Letters in Mathematical Physics (2023) 113:18
https://doi.org/10.1007/s11005-023-01638-2

Ruminations on matrix convexity and the strong subadditivity of quantum entropy

Michael Aizenman[1,2] · **Giorgio Cipolloni**[3]

To E.H. Lieb and M.B. Ruskai, in celebration of their 1973 proof of the SSA.

Received: 18 October 2022 / Revised: 16 January 2023 / Accepted: 17 January 2023 /
Published online: 3 February 2023

Abstract
The familiar second derivative test for convexity, combined with resolvent calculus, is shown to yield a useful tool for the study of convex matrix-valued functions. We demonstrate the applicability of this approach on a number of theorems in this field. These include convexity principles which play an essential role in the Lieb–Ruskai proof of the strong subadditivity of quantum entropy.

Keywords Matrix convexity · Quantum entropy · Strong subadditivity · Parallel sums

MSC codes 81Q05 · 82B10 · 26B25 · 47H05

1 Introduction

1.1 A bit of background

A real-valued function $f : (a, b) \to \mathbb{R}$ is said to be convex on n-matrices if and only if for any pair of self-adjoint $n \times n$ matrices A_0, A_1 with spectrum in (a, b), and for any $\lambda \in (0, 1)$ the following holds

$$f((1-\lambda)A_0 + \lambda A_1) \leq (1-\lambda) f(A_0) + \lambda f(A_1) \tag{1.1}$$

✉ Michael Aizenman
aizenman@princeton.edu

Giorgio Cipolloni
gc4233@princeton.edu

1 Departments of Physics and Mathematics, Princeton University, Princeton, NJ 08544, USA

2 Weizmann Institute of Science, Rehovot, Israel

3 Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544, USA

in the sense of quadratic forms. We denote by $\mathcal{B}_n(a,b)$ the (convex) set of self-adjoint matrices with spectrum in (a,b) and, adapting the notation of [21], by $\mathcal{C}_n(a,b)$ the class of the corresponding n-matrix convex functions, with $\mathcal{C}_\infty(a,b) := \cap_{n=1}^{\infty}\mathcal{C}_n(a,b)$. The function f is called concave if $-f$ is convex.

Some of the familiar implications of convexity of a real-valued functions extend directly to its matrix version. Among those is the Jensen inequality that states that for any function f which is convex over a convex domain D, and any probability measure $\rho(dM)$ of a compact support within D

$$\int f(M)\rho(dM) \geq f\left(\int M\,\rho(dM)\right). \tag{1.2}$$

However, the matrix version of convexity turns to be a far more confining notion. In particular, even for $n=2$ only quadratic functions meet the defining condition with no restriction on the self-adjoint matrices' spectrum [9]. A similar phenomenon holds for monotonicity, as was discovered earlier by Charles Loewner for matrix monotone functions. Among Loewner's remarkable contributions is the realization that upon limiting the requirement of matrix-monotonicity to matrices of spectrum confined to a subinterval of $\mathbb{R}$ one gets a richer class of functions. He showed that any f that is monotone over matrices with spectrum in (a,b) is analytic over (a,b), has an analytic extension off the real line with the Herglotz property, and admits a corresponding integral representation [19] (cf. [21]).

Building on Loewner's theory, Kraus [11] and later Bendat and Sherman [2] have developed a related characterization of matrix convex functions. Its explicit statement is found below in Theorem 3.2 which presents the easy half of the statement. Fundamental role in their studies was played by the monotonicity and implied properties, of Loewner's *divided difference* matrices $L=(L_{i,j})_{i,j=1}^k$, of matrix elements

$$L_{i,j} = \frac{f(x_i)-f(x_j)}{x_i-x_j} \quad \text{(interpreted as } f'(x_i) \text{ in case } i=j\text{)}, \tag{1.3}$$

associated with collections of sites $x_1 < x_2 < \cdots < x_k \subset (a,b)$. The link between the two notions established in [2, 11] is that f is matrix convex if and only if for all (equivalently for some [23]) $y \in (a,b)$ the function $g(x) := (f(x)-f(y))/(x-y)$ is matrix monotone over (a,b).

Aside from the analytical challenges described above, convexity has been of interest as a key tool for a myriad of variational problems. It is then natural that matrix convexity drew attention also in the theory of entropy functions of quantum statistical mechanics and quantum information theory, and somewhat independently in discussions of matrix networks. For such applications of convexity, the more relevant challenge has been to establish convexity statement for specific functions of interest.

Examples of results whose development was driven by such applications are mentioned below. Among those is a theorem of Lieb [14], which proved the Wigner–Yanase–Dyson conjecture [25] and played a key role in the Lieb–Ruskai proof of strong subadditivity of entropy [16, 17], a property commonly referred to as SSA. From another direction, matrix convexity showed up in the work of Anderson and

Duffin [1] on parallel sums (equivalently harmonic means), an extension of which can be found in the Kubo–Ando [12] theory of operator means[1]

Given how rich is the theory related to the above, we refer the reader for further details to the recent and thorough book by Simon [21].

