^2 = \prod_{i=1}^{r+1} \lambda_i((x + \epsilon y)(x + \epsilon y)^*).$$

As $(\sigma_1(\wedge^{r+1}(\xi+\epsilon y)))^2 \le a^2\epsilon^4$, we deduce that $d''_{r+1}=0$.

It is left to show that if $y_{p,q}\neq 0$ for some $p,q>r$, then $d''_{r+1}>0$. Clearly,

$$\| \wedge^{r+1}(x+\epsilon y)\|_2 \ge |\det(x+\epsilon y)[\alpha,\beta]| \ge d_1\dots d_r|y_{p,q}|\frac{|\epsilon|}{2}$$

for some small value of ϵ. (The first inequality follows from the fact the ℓ_2 norm of a matrix is not less than the absolute value of any of its entries.) This shows that $d''_{r+1}>0$. □

The lemma above implies that for the purpose of calculating local minima, without loss of generality, we can always take the directional derivatives in a direction with $y_{22}=0$. In the lemma above, however, we did not impose the normalization condition $\mathrm{Tr}(xx^*)=1$. As we show in the next lemma, it does not affect the result that $D^2_yH=\infty$ if and only if $y_{22}=0$.

Lemma 2 *Let $x,y\in \mathrm{C}^{m\times n}$, with $\mathrm{Tr}(xx^*)=1$ and $y\neq 0$. Consider the matrix*

$$x(y,\epsilon) := \frac{1}{\sqrt{\mathrm{Tr}((x+\epsilon y)(x+\epsilon y)^*)}}(x+\epsilon y)\ ,$$

which is always defined for small $|\epsilon|$. Then $\frac{d}{d\epsilon}H(x(y,\epsilon))|_{\epsilon=0}\in \mathrm{R}$, and $\frac{d^2}{d\epsilon^2}H(x(y,\epsilon))|_{\epsilon=0}=\infty$ if and only if $D^2_y(f)=\infty$.

Proof. The functions $h_1(\epsilon):=(\mathrm{Tr}((x+\epsilon y)(x+\epsilon y)^*))^{-1}$ and $h_2(\epsilon):=\log\mathrm{Tr}((x+\epsilon y)(x+\epsilon y)^*)$ are analytic in the neighborhood of $\epsilon=0$, and clearly

$$H(x(y,\epsilon)) = h_1(\epsilon)f(\epsilon)+h_2(\epsilon). \tag{4}$$

As $h_1(0)=1$ we obtain

$$\frac{d}{d\epsilon}H(x(y,\epsilon))|_{\epsilon=0} = D_y(f)+h_1'(0)H(x)+h_2'(0)\in \mathrm{R},$$

while $\frac{d^2}{d\epsilon^2}H(x(y,\epsilon))|_{\epsilon=0}$ consists of $D^2_y(f)$ plus finite terms. The lemma follows. □

The two lemmas above imply the following characterization of the local commutative condition discussed in the introduction.

Lemma 3 *Let the assumptions of Lemma 1 hold. Assume that x,y are in the form (2). Then xx^* commutes with xy^* if and only if $y_{21}=0$ and $x_{11}x_{11}^*$ commutes with $x_{11}y_{11}^*$ (which is equivalent to $x_{11}^*x_{11}y_{11}^* = y_{11}^*x_{11}x_{11}^*$.)*

Proof. Write x and y as in (2). The assumption that x_{11} is invertible, and xx^* commutes with xy^* is equivalent to $y_{21}=0$ and $x_{11}x_{11}^*$ commutes with $x_{11}y_{11}^*$. So $x_{11}x_{11}^*x_{11}y_{11}^* = x_{11}y_{11}^*x_{11}x_{11}^*$. Divide both sides of this equalities by x_{11} to obtain the lemma. □

In particular, the above lemma together with the theorem in Section 3 imply that local additivity holds for subspaces consisting of matrices y as in Eq. (2), with y_{11} diagonal, $y_{21}=0$, and y_{12} arbitrary.

2 First derivative of entropy functions under tensor products

All of the results in this section are due to the "Quantum Information Group" participating in the workshop "Geometry and representation theory", held at the American Institute for Mathematics [4]; we record the results here for completeness.

For a given function $f(t)$ as defined above, let $D_y f(x)$ denote the derivative of f in the y direction:

$$D_y f(x) = \frac{d}{d\epsilon} f(x+\epsilon y)\big|_{\epsilon=0}.$$

Then x is a critical point if and only if $D_y f(x) = 0$ for every y. Since we are interested in local minima in K subject to $\mathrm{Tr}[xx^*] = 1$, we restrict y to the tangent space $\{y \in K : D_y \mathrm{Tr}[xx^*] = 0\} = \{y \in K : \mathrm{Tr}[xy^* + yx^*] = 0\}$. Also, we restrict our attention to functions $f(x)$ which depend only on xx^*. Since xx^* is invariant under $x \mapsto ix$, we may ignore $y = ix$. That is, $x \in K$ is critical if and only if $D_y f(x) = 0$ for every y in the orthogonal subspace

$$x^{\perp} := \{y \in K : \mathrm{Tr}[xy^*] = 0\}.$$

Under tensor products, the orthogonal subspace has the following decomposition:

$$(x_1 \otimes x_2)^{\perp} \;=\; \langle x_1 \rangle \otimes x_2^{\perp} \;\;\oplus\;\; x_1^{\perp} \otimes \langle x_2 \rangle \;\;\oplus\;\; x_1^{\perp} \otimes x_2^{\perp}.$$

For a function $f(x)$ depending only on xx^*, a point $x \in K$ is critical in K if and only if $D_y f(x) = 0$ for every $y \in x^{\perp}$. In general, given a univariate differentiable function F, a Taylor series expansion of F shows that the matrix function $a \mapsto \mathrm{Tr}[F(a)]$ has directional derivative

$$\frac{d}{d\epsilon} \mathrm{Tr}[F(a+\epsilon b)]\big|_{\epsilon=0} = \mathrm{Tr}[F'(a)b].$$

We are interested in the case $a = xx^*$ and $b = xy^* + yx^*$: if $f(x) = \mathrm{Tr}[F(xx^*)]$, then

$$D_y f(x) = \mathrm{Tr}[F'(xx^*)(xy^* + yx^*)].$$

This derivative is zero for all $y \in x^{\perp}$ if and only if $\mathrm{Tr}[F'(xx^*)xy^*] = 0$ for all $y \in x^{\perp}$.

Theorem 1 *Let F be a differentiable univariate function such that $F'(a_1 \otimes a_2)$ is in the span of*

$$\{F'(a_1) \otimes F'(a_2), F'(a_1) \otimes I, I \otimes F'(a_2), I \otimes I\}.$$

If x_1 and x_2 are critical points of $f(x) = \mathrm{Tr}[F(xx^*)]$ subject to $\mathrm{Tr}[xx^*] = 1$, then so is $x_1 \otimes x_2$.

