Yet another proof of the joint convexity of relative entropy

Mary Beth Ruskai

Received: 3 January 2022 / Revised: 14 May 2022 / Accepted: 25 June 2022 /
Published online: 12 August 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
The joint convexity of the map \((X, A) \mapsto X^* A^{-1} X\), an integral representation of operator convex functions, and an observation of Ando are used to obtain a simple proof of both the joint convexity of relative entropy and a trace convexity result of Lieb. The latter was the key ingredient in the original proof of the strong subadditivity of quantum entropy.

1 Introduction

In their influential book *Quantum Computation and Quantum Information*, Nielsen and Chuang [17, Appendix 6] assert that “no transparent proof of strong subadditivity” of quantum entropy is known. Since then, there have been a number of simpler proofs, e.g., [4, 23] as well as more transparent expositions [3, 21, 22] of earlier arguments, including Lieb’s key result [12] on the joint concavity of the map \((A, B) \mapsto \text{Tr} K^* A^p K B^{1-p}\) for \(p \in (0, 1)\) and \(A, B \geq 0\) positive semi-definite matrices. Indeed, Simon’s recent book [25] on Loewner’s Theorem includes three different proofs of Lieb’s result!

When Lieb and the author were working on strong subadditivity (SSA) in 1971-72, they proved [16] that the map \((X, A) \mapsto X^* A^{-1} X\) is jointly convex in \((X, A)\) with \(A\) positive.\(^1\) They were able to use this result to prove a number of special cases of SSA, but it seemed insufficient for the general result. Later, the author realized that the strategy used in [16] could also be used to prove related results, e.g., [7, Appendix] and [23, Section 2.4], which lead to simple proofs of SSA.

\(^1\) While refereeing a paper in 2010, the author was surprised to read that “the joint convexity of \((X, A) \mapsto X^* A^{-1} X\) was proved by Kiefer in 1959 and rediscovered by Lieb and Ruskai.” We learned this just in time to add a reference to Kiefer [8, Lemma 3.2] to the final version of [7].
A few years after Lieb’s paper [12], Ando [1, 2] gave two different proofs of Lieb’s key result. After reading a draft of [25], the author realized that Ando’s tensor product approach could be combined with the joint convexity [8, 16] of \( (X, A) \mapsto X^* A^{-1} X \) to prove the joint convexity of relative entropy as well as Lieb’s result [12]. Historically, Lieb’s concavity result was a key step in the first proof of SSA. However, by proving the joint convexity of relative entropy directly [7, 23] one can prove SSA without Lieb’s result. In this note, we present a concise proof of both Lieb’s result and the joint convexity of relative entropy. We conclude by using the latter to give a simple proof of the monotonicity of relative entropy under partial traces and, hence, SSA.

2 Preliminaries

A mixed state for a quantum system associated with a finite dimensional Hilbert space \( \mathcal{H} \) is given by a density matrix, i.e., a positive semi-definite matrix \( \rho \) with \( \text{Tr} \rho = 1 \). The quantum entropy of a state \( \rho \) was defined by von Neumann as \( S(\rho) \equiv -\text{Tr} \rho \log \rho \). Umegaki [26] defined the relative entropy for a pair of states as \( H(\rho, \gamma) \equiv \text{Tr} (\rho \log \rho - \rho \log \gamma) \). Note that the relative entropy is well-defined if \( \ker \gamma \subseteq \ker \rho \) and easily extended to arbitrary matrices \( A, B \geq 0 \). Expositions of these concepts can be found in [17, 18, 28].

A quantum system with two (or more) subsystems is described by a tensor product of two (or more) Hilbert spaces \( \mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2 \). If \( \rho_{12} \) is a density matrix on \( \mathcal{H}_{12} \), then the partial trace gives density matrices \( \rho_1 = \text{Tr}_2 \rho_{12} \) and \( \rho_2 = \text{Tr}_1 \rho_{12} \) on \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \), respectively.