1.2 Outline of the paper's contents

Our ruminations on matrix convexity start with the discussion of a local criterion for concavity, and show how it allows short proofs of some of the known theorems, which were originally derived by other means. These include concavity of parallel sums, and from that a statement known as Lieb concavity, that plays a key role in different proof of the above-mentioned properties of quantum entropy functionals. That discussion does not add to the proof plan which was laid out in the original work of Lieb and Ruskai [17], beyond what may perhaps be received as a simplification in the derivation of the enabling results.

The proof of SSA presented here passes through the Lieb–Ruskai concavity of the conditional entropy of a composite system as function of the state. More explicitly, in the notation that is explained in Sect. 5, $S(1|2) \equiv S(\rho_{12}) - S(\rho_2)$ is a concave function of the state operator ρ—a statement which is also of independent interest [7].

2 Local test of convexity

While some convexity relations are perhaps better grasped through convexity's non local expressions, such as the Jensen inequality (1.2) and its non-abelian extensions (cf. [5, 8]), a well-known sufficient condition for it is the positivity of the second derivative (employed in different ways in [3, 9, 11, 14]). Here we shall employ the following simply stated version of such a criterion for matrix valued functions.

Proposition 2.1 *A sufficient condition for $f : (a, b) \to \mathbb{R}$ to be in $\mathcal{C}_n(a, b)$ is that for any matrix $M \in \mathcal{B}_n(a, b)$ and any bounded self-adjoint Q of equal rank the matrix-valued function $f(M + tQ)$ is twice differentiable at $t = 0$ and satisfies*

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f\big(M + tQ\big)\Big|_{t=0} \geq 0 \tag{2.1}$$

in the sense of quadratic forms.

Proof First one may note (by considering the case $Q = \mathbb{1}$) that the assumed condition requires $f(x)$, as a function of a real variable, to be twice differentiable. Taking that as granted, the passage from the local condition (2.1) to the global (1.1) (the defining

[1] Some of these developments have proceeded on parallel tracks: the paper of Anderson and Duffin [1], where matrix concavity of parallel sums is noted and proven, has neither cited the earlier theory nor was its value recognized in early works on quantum entropy. A contributing factor may have been that convexity was mentioned in [1] only in passing, in one (#24) of many theorems, and was not mentioned in the work's summary.

condition of matrix convexity) can be explained by the observation that for any given A_0, A_1 and $\lambda \in (0, 1)$, and any twice differentiable function f

$$[(1-\lambda)f(A_0)+\lambda f(A_1)] - f(A_\lambda) = \int_0^1 K_\lambda(t)\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(A_t)\,\mathrm{d}t, \tag{2.2}$$

with $A_\lambda = (1-\lambda)A_0 + \lambda A_1 \equiv A_0 + \lambda(A_1 - A_0)$ and

$$K_\lambda(t) = \begin{cases} (1-\lambda)t & 0 \le t \le \lambda \\ (1-t)\lambda & \lambda \le t \le 1 \\ 0 & t \in \mathbb{R} \setminus [0,1] \end{cases}. \tag{2.3}$$

The relation (2.2) is implied through integration by parts, and the observation that being continuous with piecewise constant derivative the function $K_\lambda(t)$ satisfies

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} K_\lambda(t) \;=\; (1-\lambda)\delta(t) - \;\delta(t-\lambda) + \lambda\delta(t-1) \tag{2.4}$$

in the distributional sense, and vanishes beyond (0, 1). (To simplify the integration by parts it is convenient to extend the integral to $[-a, 1+a]$ at an arbitrary $a > 0$.) □

Likewise, to conclude the joint convexity of $f(A_1, \ldots, A_k)$ it suffices to establish the analogous version of (2.1) for function $f((M_1, \ldots, M_k) \;+\; t\,(Q_1, \ldots, Q_k))$. In this case the condition should be verified for arbitrary k-tuples $(M_1, \ldots, M_k)$ within the relevant convex domain and for arbitrary k-tuples of bounded self-adjoint matrices Q_j.

As an alert to a subtlety let us note that while $f(\cdot)$ is a function of a single variable, when that variable is a matrix the domain's tangent space at specified M is n^2-dimensional. In (2.1), the multivariate nature of the local test shows up in the requirement that the second derivative's positivity (in the quadratic form sense) holds for all directions of the linear explorations (expressed by Q).

An alternative approach to the local positivity condition is to express it through the multivariate second derivative of the mapping $M \to f(M)$, as is explained and applied in [9] (in terms of a generalized Hessian and Fréchet differentials). In comparison, the condition presented in (2.1) is simpler to state. It also facilitates the analysis when the computation of the second derivative in t yields a recognizably positive algebraic expression. Examples of that are found among the functions which were successfully analyzed in [6, 10, 14], as well as those discussed here.