Proof. Let $x = x_1 \otimes x_2$. It suffices to show that if $D_{y_i} f(x_i) = 0$ for all $y_i \in x_i^{\perp}$, then $D_y f(x) = 0$ for all $y \in x^{\perp}$. That is, if $\mathrm{Tr}[F'(x_i x_i^*)x_i y_i^*] = 0$, then $\mathrm{Tr}[F'(xx^*)xy^*] = 0$.

First, suppose $y = y_1 \otimes y_2$, for some arbitrary y_1 and y_2, and consider the term in $F(xx^*)$ proportional to $F'(x_1 x_1^*) \otimes F'(x_2 x_2^*)$: we have

$$\begin{aligned}&\mathrm{Tr}[(F'(x_1 x_1^*) \otimes F'(x_2 x_2^*))\,(xy^*)] \\ &\quad = \mathrm{Tr}[F'(x_1 x_1^*)x_1 y_1^*]\,\mathrm{Tr}[F'(x_2 x_2^*)x_2 y_2^*],\end{aligned}$$

which is 0 provided that either $y_1 \in x_1^{\perp}$ or $y_2 \in x_2^{\perp}$ (or both). Likewise, for the term proportional to $F'(x_1 x_1^*) \otimes I$,

$$\mathrm{Tr}[(F'(x_1 x_1^*) \otimes I)\,(xy^*)] = \mathrm{Tr}[F'(x_1 x_1^*)x_1 y_1^*]\,\mathrm{Tr}[x_2 y_2^*],$$

which again is 0 if either $y_1 \in x_1^{\perp}$ or $y_2 \in x_2^{\perp}$. Similarly, $\mathrm{Tr}[(I \otimes F'(x_2 x_2^*))\,(xy^*)] = 0$ and $\mathrm{Tr}[(I \otimes I)\,(xy^*)] = 0$. Combining the terms which make up $F'(xx^*)$, we see that $\mathrm{Tr}[F'(xx^*)(xy^*)] = 0$ whenever $y = y_1 \otimes y_2$ satisfies $y_1 \in x_1^{\perp}$ or $y_2 \in x_2^{\perp}$.

Now an arbitrary element $y \in x^{\perp}$ can be written as a linear combination of terms of the form $x_1 \otimes y_2$, $y_1 \otimes x_2$, and $y_1 \otimes y_2$, with $y_i \in x_i^{\perp}$. For each of these terms either the first or second component of the tensor product is in $x_i^{\perp}$. Therefore $\mathrm{Tr}[F'(xx^*)xy^*] = 0$ for all $y \in x^{\perp}$. □

Our main interest is in the function $x \mapsto -\mathrm{Tr}[xx^* \ln xx^*]$, which is proportional to the usual von Neumann entropy of the matrix xx^*. Letting $F(t) = -t \ln t$, so that $F'(t) = -(1 + \ln t)$, we have

$$\begin{aligned} F'(a_1 \otimes a_2) &= -I - \ln(a_1 \otimes a_2) \\ &= -I \otimes I - \ln(a_1) \otimes I - I \otimes \ln(a_2) \\ &\in \mathrm{span}\left\{I \otimes I, F'(a_1) \otimes I, I \otimes F'(a_2)\right\}. \end{aligned}$$

(Here we used the fact that $\ln(a_1 \otimes a_2) = \ln(a_1) \otimes I + I \otimes \ln(a_2)$.) Thus the hypotheses of Theorem 1 are satisfied, and so critical points of $x \mapsto -\mathrm{Tr}[xx^* \ln xx^*]$ are closed under tensor products.

Another important class of entropy functions are the p-norms:

$$x \mapsto \|xx^*\|_p^p = \mathrm{Tr}[(xx^*)^p].$$

Letting $F(t) = t^p$, so $F'(t) = pt^{p-1}$, we have

$$F'(a_1 \otimes a_2) = p(a_1 \otimes a_2)^{p-1} = \frac{1}{p} F'(a_1) \otimes F'(a_2).$$

Again $F(t)$ is in the form of Theorem 1. Thus for both the von Neumann entropy and the p-norms, the tensor product of critical points (subject to $\mathrm{Tr}[xx^*] = 1$) are again critical points.

3 Second derivative of the von-Neumann entropy

In this section we show that under the local commutativity condition, if $x_1 \in K_1$ and $x_2 \in K_2$ are nonsingular strong local minima of

$$x \mapsto -\mathrm{Tr}[xx^* \log xx^*]$$

subject to $\mathrm{Tr}[xx^*] = 1$, then $x_1 \otimes x_2$ is also a strong local minimum in $K_1 \otimes K_2$. More precisely, we assume that if $y_i \in K_i$ is orthogonal to x_i, then $x_i x_i^*$ and $x_i y_i^*$ commute. Throughout this section we will also assume without loss of generality that $\mathrm{Tr}[yy^*] = 1$.

In this section we work with the normalized entropy function

$$\mathrm{H}(x) := -\mathrm{Tr}\left[\frac{xx^*}{\|x\|^2} \log \frac{xx^*}{\|x\|^2}\right].$$

A point x is a strong local minimum of H on $\{x : \mathrm{Tr}[xx^*] = 1\}$ if and only if for every y orthogonal to x, the second directional derivative $D_y^2 \mathrm{H}(x)$ is positive.

Lemma 4 *Assume xx^* and xy^* commute. Then*

$$\begin{aligned} D_y^2 \mathrm{H}(x) = 2\mathrm{Tr}\left[xx^* \log xx^*\right] - 2\mathrm{Tr}\left[yy^* \log xx^*\right] \\ -\mathrm{Tr}[(xy^* + yx^*)^2 (xx^*)^{-1}], \end{aligned}$$

where the last trace is taken over the support of xx^*.

Proof. For convenience define $a = xx^*$, $b = xy^* + yx^*$, and $c = yy^*$, so that

$$(x + \epsilon y)(x + \epsilon y)^* = a + \epsilon b + \epsilon^2 c.$$

Note that $\mathrm{Tr}[a] = \mathrm{Tr}[c] = 1$ and $\mathrm{Tr}[b] = 0$, so $\mathrm{Tr}[(x + \epsilon y)(x + \epsilon y)^*] = 1 + \epsilon^2$. Then

$$\begin{aligned}\mathrm{H}(x + \epsilon y) &= -\mathrm{Tr}\left[\frac{a + \epsilon b + \epsilon^2 c}{1 + \epsilon^2} \log \frac{a + \epsilon b + \epsilon^2 c}{1 + \epsilon^2}\right] \\ &= -\frac{\mathrm{Tr}[(a + \epsilon b + \epsilon^2 c)\log(a + \epsilon b + \epsilon^2 c)]}{1 + \epsilon^2} + \log(1 + \epsilon^2).\end{aligned}$$