The strong subadditivity (SSA) inequality for a state \( \rho_{123} \) on \( \mathcal{H}_{123} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \) is

\[
S(\rho_{123}) + S(\rho_2) \leq S(\rho_{12}) + S(\rho_{23}) \tag{1}
\]

SSA was conjectured by Lanford and Robinson [10] in 1968 and proved by Lieb and Ruskai [14, 15] in 1973. Equivalent formulations can be found in [15, 18, 21, 28].

Our argument relies heavily on well-known results about operator convex functions. A function \( f : (a, b) \mapsto \mathbb{R} \) is operator monotone if \( A \leq B \Rightarrow f(A) \leq f(B) \) for self-adjoint \( A, B \). It is operator convex if \( f(tA + (1 - t)B) \leq tf(A) + (1 - t)f(B) \) holds as an operator inequality for \( t \in [0, 1] \). A function \( f(x) \) is operator monotone if it can be analytically continued to the upper half plane and maps the upper half plane into the upper half plane [25, Theorem 2.7]. It is operator convex if a suitable difference quotient is operator monotone [25, Theorem 9.1].

**Theorem 1** Define \( g_p(x) \) and \( h_q(x) \) on \( (0, \infty) \) by

\[
g_p(x) \equiv \begin{cases} 
\frac{1}{1-p}(x - x^p) & p \in (0, 1) \cup (1, 2] \\
x \log x & p = 1
\end{cases} \tag{2a}
\]
Then, \( g_p \) and \( h_q \) are operator convex for \( p \in (0, 2] \) and \( q \in [-1, 1) \), respectively.

**Proof** The function \( f(x) = x \log x \) is operator convex because \( \frac{f(x) - f(0)}{x - 0} = \log x \) is operator monotone. Since \( f(x) = x \) is linear, the operator convexity of \( g_p(x) \) depends on the behavior of \( x^p \). Observe that \( x^t \) is operator monotone for \( t \in (0, 1) \) and \( -x^t \) is operator monotone for \( t \in [-1, 0] \) so that \( x^p \) is operator convex for \( p \in (1, 2] \) and \( -x^p \) is operator convex for \( p \in (0, 1) \). Combined with the sign change at \( p = 1 \), this implies that \( g_p(x) \) is operator convex.

Next, observe that \( h_q(x) = xg_{1-q}(x^{-1}) \). When \( g \) is operator convex and \( g(1) = 0 \), then \( \frac{g(x)}{x-1} \) is operator monotone. Since \( x^{-1} \) maps the upper half plane to the lower half plane, \( \frac{xg(x^{-1})}{x-1} = \frac{-g(x)}{1-x^{-1}} \) is operator monotone and \( xg(x^{-1}) \) is operator convex. Thus, \( h_q(x) \) is operator convex for \( q \in [-1, 0) \cup (0, 1) \). \( \square \)

### 3 Joint convexity framework

**Theorem 2** (Kiefer [8, Lemma 3.2] ) Let \( A \) be positive semi-definite with ker \( A \subseteq \ker XX^* \). Then, the map \((X, A) \mapsto X^* A^{-1} X \) is jointly convex in \( X, A \).

**Proof** Kiefer considered \( A_j > 0 \) positive definite. We give an argument based on [16]. Let \( M_j = t_j^{1/2}(A_j^{-1/2}X_j - A_j^{1/2} \Lambda) \) and \( \Lambda = (\sum_j t_j A_j)^{-1}(\sum_j t_j X_j) \) with \( t_j > 0 \) and \( \sum_j t_j = 1 \). Then,

\[
0 \leq \sum_j M_j^* M_j = \sum_j t_j X_j^* A_j^{-1} X_j - (\sum_j t_j X_j^*) \Lambda - \Lambda^* (\sum_j t_j X_j)
+ \Lambda^* (\sum_j t_j A_j) \Lambda
= \sum_j t_j X_j^* A_j^{-1} X_j - (\sum_j t_j X_j^*) (\sum_j t_j A_j)^{-1}(\sum_j t_j X_j)
\]

which proves joint convexity for \( A_j > 0 \). If some \( A_j \geq 0 \) is singular and \( \ker A_j \subseteq \ker X_j X_j^* \), then \( \lim_{\epsilon \to 0} X_j^* (A_j + \epsilon I) X_j \) exists. Thus, one can repeat the argument above with \( A_j \) replaced by \( A_j + \epsilon I \) and take the limit \( \epsilon \to 0 \). \( \square \)