3 Computing with resolvents

The differentiation of functions of the form $f(A(t))$ with $A(t)$ varying smoothly over $\mathcal{B}_n(a, b)$ is complicated by the non-commutativity of the matrix product. More explicitly: even if f is a smooth function over $\mathbb{R}$, its value over matrices in the general case is defined through the spectral representation of the time-dependent $A(t)$, which for different times is diagonalized at different frames. Nevertheless, the test is manageable for a number of relevant classes of functions. As was noted in [6], in addition to the trivial case of quadratic polynomials, simple cases include $f(x) = 1/(x-u)$ with $u \in \mathbb{R}\backslash(a,b)$, which produces resolvent operators. Through convex combinations, such functions are the building blocks of Herglotz/Pick functions.

Lemma 3.1 *For any $I = (a,b) \subset R$ and $u \in \mathbb{R} \setminus I$, the following function is in $\mathcal{C}_\infty(a,b)$*

$$f_u(z) = \begin{cases} \frac{1}{z-u} & u < a \\ \frac{1}{u-z} & u > b \end{cases} =: \frac{sgn_{(a,b)}(u)}{u-z}. \tag{3.1}$$

Proof Applied to matrices, $f_u(A) = \pm(u\mathbb{1} - A)^{-1}$ corresponds to the resolvent operator. In this case, the differential calculus is easy to manage through the resolvent expansion. To streamline its presentation, we denote (locally within this proof)

$$R(t) := \frac{1}{u\mathbb{1} - A(t)}, \tag{3.2}$$

with $A(t) := A + tQ$.

Differentiation of the resolvent to any order is facilitated by the resolvent identity:

$$\frac{1}{A+\Delta A} - \frac{1}{A} = -\frac{1}{A}\,\Delta A\,\frac{1}{A+\Delta A}. \tag{3.3}$$

This yields the exact and simple expression

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\,R(t) = 2R(t)\;Q\;R(t)\;QR(t) = X^*(t)\;X(t), \tag{3.4}$$

with $X(t) := \sqrt{R(t)}\;QR(t)$. The claim then follows by the clear positivity of (3.4) together with Proposition 2.1. □

Theorem 3.2 *For any $-\infty \le a < b \le +\infty$, if a function $f : (a,b) \to \mathbb{R}$ admits a representation of the form*

$$f(z) = \alpha + \beta z + \gamma z^2 + \int_{-\infty}^{a} \frac{(z-c)(1+uz)}{u-z}\mu(\mathrm{d}u) + \int_{b}^{\infty} \frac{(c-z)(1+uz)}{z-u}\mu(\mathrm{d}u), \tag{3.5}$$

with $\alpha, \beta \in \mathbb{R}$, $\gamma \ge 0$, $c \in (a,b)$, and μ a finite positive measure on $\mathbb{R} \setminus (a,b)$, then f is in $\mathcal{C}_\infty(a,b)$.

Proof A function of the form (3.5), when restricted to a compact subset of (a, b), is an integral over a bounded measure of uniformly bounded terms, corresponding to the following elementary functions (of z)

$$\begin{aligned}\frac{(z-c)(1+uz)}{u-z} &= \frac{(u-c)(1+u^2)}{u-z} - uz + uc - (1+u^2) \\ &= (1+u^2)(u-c)\mathrm{sgn}_{(a,b)}(u)f_u(z) + -uz + uc - (1+u^2).\end{aligned} \tag{3.6}$$

By Lemma 3.1 $f_u(A(t))$ has positive second derivative, and the same holds for their affine combination. Thus $f \in \mathcal{C}_\infty(a, b)$. □

The condition (3.5) assumed here is not unnatural since by the theorem of Kraus–Bendat–Sherman–Uchiyama [2, 11, 23] any function f in $\mathcal{C}_\infty(a, b)$ admits such a representation (cf. [21]). Thus, while Theorem 3.2 is one of the easiest statements to prove in the existent theory of matrix convexity, ipso-facto, it covers all $\mathcal{C}_\infty(a, b)$. It should be added that the collection of functions for which the local test is to some extent manageable includes also the exponential function $f(x) = e^{\alpha x}$, though in this case the resulting algebra is a bit less elementary when the real variable x is replaced by a matrix M. The exponential case played an important role in Lieb's original concavity work [14, Corollary 6.1]. An alternative proof based on the Herglotz / Pick representation was presented in the subsequent work of Epstein [6] and its extension by Hiai [10].

Proofs of matrix convexity through Herglotz condition, or resolvent analysis, go back to early works on the subject, including those quoted above. Yet some novelty may potentially be found in our use of resolvent analysis for simple proofs of *joint concavity* in more than one matrix variable, which is presented next.

4 Resolvent-based proofs of joint concavity in multiple matrix variables

4.1 Joint concavity of the parallel sums

Turning to specific statements of interest, we next present a simple local proof of the concavity of the *parallel sums* (so-called since the sum coincides with the addition rule for resistors connected in parallel).[2] This statement was formulated and proved for $k = 2$ by Anderson and Duffin [1] using an algebraic argument. The proof of its more general version, using the second derivative criterion, is presented here as a demonstration of the resolvent-based approach.