Up to a second order in ε this expression becomes

$$\begin{aligned}&\mathrm{H}(x + \epsilon y) = -\mathrm{Tr}\left[a \log\left(a + \varepsilon b + \varepsilon^2 c\right)\right] \\ &- \varepsilon \mathrm{Tr}\left[b \log\left(a + \varepsilon b\right)\right] + \varepsilon^2\big(1 + \mathrm{Tr}\left[a \log a\right] - \mathrm{Tr}\left[c \log a\right]\big).\end{aligned}$$

Therefore, the second order directional derivative can be expressed in the following way:

$$\begin{aligned}&D_y^2 \mathrm{H}(x) = -\frac{d^2}{d\varepsilon^2} \mathrm{Tr}\left[a \log\left(a + \varepsilon b + \varepsilon^2 c\right)\right]\Big|_{\varepsilon=0} - \\ &2\frac{d}{d\varepsilon} \mathrm{Tr}\left[b \log\left(a + \varepsilon b\right)\right]\Big|_{\varepsilon=0} + 2\big(1 + \mathrm{Tr}\left[a \log a\right] - \mathrm{Tr}\left[c \log a\right]\big).\end{aligned}$$

To calculate the derivative expressions above, we will express the log function by its Taylor series:

$$\log(a + \varepsilon b) = \log[I - (I - a - \varepsilon b)] = -\sum_{n=1}^{\infty} \frac{(I - a - \varepsilon b)^n}{n}.$$

Without loss of generality (see Lemma 1, in the last equality we assumed that a is invertible, so that for sufficiently small ε also $a + \varepsilon b$ is invertible and therefore $I - a - \varepsilon b < I$. To calculate the derivative of $\mathrm{Tr}\left[b \log\left(a + \varepsilon b\right)\right]$ at $\varepsilon = 0$, we only need to take terms proportional to ε in the expansion of the logarithm. Assuming a and b commute,

$$\begin{aligned}\frac{d}{d\varepsilon} \mathrm{Tr}\left[b \log\left(a + \varepsilon b\right)\right]\Big|_{\varepsilon=0} &= \sum_{n=1}^{\infty} \mathrm{Tr}\left[b^2 (I - a)^{n-1}\right] \\ &= \mathrm{Tr}\left[b^2 a^{-1}\right].\end{aligned}$$

To calculate the second derivative of $\mathrm{Tr}[a \log(a + \varepsilon b + \varepsilon^2 c)]$ we need only take the terms proportional to ε^2. Again assuming a and b commute and a is invertible,

$$\begin{aligned}&\frac{d^2}{d\varepsilon^2} \mathrm{Tr}\left[a \log\left(a + \varepsilon b + \varepsilon^2 c\right)\right]\Big|_{\varepsilon=0} \\ &= -\sum_{n=2}^{\infty} \frac{2}{n}\binom{n}{2} \mathrm{Tr}\left[a(I - a)^{n-2} b^2\right] + 2\sum_{n=1}^{\infty} \mathrm{Tr}\left[a(I - a)^{n-1} c\right] \\ &= -\mathrm{Tr}\left[a^{-1} b^2\right] + 2\,\mathrm{Tr}[c] = -\mathrm{Tr}\left[a^{-1} b^2\right] + 2.\end{aligned}$$

Therefore $D_y^2 \mathrm{H}(x) = 2\mathrm{Tr}\left[a \log a\right] - 2\mathrm{Tr}\left[c \log a\right] - \mathrm{Tr}[b^2 a^{-1}]$. □

Corollary 1 *Assume xx^* and xy^* commute. Then $D_y^2\mathrm{H}(x) > 0$ if and only if*

$$\begin{aligned}&\left|\mathrm{Tr}[(xx^*)^{-1}(xy^*)^2]\right| + \mathrm{Tr}[(xx^*)^{-1}xy^*yx^*]\\&\qquad < \mathrm{Tr}[xx^* \log xx^*] - \mathrm{Tr}[yy^* \log xx^*] \ ,\end{aligned}$$

where $(xx^*)^{-1}$ is the inverse over the support of xx^*.

Proof. Expand $(xy^* + yx^*)^2$ into four terms, noting that xy^* and yx^* commute with $(xx^*)^{-1}$. Then $D_y\mathrm{H}(x) > 0$ if and only if

$$\begin{aligned}&\mathrm{Tr}[(xx^*)^{-1}(xy^*)^2] + \mathrm{Tr}[(xx^*)^{-1} + (yx^*)^2] + 2\,\mathrm{Tr}[(xx^*)^{-1}xy^*yx^*]\\&< -2\,\mathrm{Tr}[yy^* \log xx^*] + 2\,\mathrm{Tr}[xx^* \log xx^*].\end{aligned}$$

The first two terms on the LHS are twice the real part of $\mathrm{Tr}[(xx^*)^{-1}(xy^*)^2]$; the largest value of these two terms over all phases of y is $2\left|\mathrm{Tr}[(xx^*)^{-1}(xy^*)^2]\right|$. □

For convenience, denote the terms in Corollary 1 as follows:

$$\begin{aligned}a(x,y) &:= \left|\mathrm{Tr}[(xx^*)^{-1}(xy^*)^2]\right| ,\\ b(x,y) &:= \mathrm{Tr}[(xx^*)^{-1}xy^*yx^*],\\ c(x) &:= \mathrm{Tr}[xx^* \log xx^*],\\ d(x,y) &:= \mathrm{Tr}[yy^* \log xx^*],\end{aligned} \tag{5}$$

so $D_y^2\mathrm{H}(x) > 0$ if and only if $a + b < c - d$. Each of these terms behaves nicely under tensor products:

$$\begin{aligned}a(x_1 \otimes x_2, y_1 \otimes y_2) &= a(x_1, y_1)a(x_2, y_2), \qquad (6)\\ b(x_1 \otimes x_2, y_1 \otimes y_2) &= b(x_1, y_1)b(x_2, y_2),\\ c(x_1 \otimes x_2) &= c(x_1) + c(x_2),\\ d(x_1 \otimes x_2, y_1 \otimes y_2) &= d(x_1, y_1) + d(x_2, y_2).\end{aligned}$$

We can also bound the size of some of these terms for any x and y such that $\mathrm{Tr}[xy^*] = 0$ and $\mathrm{Tr}[xx^*] = \mathrm{Tr}[yy^*] = 1$. First, we claim $b \in [0, 1]$. To see this, note that $P = x^*(xx^*)^{-1}x$ is a projection matrix, so

$$b = \mathrm{Tr}[y^*yx^*(xx^*)^{-1}x] = \|Py^*\|^2,$$

and $0 \le \|Py^*\|^2 \le \|y^*\| = 1$. Second, we claim $a \in [0, b]$. To see this, note that without loss of generality (see Lemma 1) we can assume that xx^* is invertible and therefore positive definite, so $(xx^*)^{-1/2}$ exists and commutes with xy^*, and so $(xx^*)^{-1}(xy^*)^2 = ((xx^*)^{-1/2}xy^*)^2$. By Cauchy-Schwarz,

$$\begin{aligned}a &= \left|\mathrm{Tr}[((xx^*)^{-1/2}xy^*)^2]\right|\\ &\le \mathrm{Tr}[((xx^*)^{-1/2}xy^*)((xx^*)^{-1/2}xy^*)^*] = b.\end{aligned}$$

Thirdly, we claim that $c \le 0$, since it is the negative of the entropy function. We are now ready to prove the main result of this paper.