Henceforth, for simplicity, we consider only positive definite matrices \( A, B > 0 \). The reader can readily discern the conditions under which some results extend to positive semi-definite matrices.

**Lemma 3** (Ando [1, 2]) Let \( A, B \) be \( m \times m \) and \( n \times n \) matrices, respectively, and let \( K \) be an \( m \times n \) matrix considered as a vector in \( \mathbb{C}_m \otimes \mathbb{C}_n \). Then,

\[
\text{Tr} K^* A K B = \langle K, (A \otimes B^T) K \rangle .
\]

(4)
Theorem 4 Let $g : (0, \infty) \mapsto \mathbb{R}$ be an operator convex function with $g(1) = 0$. Then, for $A, B > 0$ the map $(A, B) \mapsto (I \otimes B)g(A \otimes B^{-1})$ is jointly convex.

Proof It follows from [25, Example 12.8] that the function $g(x)$ can be written in the form

$$g(x) = b(x - 1) + c(x - 1)^2 + \int_0^\infty \frac{(x - 1)^2}{x + s}d\mu(s)$$

(5)

with $c \geq 0$ and $\mu(s)$ a positive measure on $(0, \infty)$ with $\int_0^\infty \frac{1}{1+s}d\mu(s) < \infty$. Then,

$$(I \otimes B)g(A \otimes B^{-1}) = b(A \otimes I - I \otimes B) + c(A \otimes I - I \otimes B) \frac{1}{I \otimes B}(A \otimes I - I \otimes B) + \int_0^\infty (A \otimes I - I \otimes B) \frac{1}{A \otimes I + sI \otimes B}(A \otimes I - I \otimes B) d\mu(s)$$

(6)

The first term is linear in $A, B$; and the remaining terms are jointly convex in $A, B$ by Theorem 2. Since $c \geq 0$ and $\mu$ is a positive measure, the result follows. \qed

Remark $B \mapsto B^T$ is linear, $B \mapsto \overline{B}$ is affine and $B > 0$ implies $B^T = \overline{B}$. Therefore, $(A, B) \mapsto (I \otimes \overline{B})g(A \otimes \overline{B}^{-1})$ is also jointly convex in $(A, B)$. Thus, in the applications which follow, one can replace $B^T$ by $B$ on the right in (4).
convexity also holds for

$$(A, B) \mapsto \frac{1}{q} \operatorname{Tr} \left( K^* KB - K^* A^q KB^{1-q} \right).$$

(d) The map $(A, B) \mapsto -(I \otimes B) \log(A \otimes B^{-1}) = (I \otimes B)(I \otimes \log B - \log A \otimes I)$

is jointly convex in $A, B > 0$ which implies that joint convexity also holds for

$$(A, B) \mapsto \operatorname{Tr} \left( K^* K (B \log B) - K^* (\log A) K B \right).$$

5 Discussion

Note that (a) and (c) imply that $1/p \left( 1 - p \right) \operatorname{Tr} K^* A^p K B^{1-p}$ is jointly concave in $A, B > 0$ for all $p \in [-1, 2]$ (extended by continuity at $p = 0, 1$). Thus, $p \in (0, 1)$ gives Theorem 1 in [12] when $r = 1 - p$. Ando [1, 2] also showed that $\operatorname{Tr} K^* A^p K B^{1-p}$ is jointly convex for $p \in [1, 2]$. Hasegawa [5] seems to have been the first to realize that Lieb and Ando’s results can be extended to all $p \in [-1, 2]$ in the form stated here, as also observed in [7].