Theorem 4.1 *The mappping of k-tuples of strictly positive matrices* $(A_1, \ldots, A_k)$

$$(A_1, \ldots, A_k) \mapsto \frac{1}{A_1^{-1} + \cdots + A_k^{-1}} \tag{4.1}$$

[2] The parallel sum equals half of their harmonic average, and as such is one of a number of interesting examples of operator means [20].

is jointly concave.

Proof The proof proceeds by verifying the second derivative criterion for the parallel sum in (4.1) with $A_j(t) = A_j + tQ_j$. To shorten the relevant expressions, we denote (for use within this proof)

$$R(t) := A_1(t)^{-1} + \cdots + A_k(t)^{-1}, \tag{4.2}$$

and let $R'(t) := \mathrm{d}R(t)/\mathrm{d}t$, $R''(t) := \mathrm{d}^2R(t)/\mathrm{d}t^2$. The existent t–dependence will be omitted in lengthier displayed equations.

From the resolvent identity (3.3), one gets

$$R' = -\sum_j \frac{1}{A_j} Q_j \frac{1}{A_j}, \qquad R'' = 2\sum_j \frac{1}{A_j} Q_j \frac{1}{A_j} Q_j \frac{1}{A_j}. \tag{4.3}$$

We thus obtain

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2}\frac{1}{R} &= 2\,\frac{1}{R}\,R'\,\frac{1}{R}\,R'\,\frac{1}{R} \;-\; \frac{1}{R}\,R''\,\frac{1}{R}\\
&= -2\left(\sum_j \frac{1}{R}\,\frac{1}{A_j}\,Q_j\,\frac{1}{A_j}\,Q_j\,\frac{1}{A_j}\,\frac{1}{R} - \sum_{j,m}\frac{1}{R}\,\frac{1}{A_j}\,Q_j\,\frac{1}{A_j}\,\frac{1}{R}\,\frac{1}{A_m}\,Q_m\,\frac{1}{A_m}\,\frac{1}{R}\right)\\
&= -2\sum_{j,m} Y_j^*\left(\delta_{j,m} - T_{jm}\right)Y_m,
\end{aligned} \tag{4.4}$$

where

$$Y_j := \frac{1}{\sqrt{A_j}} Q_j \frac{1}{A_j}\frac{1}{R}, \qquad T_{j,m} := \frac{1}{\sqrt{A_j}}\frac{1}{R}\frac{1}{\sqrt{A_m}}. \tag{4.5}$$

Viewing $T = [T_{j,m}]$ as a matrix of operators (matrices), we note that it has the self-adjoint projection properties:

$$T^\dagger = T, \quad T^2 = T \quad (\text{in the sense that} \sum_k T_{j,k}\, T_{k,m} = T_{j,m}). \tag{4.6}$$

Hence $\mathbb{1} - T = [\mathbb{1} - T]^\dagger\,[\mathbb{1} - T]$. It follows that the expression in the last line of (4.4) is a negative matrix, and thus the concavity criterion of Proposition 2.1 is met. □

4.2 Lieb concavity

From the concavity of parallel sums, one may deduce the following result on the joint concavity of powers' tensor products. This statement forms a particular case of Corollary 6.1 in Lieb's [14]. In [10], one finds such a result proven using Pick function analysis (Corollary 2.2 there), and also a more complete discussion of statements in this vein.

Theorem 4.2 *For any real numbers $p_1, \dots, p_k$ obeying*

$$p_j \geq 0, \qquad \sum_{j=1}^{k} p_j \leq 1, \tag{4.7}$$

the mapping

$$(A_1, A_2, \dots, A_k) \mapsto A_1{}^{p_1} \otimes A_2{}^{p_2} \otimes \cdots \otimes A_k{}^{p_k} \tag{4.8}$$

is jointly concave on k-tuples of strictly positive matrices.

Proof It is convenient to present the tensor product as a regular product of commuting operators,

$$A_1{}^{p_1} \otimes A_2{}^{p_2} \otimes \cdots \otimes A_k{}^{p_k} = \widetilde{A}_1{}^{p_1} \cdot \widetilde{A}_2{}^{p_2} \cdot \cdots \cdot \widetilde{A}_k{}^{p_k} \tag{4.9}$$

with $\widetilde{A}_j := \mathbb{1} \otimes \cdots \otimes \mathbb{1} \otimes A_j \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1}$.

Let us first prove the claim for sequences of strictly positive numbers with $\sum_{j=1}^{k} p_j = 1$, $k \geq 2$. In that case, for any collection of commuting matrices $\widetilde{A}_j$

$$\begin{aligned} A_1{}^{p_1} \otimes A_2{}^{p_2} \otimes \cdots \otimes A_k{}^{p_k} &= \widetilde{A}_1{}^{p_1} \cdot \ldots \cdot \widetilde{A}_k{}^{p_k} \\ &= C_k(p_1, p_2, \dots, p_k)^{-1} \int_0^\infty \cdots \int_0^\infty \left(\frac{1}{\widetilde{A}_1} + \frac{u_2}{\widetilde{A}_2} + \cdots + \frac{u_k}{\widetilde{A}_k} \right)^{-1} \prod_{j=2}^{k} \left(u_j^{p_j} \frac{\mathrm{d}u_j}{u_j} \right). \end{aligned} \tag{4.10}$$

with the finite constant

$$C_k(p_1, \dots, p_k) := \int_0^\infty \cdots \int_0^\infty \frac{1}{1 + u_2 + \cdots + u_m} \prod_{j=2}^{k} \left(u_j^{p_j} \frac{\mathrm{d}u_j}{u_j} \right). \tag{4.11}$$