Theorem 2 *Suppose x_1 and x_2 are strong local minima of $x \mapsto -\operatorname{Tr}[xx^* \log xx^*]$ subject to $\operatorname{Tr}[xx^*] = 1$ and $x_i \in K_i$, where K_i is a subspace. Further assume that for every $y_i \in K_i$, the matrices $x_i x_i^*$ and $x_i y_i^*$ commute. Then $x := x_1 \otimes x_2$ is a strong local minimum in $K_1 \otimes K_2$.*

Proof. We show that under the hypotheses of the theorem, if $D^2_{y_i}\mathrm{H}(x_i)$ is positive for every $y_i \in x_i^\perp$, then $D^2_y\mathrm{H}(x)$ is positive for every $y \in x^\perp$. We break the proof into several cases depending on y.

First, suppose y is a tensor product.

Case $y = x_1 \otimes y_2$, $y_2 \in x_2^\perp$: Since $y_2 \in x_2^\perp$ and x_2 is a strong local minimum, we know that

$$a(x_2, y_2) + b(x_2, y_2) < c(x_2) - d(x_2, y_2).$$

It is also easy to see from the expressions (5) that

$$a(x_1, x_1) = b(x_1, x_1) = 1, \quad c(x_1) = d(x_1, x_1).$$

So, using the expressions for tensors in (6), we have

$$\begin{aligned} a(x, y) + b(x, y) &= a(x_2, y_2) + b(x_2, y_2) \\ &\le c(x_2) - d(x_2, y_2) \\ &= c(x) - d(x, y). \end{aligned}$$

Thus the second directional derivative is positive for this choice of y.

Case $y = y_1 \otimes x_2$, $y_1 \in x_1^\perp$: This case is similar to $y = x_1 \otimes y_2$.

Case $y = y_1 \otimes y_2$, $y_i \in x_i^\perp$: Here we require the arithmetic-geometric mean inequality. For two terms $a_1, a_2 \le 1$,

$$a_1 a_2 \le \Big(\frac{a_1 + a_2}{2}\Big)^2 \le \frac{1}{2}(a_1 + a_2).$$

In particular, $a(x_1, x_1)a(x_2, y_2) \le a(x_1, y_1) + a(x_2, y_2)$ and similarly for b. Now, since $y_i \in x_i^\perp$, we have $a(x_i, y_i) + b(x_i, y_i) < c(x_i) - d(x_i, y_i)$. Combining these inequalities we get $a(x, y) + b(x, y) \le c(x) - d(x, y)$.

Next, we consider cases where y is a linear combination of terms.

Suppose y is in $x_1^\perp \otimes x_2^\perp$. In this case, we break y into two orthogonal pieces according to the projection matrix $P = x^*(xx^*)^{-1}x$. Let $P_i = x_i^*(x_i x_i^*)^{-1}x_i$: this is the projection matrix onto the range of x_i^*, which we denote $R(x_i^*)$. Then $P = P_1 \otimes P_2$ is the projection matrix onto the range $R(x^*) = R(x_1^*) \otimes R(x_2^*)$. Write y as a direct sum:

$$y = \alpha u + \beta v,$$

where $u^* \in R(x^*)$ (so $Pu^* = u^*$), and $Pv^* = 0$. The normalizations are chosen so that $\alpha \in R$ and $\beta \in R$ satisfy $\alpha^2 + \beta^2 = 1$, and $\|u\|^2 = \|v\|^2 = 1$. We deal with the u and v components separately.

Case $y = u \in (x_1^\perp \otimes x_2^\perp) \cap R(x^*)$: Here we have $b(x, u) = \|Pu^*\|^2 = \|u^*\|^2 = 1$. Note that if y_i is in $x_i^\perp$, then

$$\operatorname{Tr}[x_i P_i y_i^*] = \operatorname{Tr}[x_i x_i^*(x_i x_i^*)^{-1} x_i y_i^*] = \operatorname{Tr}[x_i y_i^*] = 0,$$

so $P_i y_i^*$ is also in $x_i^\perp$. If we write $u = \sum_j y_{1j} \otimes y_{2j}$ with $y_{ij} \in x_i^\perp$, so that

$$u = Pu = \sum_j P_1 y_{1j} \otimes P_2 y_{2j},$$

then $P_i y_{ij}$ is in $x_i^\perp \cap R(x_i^*)$, and it follows that u is in $(x_1^\perp \cap R(x_1^*)) \otimes (x_2^\perp \cap R(x_2^*))$. Now perform a Schmidt decomposition of u with respect to this tensor space: we get

$$u = \sum_j \alpha_j u_{1j} \otimes u_{2j},$$

where $u_{ij} \in x_i^\perp \cap R(x_i^*)$, $\mathrm{Tr}[u_{ij} u_{ik}^*] = \delta_{jk}$, $\alpha_j \geq 0$, and $\sum_j \alpha_j^2 = 1$. Since u_{ij} is in $R(x_i^*)$, we have $b(x_i, u_{ij}) = 1$. Since u_{ij} is in $x_i^\perp$, we know

$$a(x_i, u_{ij}) + b(x_i, u_{ij}) \leq c(x_i) - d(x_i, u_{ij}), \tag{7}$$

and also $0 \leq a(x_i, u_{ij}) \leq 1$. Under this decomposition, we also have

$$d(x, u) = \sum_j \alpha_j^2 (d(x_1, u_{1,j}) + d(x_2, u_{2,j})). \tag{8}$$

Therefore, from (7) and (8),

$$\begin{aligned} a(x,u) &+ b(x,u) \\ &\leq 2b(x,u) \\ &= \sum_j \alpha_j^2 [b(x_1, u_{1,j}) + b(x_2, u_{1,2})] \\ &\leq \sum_j \alpha_j^2 [c(x_1) - d(x_1, u_{1,j})] + c(x_2) - d(x_2, u_{1,2}) \\ &= c(x) - d(x,u). \end{aligned}$$

Case $y = v \in x_1^\perp \otimes x_2^\perp$, $Pv^* = 0$: We know that $0 \leq a(x, v) \leq b(x, v) = \|Pv^*\| = 0$, and so $a(x, v) = b(x, v) = 0$. Perform a Schmidt decomposition of v with respect to the space $x_1^\perp \otimes x_2^\perp$:

$$v = \sum_j \beta_j v_{1j} \otimes v_{2j},$$

where $v_{ij} \in x_i^\perp$, $\mathrm{Tr}[v_{ij} v_{ik}^*] = \delta_{jk}$ and $\sum_j \beta_j^2 = 1$. Since v_{ij} is in $x_i^\perp$, we have $0 \leq a(x_i, v_{ij}) \leq b(x_i, v_{ij})$ and

$$0 \leq a(x_i, v_{ij}) + b(x_i, v_{ij}) \leq c(x_i) - d(x_i, v_{ij}). \tag{9}$$

It follows quickly that $a(x, v) + b(x, v) \leq c(x) - d(x, v)$.