The choice $K = I$ in (b) gives the joint convexity of the relative entropy $H(A, B)$, while $K = I$ in (d) gives the joint convexity of $H(B, A)$. More generally, $(A \otimes I) g_p(A^{-1} \otimes B) = (I \otimes B) h_{1-p}(A \otimes B^{-1})$ so that (c) and (d) are somewhat redundant. However, using both $g_p(x)$ and $h_q(x)$ make clear how the results vary for subintervals $[0, 1), (-1, 0), (1, 2]$.

The linear term $\operatorname{Tr} K^* A K$ in (a) arises because $g_p(x)$ was defined so that $\lim_{p \to 1} g_p(x) = x \log x$. Moreover, the Wigner–Yanase–Dyson (WYD) entropy [29] for a density matrix $\gamma$ and $K = K^*$ was defined as $\frac{1}{2} \operatorname{Tr} [K, \gamma^p][K, \gamma^{1-p}] = \operatorname{Tr} K^{1/2} K K^* K^{1/2} - \operatorname{Tr} K^* K$, which contains a linear term. Wigner and Yanase proved concavity in $\gamma$ at $p = \frac{1}{2}$, and Dyson suggested the extension to $p \in (0, 1)$. After dropping the linear term, Lieb [13] proved a generalization of the WYD conjecture as described above. The expressions in (a) and (c) can be regarded as a WYD version of relative entropy.

In most applications, one chooses $K = I$. However, Kim [9] showed 2 that by using a judicious choice of $K$ he could strengthen some related inequalities to operator inequalities on one subspace of a tensor product. Using $K = I_1 \otimes I_2 \otimes |\phi\rangle \langle \phi|$ with $\phi \in \mathcal{H}_3$, he proved that

$$\operatorname{Tr}_{12} \rho_{123} \left[ \log \rho_{123} - \log \rho_{12} - \log \rho_{23} + \log \rho_{2} \right] \geq 0$$

holds as an operator inequality on $\mathcal{H}_3$. For examples related to (a) and (c) see [24].

---

2 Kim’s paper [9] was posted on arxiv.org very close the 40th anniversary of the actual completion of the proof of SSA in October, 1972.
6 Generalizations of relative entropy

The examples above are special cases of a generalization of relative entropy [6, 7, 11] introduced by Petz [18, 19] and sometimes called quasi-entropies or $f$-divergences. These generalizations were defined using the relative modular operator $L_AR^{-1}_B$ where $L_A(X) = AX$ and $R_B(X) = XB$ denote left and right multiplication. Let $\mathcal{G} = \{g : (0, \infty) \mapsto \mathbb{R} : g \text{ is operator convex and } g(1) = 0\}$. Then, for any $g \in \mathcal{G}$ and $A, B > 0$, define $H_g(K, A, B) = \text{Tr} K^* g(L_AR^{-1}_B)R_B K$. It follows from Lemma 3 that

$$H_g(K, A, B) = \langle K, (I \otimes B) g[(A \otimes I)(I \otimes B^{-1})]K \rangle.$$  \hfill (8)

Since $g(x)$ has an integral representation as in (5), it follows immediately from Theorem 4 that the map $(A, B) \mapsto H_g(K, A, B)$ is jointly convex for all $g \in \mathcal{G}$.

Ando proved the joint concavity of $(A, B) \mapsto (I \otimes B) f[(A \otimes I)(I \otimes B^{-1})]$ when $f$ is a positive monotone operator function on $(0, \infty)$ in [2, Theorem 6] and used this to prove Lieb’s concavity result. However, $f \notin \mathcal{G}$ because $f(1) \neq 0$. He then [2, Theorem 7] extended his result to arbitrary monotone operator functions on $(0, \infty)$ and, in particular [2, Corollary 7.1], to $f(x) = \log x$ which implies the joint convexity of relative entropy.