The convergence of the above integral (which is suggested by simple power counting) can be seen by transforming it and changing the variables of integration to $v_j := t\, u_j$ as follows:

$$\begin{aligned} C_k(p_1, \dots, p_k) &= \int_{\mathbb{R}_+^k} e^{-t(1+u_2+\cdots+u_k)} \prod_{j=2}^{k} \left(u_j^{p_j} \frac{\mathrm{d}u_j}{u_j} \right) \mathrm{d}t \\ &= \prod_{j=1}^{k} \left(\int_0^\infty e^{-v_j} v_j^{p_j} \frac{\mathrm{d}v_j}{v_j} \right) < \infty. \end{aligned} \tag{4.12}$$

By Theorem 4.1, for each $\{u_1, \dots, u_k\}$ the integrand in (4.10) is a jointly concave function of $\{A_1, \dots, A_k\}$. Combined with the uniform convergence of the integral over compact sets of k-tuples of A_j, this allows to deduce the concavity of the tensor product which is on the left side of (4.10).

The statement's extension to $\sum_{j=1}^{k} p_j < 1$ can be deduced from the case $k+1$ with $\sum_{j=1}^{k+1} p_j = 1$, upon setting $\widetilde{A}_{k+1} \equiv \mathbb{1}$. Once that is established, the case where some $p_j = 0$ easily follows from the validity of this statement for $\sum_{j=1}^{k} p_j < 1$ with strictly positive j, at smaller values of k. □

In addition to its intrinsic value, Theorem 4.2 is of interest as a gateway (or one of such) to the following theorem of E.H. Lieb, for which an elementary proof is presented below.

Corollary 4.3 (Lieb [14]) *For each fixed matrix K, and $p, r \in [0, 1]$ with $p + r \leq 1$, the following function of two matrix variables*

$$(A, B) \mapsto \mathrm{Tr}\big[A^p K^* B^r K\big] \tag{4.13}$$

is jointly concave over $\mathcal{B}_n(0, \infty) \times \mathcal{B}_n(0, \infty)$.

Proof The latter follows from the $k = 2$ case of Theorem 4.2 through the observation that for any $K : \mathcal{H}_1 \mapsto \mathcal{H}_2$ there exists a vector $|\mathcal{K}\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ (in Dirac's notation) such that for all $A : \mathcal{H}_1 \mapsto \mathcal{H}_1$ and $B : \mathcal{H}_2 \mapsto \mathcal{H}_2$

$$\mathrm{Tr}[A^p K^* (B^t)^r K\big] = \langle \mathcal{K} | A^p \otimes B^r | \mathcal{K} \rangle \, . \tag{4.14}$$

□

An immediate implication of Lieb concavity, Corollary 4.3, and the context in which the result appeared first, was Lieb's proof of the Wigner–Yanase–Dyson conjecture [14]. It states that the Wigner–Yanase [25] "p-skew information" of a density matrix ρ, with respect to an operator K, is concave in ρ. This quantity is defined as

$$I_p(\rho, K) := \frac{1}{2} \mathrm{Tr}[\rho^p, K^*][\rho^{1-p}, K] \; \Big(= \mathrm{Tr} K \rho^p K \rho^{1-p} - Tr K \rho K\Big) , \tag{4.15}$$

where $[A, B] := AB - BA$ denotes the commutator. Its convexity as a function of ρ follows readily from Corollary 4.3 with $A = B = \rho$ and $r = 1 - p$.

4.3 The Kubo–Ando theorem

Theorem 4.2 yields also a streamlined proof of the following theorem of Kubo and Ando [12]. A simplification of the original derivation was presented by Effros [5] using the Hansen–Pedersen Jensen operator inequality [8]. The proof given below takes a short-cut through Loewner's classification theorem, but given that it may seem shorter.

Theorem 4.4 (Kubo-Ando [12]) *If $f : (0,\infty) \to \mathbb{R}$ is operator monotone, then for any n the function*

$$(A, B) \mapsto B^{1/2} f\big(B^{-1/2} A B^{-1/2}\big) B^{1/2} \tag{4.16}$$

is operator convex on $\mathcal{B}_n(0,\infty) \times \mathcal{B}_n(0,\infty)$.