Next we deal with a combination of u and v.

Case $y \in x_1^\perp \otimes x_2^\perp$: Write $y = \alpha u + \beta v$, where $u^* \in R(x^*))$, $Pv^* = 0$, $\alpha^2 + \beta^2 = 1$, and $\|u\|^2 = \|v\|^2 = 1$. Then since $uv^* = uPv^* = 0$, we have

$$yy^* = \alpha^2 uu^* + \beta^2 vv^*,$$

from which it follows that

$$b(x,y) = \alpha^2 b(x,u) + \beta^2 b(x,v), \tag{10}$$

$$d(x,y) = \alpha^2 d(x,u) + \beta^2 d(x,v). \tag{11}$$

(In fact, $b(x,u) = 1$ and $b(x,v) = 0$.) Combining (10) and (11) with the results for u and v from the previous cases, we get

$$\begin{aligned} a(x,y) + b(x,y) &\leq 2b(x,y) \\ &= \alpha^2 2b(x,u) + \beta^2 2b(x,v) \\ &\leq \alpha^2[c(x) - d(x,u)] + \beta^2[c(x) - d(x,v)] \\ &= c(x) - d(x,y). \end{aligned}$$

Finally, we have the case where y is an arbitrary element of $x^{\perp}$.

Case $y \in x^{\perp}$: Here y may be written in the form

$$y = \alpha x_1 \otimes y_2 + \beta y_1 \otimes x_2 + \gamma y',$$

where $y_i \in x_i^{\perp}$ and $y' \in x_1^{\perp} \otimes x_2^{\perp}$, with real constants satifying $\alpha^2 + \beta^2 + \gamma^2 = 1$. Expanding out terms of yy^* and simplifying, we find that most terms disappear under trace:

$$\begin{aligned} d(x,y) = \alpha^2[c(x_1) + d(x_2,y_2)] + \beta^2[c(x_2) + d(x_1,y_1)] \\ + \gamma^2 d(x,y'), \end{aligned} \tag{12}$$

$$b(x,y) = \alpha^2 b(x_2,y_2) + \beta^2 b(x_1,y_1) + \gamma^2 b(x,y'), \tag{13}$$

$$\begin{aligned} a(x,y) = \Big|\alpha^2 \operatorname{Tr}[(x_2x_2^*)^{-1}(x_2y_2^*)^2] \\ +\beta^2 \operatorname{Tr}[(x_1x_1^*)^{-1}(x_1y_1^*)^2] + \gamma^2 \operatorname{Tr}[(xx^*)^{-1}(x(y')^*)^2]\Big|. \end{aligned} \tag{14}$$

The expression for $d(x,y)$ requires the observation that $\operatorname{Tr}[x_iy_i^* \log x_ix_i^*] = 0$, because the first directional derivative of $D_{y_i}\mathrm{H}(x_i)$ is 0 when x_i is a local minimum. The expression for $a(x,y)$ is bounded as follows:

$$\begin{aligned} a(x,y) &= \Big|\alpha^2 \operatorname{Tr}[(x_2x_2^*)^{-1}(x_2y_2^*)^2] \\ &+ \beta^2 \operatorname{Tr}[(x_1x_1^*)^{-1}(x_1y_1^*)^2] + \gamma^2 \operatorname{Tr}[(xx^*)^{-1}(x(y')^*)^2]\Big| \\ &\leq \alpha^2 \left|\operatorname{Tr}[(x_2x_2^*)^{-1}(x_2y_2^*)^2]\right| \\ &+ \beta^2 \left|\operatorname{Tr}[(x_1x_1^*)^{-1}(x_1y_1^*)^2]\right| + \gamma^2 \left|\operatorname{Tr}[(xx^*)^{-1}(x(y')^*)^2]\right| \\ &= \alpha^2 a(x_2,y_2) + \beta^2 a(x_1,y_1) + \gamma^2 a(x,y'). \end{aligned} \tag{15}$$

Combining (12), (13) and (15), we get $a(x,y) + b(x,y) \leq c(x) - d(x,y)$. □

4 The second derivative of the 2-norm

In this section we focus on the 2-norm since its second directional derivative has an elegant analytical form. We prove that if K_1 and K_2 are subspaces of matrices, at least one of which has dimension 2, and $x_1 \in K_1$, $x_2 \in K_2$ are strong local maxima of the 2-norm function

$$x \mapsto \operatorname{Tr}[(xx^*)^2]$$

subject to $\mathrm{Tr}[xx^*] = 1$, then $x_1 \otimes x_2$ is also a strong local maximum in $K_1 \otimes K_2$. Since it is known that the 2-norm is not globally additive, this result sheds some light on the possibility that there exist functions that are locally additive while they are not globally additive.

We will work with the normalized function

$$\mathrm{H}_2(x) := \mathrm{Tr}\left[\left(\frac{xx^*}{\|x\|^2}\right)^2\right] = \frac{\mathrm{Tr}[(xx^*)^2]}{[\mathrm{Tr}(xx^*)]^2}.$$

As before, we consider $(x+\epsilon y)(x+\epsilon y)^* = xx^* + \epsilon(xy^* + yx^*) + \epsilon^2(yy^*)$, where $\mathrm{Tr}[xx^*] = \mathrm{Tr}[yy^*] = 1$ and $\mathrm{Tr}[xy^* + yx^*] = 0$. Noting that

$$\begin{aligned}[(x+\epsilon y)(x+\epsilon y)^*]^2 &= (xx^*)^2 + 2\epsilon xx^*(xy^* + yx^*) \\ &+\epsilon^2[2xx^*yy^* + (xy^* + yx^*)^2] + O(\epsilon^3),\end{aligned}$$

and that $\mathrm{Tr}[(x+\epsilon y)(x+\epsilon y)^*]^2 = (1+\epsilon^2)^2$, we have that up to second order in ϵ,

$$\begin{aligned}\mathrm{H}_2(x+\epsilon y) &= \mathrm{Tr}[(xx^*)^2] + 2\epsilon\,\mathrm{Tr}[xx^*(xy^* + yx^*)] \\ &+ \epsilon^2\,\mathrm{Tr}[2xx^*yy^* + (xy^* + yx^*)^2 - 2(xx^*)^2].\end{aligned}$$