The argument used in the proof of Theorem 1 to show that $h_g(x)$ is operator convex can also be used to show that $\tilde{g}(x) = xg(x^{-1})$ is operator convex for all $g \in \mathcal{G}$. Then, $H_{\tilde{g}}(K, A, B) = H_g(K, B, A)$. Moreover, the function $k(x) = [g(x) + xg(x^{-1})](x - 1)^{-2}$ is well-defined, operator convex, and operator monotone decreasing [6]. Petz [20] showed that there is a one-to-one correspondence between $k(x)$ and Riemannian metrics which generalize the classical Fisher-Rao information and decrease under the action of completely positive, trace-preserving maps which describe quantum channels. (See also [10, Theorem II.14].)

7 Proof of strong subadditivity

It is now well-known [3, 15, 21, 25, 27, 28] that these convexity results can be used in a number of different ways to prove strong subadditivity of quantum entropy, as well as the fact that relative entropy decreases under partial traces and under the action of quantum channels, i.e., completely positive trace-preserving maps. For completeness, we conclude by showing that joint convexity of $H(\rho, \gamma)$ implies monotonicity under partial traces and SSA.

To prove $H(\rho_1, \gamma_1) \leq H(\rho_{12}, \gamma_{12})$, we use the Weyl–Heisenberg matrices $X$ and $Z$ in $M_d$ with elements $x_{jk} = \delta_{j+1,k}$ (with $j + 1 \mod d$) and $z_{jk} = \omega^j \delta_{jk}$ where $\omega = e^{2\pi i/d}$. Then $X$ and $Z$ are unitary and for any matrix $P \in M_d$,

$$\sum_{m, n} X^m Z^n P Z^{-n} X^{-m} = d \text{ (Tr } P \text{) } I.$$  \hfill (9)
Now, let $U_{mn} = I_1 \otimes X^m Z^n$ where $X, Z$ act on $\mathcal{H}_2$. Then $U_{mn}$ is a unitary matrix acting on $\mathcal{H}_{12}$ and $U_{mn}^* = I_1 \otimes Z^{-n} X^{-m}$. It follows from (9) that $\frac{1}{d_2^2} \sum_{m,n} U_{mn} P_{12} U_{mn}^* = \frac{1}{d_2} P_1 \otimes I_2$. Since $H(\rho, \gamma)$ is invariant under unitary conjugations, its joint convexity implies

$$H(\rho_{12}, \gamma_{12}) = \sum_{m,n} \frac{1}{d_2^2} H(U_{mn} \rho_{12} U_{mn}^*, U_{mn} \gamma_{12} U_{mn}^*) \geq H\left(\sum_{m,n} \frac{1}{d_2^2} U_{mn} \rho_{12} U_{mn}^*, \sum_{m,n} \frac{1}{d_2^2} U_{mn} \gamma_{12} U_{mn}^*\right) = H\left(\rho_1 \otimes \frac{1}{d_2} I_2, \gamma_1 \otimes \frac{1}{d_2} I_2\right) = H(\rho_1, \gamma_1). \quad (10)$$

Then one can prove SSA by observing that

$$S(\rho_2) - S(\rho_{12}) = H(\rho_{12}, \frac{1}{d_1} I_1 \otimes \rho_2) + \log d_1 \leq H(\rho_{123}, \frac{1}{d_1} I_1 \otimes \rho_{23}) + \log d_1 = S(\rho_{23}) - S(\rho_{123}) \quad (11)$$

which is equivalent to (1).

**Acknowledgements**  It is a pleasure to thank Professor Barry Simon for sending the author a draft of [25] which reawakened the author’s awareness of old results.