Proof By Loewner's representation theorem of matrix monotone functions [19], the above function of (A, B) can be presented as

$$B^{1/2} f\big(B^{-1/2} A B^{-1/2}\big) B^{1/2} = aA + bB + \int_0^\infty \frac{1}{(tA)^{-1} + B^{-1}} \cdot \frac{1+t}{t} \mathrm{d}\nu(t) \tag{4.17}$$

at some $a, b \in \mathbb{R}$, and ν an appropriate measure on $(0,\infty)$ (cf. [21, Theorem 3.1]). The stated concavity therefore follows from the concavity of parallel sums, of Theorem 4.2. □

5 Applications to quantum entropy

5.1 Basics notions

Convexity considerations are of fundamental significance in statistical mechanics. In particular, that is so in relation to the entropy function—a term which shows up in a number of different contexts, with slightly varying meaning and properties, cf. [15].

In quantum physics, the states of a system are presented as expectation value functionals defined over the self-adjoint (bounded) operators in the corresponding separable Hilbert space $\mathcal{H}$. These are described by density operators ρ, which are positive and of trace $\mathrm{Tr}(\rho) = 1$, in terms of which the expectation value functional is the mapping

$$A \mapsto \mathrm{Tr}(A\rho). \tag{5.1}$$

The von Neumann entropy of the state is defined as

$$S(\rho) := -\mathrm{Tr}(\rho \log \rho) \equiv \mathrm{Tr} F(\rho) \tag{5.2}$$

where F is the function $F(x) = -x \log(x)$.

The basic notions and fundamental properties of entropy are explained in a pedagogical manner in [15, 17]. The entropy defined in (5.2) is non-negative, vanishes if and only if ρ is a rank-one projection, and it attains its maximal value at $\rho = \frac{1}{\dim \mathcal{H}} \mathbb{1}$.

For systems composed of two disjoint, but possibly interacting, components, the relevant Hilbert space is a tensor product, $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$. A general state on such a system would be referred to as ρ_{12}. Its restriction to a sub component, e.g., observables of the form $A \otimes \mathbb{1}$ defines the state operator ρ_1 .

The mapping

$$\rho_{12} \mapsto \rho_1 \tag{5.3}$$

can be accomplished through partial trace, or alternatively through the following relation

$$\int (\mathbb{1} \otimes U^*) \ \rho_{12} \ (\mathbb{1} \otimes U) \ \nu_2(\mathrm{d}U) = \rho_1 \otimes \frac{1}{\dim \mathcal{H}_2} \mathbb{1} \tag{5.4}$$

where ν_2 is the normalized Haar measure on the group of unitary transformations on $\mathcal{H}_2$.
Another useful mapping is the de-correlation of the two components (or "pinching"—in the terminology of [22]):

$$\rho_{12} \mapsto \rho_1 \otimes \rho_2. \tag{5.5}$$

This can be accomplished through the average over rotations by elements of the group

$$\mathcal{G}_{\rho_{12},1} := \{U_1 \otimes \mathbb{1} \ : \ [U_1, \rho_1] = 0\}. \tag{5.6}$$

Denoting the corresponding normalized Haar measure by $\nu_{\rho_{12},1}$, one has

$$\int (U^* \otimes \mathbb{1}) \ \rho_{12} \ (U \otimes \mathbb{1}) \ \nu_{\rho_{12},1}(\mathrm{d}U) = \rho_1 \otimes \rho_2 \tag{5.7}$$

(which is also true if it is the second component which is rotated[3]).

Similar relations hold for a system which is composed of three or more components. In discussing the entropies of the states induced on subsystems, it is customary to abbreviate

$$S_1 = S(\rho_1), \qquad S_{12} = S(\rho_{12}), \qquad S_{123} = S(\rho_{123}). \tag{5.8}$$

In the analogous discrete classical systems, in each state ρ the entropies of the states induced on subsystems are increasing in the subsystem size, e.g.,

$$S_1 \leq S_{12}. \tag{5.9}$$

This feature clearly does not extend to the quantum case, e.g., a composite system can be in a pure state of zero entropy, while its subsystems will not be so.

Although the analogy with the classical case in places breaks down, it is of value as a guide. Among the first generally valid properties of the above constructs is the *subadditivity of entropy*:

3 The statement is easily seen in the basis in which ρ_j is diagonal.

Theorem 5.1 (Subadditivity) *For any state of a composite system*

$$S_{12} \le S_1 + S_2. \tag{5.10}$$

Proof The statement readily follows through Jensen inequality applied to the relation (5.7)

$$S(\rho_{1,2}) \;\le\; S\,(\rho_1 \otimes \rho_2) = S_1 + S_2. \tag{5.11}$$

□

Entropy's subadditivity (5.10) can be read as saying that the increment

$$S(1|2) := S_{12} - S_2, \tag{5.12}$$

to which we refer to as the conditional entropy of component 1 given 2, satisfies

$$S(1|2) \le S_1. \tag{5.13}$$

In the analogous classical discrete systems, the state of a composite system is described by a probability distribution with weights $\rho(\omega_1, \omega_2, \ldots)$ over the space of configurations $\Omega = \Omega_1 \times \Omega_2 \times \cdots$. The probability distribution of ω_1, may be presented as the average of the conditional distribution $\rho(\omega_1|\omega_2)$. A simple calculation then shows that for classical systems $S(1|2)$ equals the average over ω_2 (with measure ρ_2) of the entropy of the conditional probability distribution $\rho(\omega_1|\omega_2)$. From this emerged the term *relative entropy*.