Then the first directional derivative of H_2 is

$$D_y\mathrm{H}_2(x) = 2\,\mathrm{Tr}[xx^*(xy^* + yx^*)].$$

By considering iy as well as y, the condition $D_y\mathrm{H}_2(x) = 0$ reduces to $\mathrm{Tr}[xx^*xy^*] = 0$. The second derivative is

$$D_y^2\mathrm{H}_2(x) = 2\,\mathrm{Tr}[(xy^* + yx^*)^2] + 4\,\mathrm{Tr}[xx^*yy^*] - 4\,\mathrm{Tr}[(xx^*)^2].$$

For strong local maxima we expect $D_y^2\mathrm{H}_2(x)$ to be negative. Expand $(xy^* + yx^*)^2$ into four terms: then

$$\mathrm{Tr}[(xy^* + yx^*)^2] = 2\,\mathrm{Re}\,\mathrm{Tr}[(xy^*)^2] + \mathrm{Tr}[x^*xy^*y],$$

where $\mathrm{Re}(z)$ denotes the real part of z. The largest value of $\mathrm{Re}\,\mathrm{Tr}[(xy^*)^2]$ over all choices of unit multiples of y is $\left|\mathrm{Tr}[(xy^*)^2]\right|$. In summary:

Lemma 5 *Define* $F(x,y) :=$

$$-\,\mathrm{Tr}[(xx^*)^2] + \left|\mathrm{Tr}[(xy^*)^2]\right| + \mathrm{Tr}[xx^*yy^*] + \mathrm{Tr}[x^*xy^*y].$$

Then x is a strong local maximum of the function $x \mapsto \mathrm{Tr}[(xx^*)^2]$, subject to $\mathrm{Tr}[xx^*] = 1$, if and only if every $y \in x^\perp$, $\mathrm{Tr}[yy^*] = 1$ satisfies $\mathrm{Tr}[xx^*xy^*] = 0$ and $F(x,y) < 0$.

Denote the terms in $F(x,y)$ as follows:

$$\begin{aligned}a(x) &:= \mathrm{Tr}[(xx^*)^2] \\ b(x,y) &:= \left|\mathrm{Tr}[(xy^*)^2]\right| \\ c(x,y) &:= \mathrm{Tr}[xx^*yy^*] \\ d(x,y) &:= \mathrm{Tr}[x^*xy^*y].\end{aligned}$$

The Cauchy-Schwartz inequality implies that $\left|\mathrm{Tr}[z^2]\right| \leq \mathrm{Tr}[z^* z]$ for any matrix z. Letting $z = xy^*$, we conclude that $0 \leq b(x,y) \leq c(x,y), d(x,y)$. Assuming $a > b + c + d$ therefore implies that $a > b, c, d$. Since $a = \mathrm{Tr}[(xx^*)^2)] \leq \mathrm{Tr}[xx^*] = 1$, we see that each of the terms a, b, c, d are in the range $[0,1]$. Furthermore, each term is multiplicative under tensor products: $a(x_1 \otimes x_2) = a(x_1)a(x_2)$, $b(x_1 \otimes x_2, y_1 \otimes y_2) = b(x_1, y_1)b(x_2, y_2)$, and so on.

From Section 2, we know that that tensor products of critical points of the 2-norm are again critical points. We can now say the same for local maxima.

Lemma 6 *Suppose* x_1 *and* x_2 *are strong local maxima of* $x \mapsto \mathrm{Tr}[(xx^*)^2]$ *subject to* $\mathrm{Tr}[xx^*] = 1$ *and* $x_i \in K_i$*, where either* K_1 *or* K_2 *has dimension* 2*. Then* $x := x_1 \otimes x_2$ *is a strong local maximum in* $K_1 \otimes K_2$*.*

Proof. Without loss of generality, assume K_1 has dimension 2, so $x_1^\perp$ has dimension 1. Let y_1 be an element of $x_1^\perp$ and let y_{2j} be elements of $x_2^\perp$. Then every element of $x^\perp$ in $K_1 \otimes K_2$ is a linear combination of vectors of the form $y_1 \otimes y_{21}$, $x_1 \otimes y_{22}$, and $y_1 \otimes x_2$. First, we check that for each y of that form, $F(x,y)$ is negative.

Case $y = y_1 \otimes y_{21}$: Here

$$F(x,y) = -a(x_1)a(x_2) + b(x_1,y_1)b(x_2,y_{21}) \\ + c(x_1,y_1)c(x_2,y_{21}) + d(x_1,y_1)d(x_2,y_{21}).$$

Since $a(x_i) > b(x_i, y_i) + c(x_i, y_i) + d(x_i, y_i)$ and each term is nonnegative, it follows that $a(x_1)a(x_2) > b(x_1,y_1)b(x_2,y_{21}) + c(x_1,y_1)c(x_2,y_{21}) + d(x_1,y_1)d(x_2,y_{21})$, and so $F(x,y)$ is negative.

Case $y = x_1 \otimes y_{22}$: Here

$$F(x,y) = -a(x_1)a(x_2) + a(x_1)b(x_2,y_{22}) \\ + a(x_1)c(x_2,y_{22}) + a(x_1)d(x_2,y_{22}).$$

But $-a(x_2) + b(x_2,y_{22}) + c(x_2,y_{22}) + d(x_2,y_{22}) < 0$ and $a(x_1) > 0$, so $F(x,y)$ is negative.

Case $y = y_1 \otimes x_2$: Similar to $y = x_1 \otimes y_{22}$.

Now consider a linear combination of the three elements of $x^\perp$, say

$$y = \alpha(y_1 \otimes y_{21}) + \beta(x_1 \otimes y_{22}) + \gamma(y_1 \otimes x_2).$$

In considering $b(x,y)$, most terms disappear under trace:

$$b(x,y) = \Big|\alpha^2 \mathrm{Tr}[(x_1 y_1^*)^2] \mathrm{Tr}[(x_2 y_{21}^*)^2] \\ + \beta^2 \mathrm{Tr}[(x_1 x_1^*)^2] \mathrm{Tr}[(x_2 y_{22}^*)^2] + \gamma^2 \mathrm{Tr}[(x_1 y_1^*)^2] \mathrm{Tr}[(x_2 x_2^*)^2]\Big| \\ \leq |\alpha|^2 b(x_1,y_1)b(x_2,y_{21}) \\ + |\beta|^2 a(x_1)b(x_2,y_{22}) + |\gamma|^2 b(x_1,y_1)a(x_2).$$

Likewise we have

$$c(x,y) = |\alpha|^2 c(x_1,y_1)c(x_2,y_{21}) + \\ |\beta|^2 a(x_1)c(x_2,y_{22}) + |\gamma|^2 c(x_1,y_1)a(x_2),$$

and similarly for $d(x,y)$. Adding together, we conclude that

$$\begin{aligned} F(x,y) \le |\alpha|^2 &\left[\neg a(x_1)a(x_2) + b(x_1,y_1)b(x_2,y_{21})\right. \\ &\left. + c(x_1,y_1)c(x_2,y_{21}) + d(x_1,y_1)d(x_2,y_{21})\right] \\ + |\beta|^2 a(x_1) &\left[-a(x_2) + b(x_2,y_{22}) + c(x_2,y_{22}) + d(x_2,y_{22})\right] \\ + |\gamma|^2 &\left[-a(x_1) + b(x_1,y_1) + c(x_1,y_1) + d(x_1,y_1)\right] a(x_2). \end{aligned}$$

The $|\alpha|^2$ term is negative by the argument given in the case $y = y_1 \otimes y_{21}$; the $|\beta|^2$ term is negative by the case $y = x_1 \otimes y_{22}$; and the $|\gamma|^2$ term is negative by the case $y = y_1 \otimes x_2$. □

If both K_1 and K_2 have dimension higher than 2, the linear combinations seem to be more difficult.