**Declarations** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

**Conflicts of Interest** The author has no conflicts of interest related to this manuscript.

**References**

1. Ando, T.: Topics on Operator Inequalities *Sapporo Lecture Notes* (1978)
2. Ando, T.: Concavity of certain maps on positive definite matrices and applications to Hadamard products. Linear Algebra Appl. **26**, 203–241 (1979)
3. Carlen, E.: Trace inequalities and quantum entropy: an introductory course *Contemp. Math.* **529**, 73–140 (AMS, 2010)
4. Effros, E.G.: “A matrix convexity approach to some celebrated quantum inequalities”. Proc. Natl. Acad. Sci. U.S.A. **106**, 1006–1008 (2009)
5. Hasegawa, H.: $\alpha$ -divergence of the noncommutative information geometry” Rep. Math. Phys. **33**, 87–93 (1993)
6. Hiai, F., Ruskai, M.B.: Contraction coefficients for noisy quantum channels. J. Math. Phys. **57**, 015211 (2016)
7. Jenčová, A., Ruskai, M.B.: A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality. Rev. Math. Phys. **22**, 1099–1121 (2010)
8. Kiefer, J.: Optimum experimental designs. J. Roy. Statist. Soc. Ser. B **21**, 272–310 (1959)
9. Kim, I.: Operator extension of strong subadditivity of entropy. J. Math. Phys. **53**, 122204 (2012)
10. Lanford, O., III., Robinson, D.W.: Mean entropy of states in quantum statistical mechanics. J. Math. Phys. **9**, 1120–1124 (1968)
11. Lesniewski, A., Ruskai, M.B.: Monotone Riemannian metrics and relative entropy on noncommutative probability spaces. J. Math. Phys. **40**, 5702–5724 (1999)
12. Lieb, E.H.: Convex trace functions and the Wigner-Yanase-Dyson conjecture. Adv. Math. 11, 267–288 (1973)
13. Inequalities: Selecta of E. Lieb, Loss, M., Ruskai, M.B. eds (Springer, 2002)
14. Lieb, E., Ruskai, M.B.: A fundamental property of quantum-mechanical entropy. Phys. Rev. Lett. 30, 434–436 (1973)
15. Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum-mechanical entropy. J. Math. Phys. 14, 1938–1941 (1973)
16. Lieb, E.H., Ruskai, M.B.: Some operator inequalities of the Schwarz type. Adv. Math. 12, 269–273 (1974)
17. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information (Cambridge University Press, 2000)
18. Ohya, M., Petz, D.: Quantum Entropy and Its Use (Springer, 1993)
19. Petz, D.: Quasi-entropies for finite quantum systems. Rep. Math. Phys. 23, 57–65 (1986)
20. Petz, D.: Monotone metrics on matrix spaces. Linear Algebra Appl. 244, 81–96 (1996)
21. Ruskai, M. B.: “Inequalities for quantum entropy: a review with conditions for equality” J. Math. Phys. 43, 4358–4375 (2002); erratum 46, 019901 (2005)
22. Ruskai, M.B.: “Lieb’s simple proof of concavity of $\text{Tr} K^pA^pB^1−p$ and remarks on related inequalities” Int. J. Quantum Inf. 3, 579–590 (2005); erratum 4, 747 (2006)
23. Ruskai, M.B.: Another short and elementary proof of strong subadditivity of quantum entropy. Rep. Math. Phys. 60, 1–12 (2007)
24. Ruskai, M.B.: Remarks on Kim’s strong subadditivity matrix inequality: extensions and equality conditions”. J. Math. Phys. 54, 102202 (2013)
25. Simon, B.: Loewner’s Theorem on Monotone Matrix Functions (Springer, 2019)
26. Umegaki, H.: Conditional expectation in an operator algebra. IV. entropy and information. Ködai Math. Sem. Rep. 14, 59–85 (1962)
27. Vershynina, A., Carlen, E., Lieb, E.: Strong subadditivity of quantum entropy. Scholarpedia 8, 30920 (2013)
28. Wehrl, A.: General properties of entropy. Rev. Mod. Phys. 50, 221–260 (1978)
29. Wigner, E.P., Yanase, M.M.: Information content of distributions. Proc. Nat. Acad. Sci. USA 49, 910–918 (1963)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.