If there is a third component, from the above description through a judicious application of the Jensen inequality one may conclude that in the classical case

$$S_{123} - S_{23} \le S_{12} - S_2. \tag{5.14}$$

This relation is referred to as the *strong subadditivity of entropy (SSA)*. Its intuitive interpretation is that as each measurement adds information (in the classical case), the incremental entropy is decreasing.

The above reasoning does not apply to quantum systems. Nevertheless, Lanford and Robinson [13] conjectured in 1968 that SSA is valid also in the quantum case. The conjecture was proven in 1973 in the joint work by Lieb and Ruskai [17]. Already in that paper the authors provide a number of proofs. Given the interest in the subject, it is not surprising that since then a number of simplifications have emerged. The following is a streamlined proof using the results presented above.

5.2 The Lieb–Ruskai concavity

A key step in the first listed proof of SSA in [17] is the following noteworthy statement (Theorem 1 in [17]), which is also of independent interest, cf. [7].

Theorem 5.2 (Lieb–Ruskai [17]) *In any composite system, the conditional entropy*

$$\rho_{12} \mapsto S(\rho_{12}) - S(\rho_2) \equiv \mathrm{Tr}\rho_{12}\big[\log \rho_{12} - \mathbb{1} \otimes \log \rho_2\big]. \tag{5.15}$$

is a concave function of the state operator ρ.

The proof (version of Lindblad [18]) starts with a slightly more general version of the statement, asserting that the following mapping is jointly concave over pairs of positive matrices

$$(A, B) \mapsto -\mathrm{Tr}\big[A(\log A - \log B)\big]. \tag{5.16}$$

The combination on the right is referred to as the relative entropy, and denoted

$$S(A|B) := -\mathrm{Tr}\big[A(\log A - \log B)\big]$$

Proof By Lieb concavity, stated above as Corollary 4.3, for every $\varepsilon > 0$ the following mapping is jointly concave over pairs of positive matrices

$$(A, B) \mapsto \frac{\mathrm{Tr}_{12}\big[A^{1-\varepsilon} B^{\varepsilon} - A\big]}{\varepsilon} \equiv \mathrm{Tr}_{12}\Big[\frac{A^{1-\varepsilon} - A}{\varepsilon} B^{\varepsilon} + A \frac{B^{\varepsilon} - \mathbb{1}}{\varepsilon}\Big]. \tag{5.17}$$

The claim follows by taking the limit $\varepsilon \searrow 0$.

To relate this with 5.15, apply the above to $A = \rho_{12}$, $B = \mathbb{1} \otimes \rho_2$. In this case

$$\begin{aligned} \mathrm{Tr}\big[A \log B)\big] &= \mathrm{Tr}_{12}\ \rho_{12}\ \log(\mathbb{1} \otimes \rho_2) = \mathrm{Tr}_{12}\ \rho_{12}\ (\mathbb{1} \otimes \log \rho_2) \\ &= \mathrm{Tr}_2\ \rho_2 \log \rho_2 = -S_2, \end{aligned} \tag{5.18}$$

where the subscript on Tr indicates the Hilbert space over which the trace is performed. Thus, under this substitution

$$S(A|B) = -\mathrm{Tr}\rho_{12}\big[\log \rho_{12} - \mathbb{1} \otimes \log \rho_2\big] = S(\rho_{12}) - S(\rho_2). \tag{5.19}$$

Since both A and B are linear functions of ρ, the just proven jointly concavity of $S(A|B)$ implies the claimed concavity of the conditional entropy as a function of ρ. □

5.3 The strong subadditivity of quantum entropy

Theorem 5.3 (Lieb-Ruskai [17]) *The von Neumann entropy of a quantum (finite dimensional) systems is strongly subadditive, in the sense that for any state of a composite system*

$$S(\rho_{123}) - S(\rho_{23}) \leq S(\rho_{12}) - S(\rho_2)\,. \tag{5.20}$$

The following is a streamlined deduction of this statement from the Lieb–Ruskai concavity, incorporating an argument of Uhlmann [24, Sect. 8].

Proof As seen in (5.4) and (5.7), any state ρ_{123} can be transformed into one in which the third component is independent from the first two through a probability average over unitary transformations which do not affect the distribution of the first two components ($U = \mathbb{1} \otimes \mathbb{1} \otimes U_3$). The resulting state is a product state of the form

$$\widetilde{\rho}_{123} = \rho_{12} \otimes \widetilde{\rho}_3, \tag{5.21}$$

where $\widetilde{\rho}_3$ could be either ρ_3 or $(\dim \mathcal{H}_3)^{-1}\mathbb{1}$.

