A matrix convexity approach to some celebrated quantum inequalities

Edward G. Effros[1]

Department of Mathematics, University of California, Los Angeles, CA 90095-1555

Edited by Richard V. Kadison, University of Pennsylvania, Philadelphia, PA, and approved November 25, 2008 (received for review November 24, 2008)

Some of the important inequalities associated with quantum entropy are immediate algebraic consequences of the Hansen–Pedersen–Jensen inequalities. A general argument is given using matrix perspectives of operator convex functions. A matrix analogue of Maréchal's extended perspectives provides additional inequalities, including a $p+q \leq 1$ result of Lieb.

quantum information theory | relative quantum entropy | Lieb | Hansen–Jensen–Pedersen | matrix convexity

In 1973, Elliott Lieb published a ground-breaking article on operator inequalities (1). This and a subsequent article by Lieb and Ruskai (2) have had a profound effect on quantum statistical mechanics, and more recently on quantum information theory. Since then, a number of attempts have been made to elucidate and extend these results. Two elegant examples are those of Nielsen and Petz (3), and Ruskai (4), which use the analytic representations for operator convex functions. In addition, Frank Hansen (5) has developed a powerful theory that utilizes *geometric means* of positive operators. The latter notion was formulated by Pusz and Woronowicz (6), and subsequently investigated by Ando (7) (see the discussion in Section 3) and by Kubo and Ando (8).

1. Introduction

We present what is arguably the simplest approach to these inequalities. This is accomplished by using matrix analogues of two elementary ideas from classical convexity theory: the Jensen inequality, and the construction of the perspective of a convex function. For the first, we employ the matricial Jensen inequality of Frank Hansen and Gert Pedersen (9, 10). As we point out in Section 5, the affine and homogeneous versions of this inequality can be proved in a relatively few lines drawn from those articles. The noncommutative analogues of perspectives are completely straightforward in the context of the left and right module operations that are standard to the subject. In Section 4 we show that the same approach may be used to quantize Maréchal's extended version of the perspective. We apply this to prove Lieb's generalized $p+q \leq 1$ inequality (see also the elegant proof in ref. 5). It should be noted that Petz gave a direct proof of the convexity of the relative entropy in the von Neumann algebra case in which he used modular theory together with Hansen–Jensen–Pedersen inequality (11).

The appearance of notions from convexity theory suggests that other geometric techniques will prove useful in the operator context. In a different direction, quantum information theory is likely to have an impact on the theory of matrix convexity. This possibility is considered in Section 4.

Since the basic difficulties are already apparent in finite dimensions, we have restricted our attention to finite matrices, and we have avoided any attempt at full generality even in that context.

2. The Classical and Matrix Notions of Perspectives

Given a convex function f defined on a convex set $K \subseteq \mathbb{R}^n$, the *perspective* g is defined on the subset

$$L = \{(x,t) : t > 0 \text{ and } x/t \in K\}$$

by

$$g(x,t) = f(x/t)t$$

(see ref. 12). It is a simple exercise to verify that $g(x,t)$ is a jointly convex function in the sense that, if $0 \leq c \leq 1$, then

$$g(cx_1 + (1-c)x_2, ct_1 + (1-c)t_2) \leq cg(x_1,t_1) + (1-c)g(x_2,t_2).$$

An elementary but important example is provided by the continuous convex function $f(x) = x\log x$, with $f(0) = 0$ defined on $[0,\infty) \subseteq \mathbb{R}$. It follows that the perspective function

$$g(x,t) = t\frac{x}{t}\log\frac{x}{t} = x\log x - x\log t$$

is jointly convex. Letting $p = (p_i)$ and $q = (q_i)$ be finite probability measures with $p_i > 0$ and $q_i > 0$, the convexity of f implies that the classical entropy

$$H(p) = -\sum p_i \log p_i$$

is concave, and the convexity of g implies that the relative entropy

$$(q,p) \mapsto H(q\|p) = \sum p_i \log p_i - p_i \log q_i$$

is jointly convex on pairs of probability measures.

We recall that if $f : I = [a,b] \to \mathbb{R}$ is continuous, and T is an $n \times n$ self-adjoint matrix with spectrum in $[a,b]$, then we can define $f_n(T)$ by spectral theory (or by using a basis in which T is diagonal). f is said to be *matrix convex* if for each $n \in \mathbb{N}$, the corresponding function f_n is convex on the self-adjoint $n \times n$ matrices with spectrum in $[a,b]$. Throughout the rest of the article we only consider $n \times n$ matrices, and we usually omit the subscript n. The following is the affine version of the Hansen–Pedersen–Jensen inequality (10) (see Section 5).