Acknowledgments

We appreciate many valuable discussions with Jon Yard. Research by GG and AR was supported by NSERC, PIMS, and iCORE.

References

1. M. B. Hastings. A counterexample to additivity of minimum output entropy. *Nature Physics* 5, 255, 2009.
2. Alexander S. Holevo. The additivity problem in quantum information theory. In *International Congress of Mathematicians. Vol. III*, pages 999–1018. Eur. Math. Soc., Zürich, 2006.
3. Peter W. Shor. Equivalence of additivity questions in quantum information theory. *Comm. Math. Phys.*, 246(3):453–472, 2004.
4. Harm Derksen, Shmuel Friedland, Gilad Gour, David Gross, Leonid Gurvits, Aidan Roy, and Jon Yard. On minimum entropy output and the additivity conjecture. Notes of Quantum Information Group, American Institute for Mathematics workshop "Geometry and representation theory of tensors for computer science, statistics and other areas", July 21-25, 2008.
5. T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, 2nd ed., New York 1980.

Appendix A Affine parametrization

Let $x \in K \subset \mathrm{C}^{m\times n}$ and assume that $xx^* > 0$, i.e. xx^* invertible, and $\mathrm{Tr}\, xx^* = 1$. When now study the second variation. Let $y \in K$ and assume that $\mathrm{Tr}(xy^* + yx^*) = 0$. We then consider

$$\begin{aligned} &A(\epsilon) := A + \epsilon B + \epsilon^2 C = (x+\epsilon y)(x+\epsilon y)^*, \\ &A = xx^*, B = xy^* + yx^*, C = yy^* \in \mathrm{H}_m. \end{aligned} \tag{A.1}$$

(Here H_m is the real space of $m \times m$ matrices.) Let $\lambda_1(\epsilon), \ldots, \lambda_m(\epsilon) > 0$ be the eigenvalues of $A(\epsilon)$, as analytic functions of ϵ, (Rellich's theorem [5]). We can assume that these eigenvalues are arranged in the following order $\lambda_1(\epsilon) \ge \ldots \ge \lambda_m(\epsilon) > 0$ for small **positive** ϵ. Let $A_1(\epsilon) = A + \epsilon B$. Arrange the analytic eigenvalues of $A_1(\epsilon)$ in the order $\mu_1(\epsilon) \ge \ldots \ge \mu_m(\epsilon) > 0$ for small positive ϵ. Clearly, $\lambda_i(\epsilon) = \mu_i(\epsilon) + O(\epsilon^2)$ for $i = 1, \ldots, m$. The following result is known, and can be deduced from the arguments in Kato [5].

Lemma 7 *Let $A, B, C \in \mathrm{H}_m$, and denote $A(\epsilon) = A + \epsilon B + \epsilon^2 C, A_1(\epsilon) = A + \epsilon B$. Assume that $\lambda_1(\epsilon), \ldots, \lambda_m(\epsilon)$ and $\mu_1(\epsilon), \ldots, \mu_m(\epsilon)$ are analytic eigenvalues of $A(\epsilon), A_1(\epsilon)$ arranged in* *a nonincreasing order for small positive ϵ. Then, there exists a unitary matrix $U \in \mathrm{C}^{m\times m}$ with the following two properties. First, $UAU^* = \mathrm{diag}(\lambda_1,\dots,\lambda_m)$. Second, if we denote $UCU^* \equiv F = [f_{ij}]_{i=j=1}^m$ then*

$$\lambda_i(\epsilon) = \mu_i(\epsilon) + \epsilon^2 f_{ii} + O(\epsilon^3) \text{ for } i = 1,\dots,m. \tag{A.2}$$

In the next proposition we use the above lemma to calculate the variation of $S(A) \equiv -\mathrm{Tr}(A\log A)$ up to second order.

Proposition 1. *Let $x, y \in \mathrm{C}^{m\times n}$ and assume that $\mathrm{Tr}(xx^*) = \mathrm{Tr}(yy^*) = 1, xx^* > 0, \mathrm{Tr}(xy^* + yx^*) = 0$. Define $A(\epsilon), A_1(\epsilon)$ as in Eq.* (A.1)*. Then*

$$S\left(\frac{A(\epsilon)}{\mathrm{Tr}\,A(\epsilon)}\right) = S(A_1(\epsilon)) + \epsilon^2\,\mathrm{Tr}\left[(xx^* - yy^*)\log xx^*\right] + O(\epsilon^3) \tag{A.3}$$

Proof. First recall that

$$\begin{aligned} S\left(\frac{A(\epsilon)}{\mathrm{Tr}\,A(\epsilon)}\right) &= \frac{1}{\mathrm{Tr}\,A(\epsilon)} S(A(\epsilon)) + \log\mathrm{Tr}\,A(\epsilon) \\ &= S(A(\epsilon)) + \epsilon^2\left[-\mathrm{Tr}(yy^*)S(xx^*) + \mathrm{Tr}(yy^*)\right] + O(\epsilon^3). \end{aligned} \tag{A.4}$$

Next we claim

$$\begin{aligned} S(A(\epsilon)) &= -\sum_{i=1}^m \lambda_i(\epsilon)\log\lambda_i(\epsilon) \\ &= -\sum_{i=1}^m (\mu_i(\epsilon) + f_{ii}\epsilon^2)\log(\mu_i(\epsilon) + f_{ii}\epsilon^2) + O(\epsilon^3) = \\ &-\sum_{i=1}^m \mu_i(\epsilon)\log\mu_i(\epsilon) - \epsilon^2(\sum_{i=1}^m f_{ii}\log\lambda_i + \sum_{i=1}^m f_{ii}) + O(\epsilon^3) \\ &= S(A_1(\epsilon)) - \epsilon^2(\mathrm{Tr}((yy^*)\log(xx^*)) + \mathrm{Tr}(yy^*)) + O(\epsilon^3). \end{aligned}$$

Combine this expression with the expression above it to deduce (A.3). □

Note that the expression $\mathrm{Tr}\left[(xx^* - yy^*)\log xx^*\right]$ can be either positive or negative. In the following we give a very simple reason why we can not ignore this term (i.e. use the affine approximation), which also yields a necessary condition $xx*$ must satisfy if x is a local minimum.