Hence $\widetilde{\rho}_{123}$ is the probability average of states whose relative entropy $S(1|2)$ equals that of ρ_{123}. From the Lieb–Ruskai concavity principle (Theorem 5.2), combined with the Jensen inequality (1.2), it follows that

$$S(\rho_{123}) - S(\rho_{23}) \leq S(\widetilde{\rho}_{123}) - S(\widetilde{\rho}_{23}). \tag{5.22}$$

However, due to the above product structure of $\widetilde{\rho}_{123}$ the contribution of the third component to its entropy is strictly additive, while its restriction to the first two components equals that of ρ. Therefore

$$S(\widetilde{\rho}_{123}) - S(\widetilde{\rho}_{23}) = S(\widetilde{\rho}_{12}) - S(\widetilde{\rho}_2) = S(\rho_{12}) - S(\rho_2). \tag{5.23}$$

□

Let us add that while our discussion is carried in the context of matrices of finite order n, by continuity arguments, spelled in [2] and [17](appendix by B. Simon), results that hold uniformly in n admit natural extensions to bounded linear operators over separable Hilbert spaces, i.e., the class $\mathcal{B}_\infty(a, b)$.

It should also be noted that the above-cited works on quantum entropy, and other contributions by mathematical physicists, spawned a growing body of interesting results concerning matrix convexity. A relevant recent review can be found in [4].

Acknowledgements We thank Ramon van Handel and Simone Warzel for useful comments on an earlier draft of this manuscript and Benjamin Bobell for motivating discussions. M.A. gratefully acknowledges the support of the Weston Visiting Professorship at the Weizmann Institute of Science (Rehovot, Israel), on a visit during which some of the work was done.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Anderson, W.N., Jr., Duffin, R.J.: Series and parallel addition of matrices. J. Math. Anal. Appl. **26**, 576–594 (1969)
2. Bendat, J., Sherman, S.: Monotone and convex operator functions. Trans. Am. Math. Soc. **79**, 58–71 (1955)
3. Brown, A.L., Vasudeva, H.L.: The calculus of operator functions and operator convexity. Diss. Math. (Rozprawy Mat.) **390**, 48 (2000)

4. Carlen, E.A.: On some convexity and monotonicity inequalities of Elliott Lieb. In: The Physics and Mathematics of Elliott Lieb, vol. 1, pp. 143–209. European Mathematical Society Press (2022)
5. Effros, E.G.: A matrix convexity approach to some celebrated quantum inequalities. Proc. Natl. Acad. Sci. USA **106**, 1006–1008 (2009)
6. Epstein, H.: Remarks on two theorems of E. Lieb. Commun. Math. Phys. **31**, 317–325 (1973)
7. Hansen, F.: Quantum entropy derived from first principles. J. Stat. Phys. **165**, 799–808 (2016)
8. Hansen, F., Pedersen, G.K.: Jensen's operator inequality. Bull. Lond. Math. Soc. **35**, 553–564 (2003)
9. Hansen, F., Tomiyama, J.: Differential analysis of matrix convex functions. II. J. Inequal. Pure Appl. Math. 10 (2009)
10. Hiai, F.: Concavity of certain matrix trace functions. Taiwan. J. Math. **5**, 535–554 (2001)
11. Kraus, F.: Über konvexe Matrixfunktionen. Math. Z. **41**, 18–42 (1936)
12. Kubo, F., Ando, T.: Means of positive linear operators. Math. Ann. **246**, 205–224 (1980)
13. Lanford, O.E., III., Robinson, D.W.: Statistical mechanics of quantum spin systems. III. Commun. Math. Phys. **9**, 327–338 (1968)
14. Lieb, E.H.: Convex trace functions and the Wigner–Yanase–Dyson conjecture. Adv. Math. **11**, 267–288 (1973)
15. Lieb, E.H.: Some convexity and subadditivity properties of entropy. Bull. Am. Math. Soc. **81**, 1–13 (1975)
16. Lieb, E.H., Ruskai, M.B.: A fundamental property of quantum-mechanical entropy. Phys. Rev. Lett. **30**, 434–436 (1973)
17. Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum-mechanical entropy. J. Math. Phys. **14**, 1938–1941 (1973). (**With an appendix by B. Simon**)
18. Lindblad, G.: Expectations and entropy inequalities for finite quantum systems. Commun. Math. Phys. **39**, 111–119 (1974)
19. Löwner, K.: Über monotone matrixfunktionen. Math. Z. **38**, 177–216 (1934)
20. Pusz, W., Woronowicz, S.L.: Functional calculus for sesquilinear forms and the purification map. Reports Math. Phys. **8**, 159–170 (1975)
21. Simon, B.: Loewner's theorem on monotone matrix functions. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 354. Springer, Cham (2019)
22. Sutter, D., Berta, M., Tomamichel, M.: Multivariate trace inequalities. Commun. Math. Phys. **352**, 37–58 (2017)
23. Uchiyama, M.: Operator monotone functions, positive definite kernel and majorization. Proc. Am. Math. Soc. **138**, 3985–3996 (2010)
24. Uhlmann, A.: Endlich-dimensionale Dichtematrizen II. Wiss. Z. Karl-Marx-Univ. Leipzig, Math.-Nat. R. **22**, 139–177 (1973)
25. Wigner, E.P., Yanase, M.M.: Information contents of distributions. Proc. Nat. Acad. Sci. U.S.A. **49**, 910–918 (1963)