Assume that we have an affine subspace of the form $A + tB$, where $\mathrm{Tr}(A) = 1, \mathrm{Tr}(B) = 0$. Here $A = xx^*, B = xy^* + yx^*$ on all $y \in K$ satisfying the condition $\mathrm{Tr}(B) = 0$ and t arbitrary real. Let Φ be the set of all $A + tB$ such that $A + tB \geq 0$. Consider the function $S(C) = -\mathrm{Tr}(C\log C)$ where $C \in \Phi$. Our assumption that A is a critical point in Φ for the $S(C)$. Since $S(C)$ is strictly concave on Φ it follows that A is a unique global MAXIMUM on Φ! So if A was a local minimum for the $H(x), x \in K, \mathrm{Tr}(xx^*) = 1$ it follows that the correction term for ϵ^2 that we have must be *strictly* positive . That is, if x is a local min then

$$\begin{aligned} &\mathrm{Tr}\left[(xx^* - yy^*)\log xx^*\right] \\ &= S(yy^*) - S(xx^*) + S(yy^*\|xx^*) > 0 \end{aligned}$$

for all $y \in x^{\perp}$ (assuming the normalization $\mathrm{Tr}(yy^*) = 1$).

Appendix B A counter example to real additivity conjecture

During the 2008 American Institute for Mathematics workshop "Geometry and representation theory" [4], Leonid Gurvits found a counterexample to the analogue of the additivity conjecture for real (rather than complex) matrices. In this appendix we generalize the counterexample to show that the additivity conjecture fails to hold for real spaces of orthogonal matrices containing the identity: there exist real subspaces $K_1 \subseteq R^{m_1 \times n_1}$ and $K_2 \subseteq R^{m_2 \times n_2}$ such $\mathrm{H}(K_1 \otimes K_2) < \mathrm{H}(K_1) + \mathrm{H}(K_2)$.

$K \subseteq R^{m \times m}$ is called an orthogonal subspace if any $0 \neq A \in K$ is of the form aQ for some scalar a and an orthogonal matrix Q. Note that if K is an orthogonal subspace then for any orthogonal matrix Q_0, the subspace $Q_0^{\top} K$ is also an orthogonal subspace. By choosing $Q_0 \in K$ we can always assume that K contains the identity matrix I_m.

The maximal size of an orthogonal subspace is given by the Radon-Hurwitz number, defined as follows. For $m \in N$, let $m = 2^b \cdot a$, with a odd, and let $b = 4c + d$ where c is a nonnegative integer and $d \in \{0, 1, 2, 3\}$. Then Radon Hurwitz number of m is $\rho(m) := 2^d + 8c$.

Theorem 3 *Let $K \subseteq R^{m \times m}$ be an orthogonal subspace. Then $k := \dim K \leq \rho(m)$, and this inequality is sharp for any $m \in N$. More precisely, assume that $I_m \in K$ and $k \geq 2$. Then K has a basis $I_m, Q_1, \ldots, Q_{k-1}$ where $Q_1, \ldots, Q_{k-1}$ is a set of skew symmetric orthogonal anticommuting matrices, i.e. $Q_i Q_j = -Q_j Q_i$ for any $1 \leq i < j \leq k-1$.*

Conversely, if $Q_1, \ldots, Q_{k-1} \in R^{m \times m}$ are $k-1$ skew symmetric orthogonal anticommuting matrices then $\mathrm{span}(I_m, Q_1, \ldots, Q_{k-1})$ is an k-dimensional orthogonal subspace.

If $Q \in R^{m \times m}$ is an orthogonal matrix, then all m singular values of Q are equal to 1. Let $Q_i \in R^{m_i \times m_i}$ be an orthogonal matrix for $i = 1, 2$. Then for any real a_1, a_2, the singular values of $a_i Q_i$ are $|a_i|$, and the singular values of $(a_1 Q_1) \otimes (a_2 Q_2)$ are all $|a_1 a_2|$.

Suppose furthermore that m_1, m_2 are even and Q_1, Q_2 are skew symmetric orthogonal matrices. Then $a_i Q_i$ has $\frac{m_i}{2}$ eigenvalues equal to $a_i\sqrt{-1}$ and $-a_i\sqrt{-1}$ for $i = 1, 2$ repectively. Furthermore, $(a_1 Q) \otimes (a_2 Q)$ is a real symmetric matrix with $\frac{m_1 m_2}{2}$ eigenvalues equal to $a_1 a_2$ and $\frac{m_1 m_2}{2}$ eigenvalues equal to $-a_1 a_2$.

Theorem 4 *Let $K \subseteq R^{m \times m}$ be an orthogonal subspace. Then $\mathrm{H}(K) = \log m$. Suppose furthermore that m_1, m_2 are even and $K_i \subset R^{m_i \times m_i}$ are orthogonal subspaces of dimension two at least for $i = 1, 2$. Then*

$$\begin{aligned} \mathrm{H}(K_1 \otimes K_2) &\leq \log \frac{m_1 m_2}{2} = \log(m_1 m_2) - \log 2 \\ &= \mathrm{H}(K_1) + \mathrm{H}(K_2) - \log 2 \end{aligned} \tag{B.1}$$

In particular, the additivity conjecture does not hold for real subspaces of matrices.

Proof. Since any matrix $x \in K$ is of the form aQ for some orthogonal Q it follows that if $\mathrm{Tr}(xx^{\top}) = 1$ then the singular values of x are all equal to $\frac{1}{m}$. Hence $\mathrm{H}(x) = \log m$ and $\mathrm{H}(K) = \log m$.

Assume now that K_1, K_2 are orthogonal spaces of dimension two at least. Without loss of generality we may assume that $I_{m_1}, Q_1 \in R^{m_1 \times m_1}$ and $I_{m_2}, Q_2 \in R^{m_2 \times m_2}$, where Q_1, Q_2 are orthogonal. Hence $I_{m_1 m_2} = I_{m_1} \otimes I_{m_2}$ and $Q_1 \otimes Q_2$ are both in $K_1 \otimes K_2$. Recall that $Q_1 \otimes Q_2$ is a symmetric matrix which has $\frac{m_1 m_2}{2}$ eigenvalues equal to 1 and -1 respectively. Hence

$Q_1 \otimes Q_2 + I_{m_1 m_2}$ is a nonnegative definite real symmetric matrices which has $\frac{m_1 m_2}{2}$ eigenvalues equal to 2 and 0 respectively. Let $x = (\frac{2}{m_1 m_2})^{\frac{1}{2}}(Q_1 \otimes Q_2 + I_{m_1 m_2})$. Then $\mathrm{Tr}(xx^\top) = 1$ and x has $\frac{m_1 m_2}{2}$ nonzero singular values all equal to $(\frac{2}{m_1 m_2})^{\frac{1}{2}}$. Hence $\mathrm{H}(K_1 \otimes K_2) \leq \mathrm{H}(x) = \log(\frac{m_1 m_2}{2})$. $\square$