Theorem 2.1. *If f is matrix convex, and A and B satisfy $A^*A + B^*B = I_n$, then*

$$f(A^*T_1A + B^*T_2B) \leq A^*f(T_1)A + B^*f(T_2)B. \qquad \textbf{[2.1]}$$

We begin with some matrix conventions. Given matrices L and R, we let $[L,R] = LR - RL$. Let us suppose that $L > 0$ and $R > 0$. If $[L,R] = 0$, i.e., the matrices commute, then we may find a basis in which both matrices are diagonalized. It follows that $LR > 0$, $[L,R^{-1}] = 0$, and we may unambiguously write $\frac{L}{R}$ for the quotient. We also recall that for any continuous function f, $f(L)$ commutes with any operator commuting with L (including L itself). By using simultaneously diagonalized matrices, it is evident that we have relations such as $\log LR^{-1} = \log L - \log R$.

Author contributions: E.G.E. wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

[1]E-mail: ege@math.ucla.edu.

Theorem 2.2. *Suppose that f is operator convex. When restricted to positve commuting matrices L, R, the "perspective function"*

$$(L,R) \mapsto g(L,R) = f\left(\frac{L}{R}\right) R \quad \textbf{[2.2]}$$

is jointly convex in the sense that if $L = cL_1 + (1-c)L_2$ *and* $R = cR_1 + (1-c)R_2$ *where* $[L_j, R_j] = 0$ $(j = 1, 2)$*, and* $0 \le c \le 1$*, then*

$$g(L,R) \le cg(L_1,R_1) + (1-c)g(L_2,R_2). \quad \textbf{[2.3]}$$

Proof: The matrices $A = (cR_1)^{1/2}R^{-1/2}$ and $B = ((1-c)R_2)^{1/2}R^{-1/2}$ satisfy $A^*A + B^*B = I$. From *Theorem 2.1*,

$$\begin{aligned}
&g(L,R)\\
&= Rf\left(\frac{L}{R}\right)\\
&= R^{1/2}f(R^{-1/2}LR^{-1/2})R^{1/2}\\
&= R^{1/2}f\left(A^*\left(\frac{L_1}{R_1}\right)A + B^*\left(\frac{L_2}{R_2}\right)B\right)R^{1/2}\\
&\le R^{1/2}\left(A^*f\left(\frac{L_1}{R_1}\right)A + B^*f\left(\frac{L_2}{R_2}\right)B\right)R^{1/2}\\
&= (cR_1)^{1/2}f\left(\frac{L_1}{R_1}\right)(cR_1)^{1/2}\\
&\quad + ((1-c)R_2)^{1/2}f\left(\frac{L_2}{R_2}\right)((1-c)R_2)^{1/2}\\
&= cg(L_1,R_1) + (1-c)g(L_2,R_2).
\end{aligned}$$

□

The following result is due to Lieb and Ruskai (2) [a related early discussion may be found in Lindblad (13)].

Corollary 2.1. *The relative entropy function*

$$(\rho,\sigma) \mapsto S(\rho||\sigma) = \text{Trace } \rho\log\rho - \rho\log\sigma$$

is jointly convex on the strictly positive $n \times n$ *density matrices* ρ*,* σ*.*

Proof: We let M_n have the usual Hilbert space structure determined by $\langle X, Y\rangle = \text{Trace } XY^*$. Given positive density matrices σ and ρ, we define operators R and L on M_n by $L(X) = \rho X$ and $R(X) = X\sigma$. Then, we have that $L(X)$ and $R(X)$ are commuting positive operators on the Hilbert space M_n. However, the function $f(x) = x\log x$ is operator convex (see ref. 14, p. 123), and thus

$$\begin{aligned}
\langle g(L,R)(I), I\rangle &= \left\langle R\left(\frac{L}{R}\right)\log\left(\frac{L}{R}\right)(I), I\right\rangle\\
&= \langle L(\log L - \log R)(I), I\rangle\\
&= \text{Trace } \rho\log\rho - \rho\log\sigma = S(\rho||\sigma)
\end{aligned}$$

is jointly convex. □

The following is due to Lieb (1). It was subsequently used by Lieb and Ruskai to prove strong subadditivity for relative entropy (2). A stronger result of Lieb is discussed in the next section.

Corollary 2.2. *If* $0 < s < 1$*, then the function*

$$F(A,B) = \text{Trace } A^sK^*B^{1-s}K$$

is jointly concave on the strictly positive $n \times n$ *matrices* A, B*.*

Proof: Since $f(t) = -t^s$ is operator convex (see ref. 14; Th.5.1.9), $g(L,R) = -L^sR^{1-s}$ is jointly convex for appropriately commuting operators. Again using the Hilbert space structure on M_n, we let $L(X) = AX$ and $R(X) = XB$. It follows that

$$(A,B) \mapsto -\text{Trace } A^sK^*B^{1-s}K = \langle g(L,R)(K^*), K^*\rangle$$

is jointly convex. □

Various generalized entropies may be handled in much the same manner.

3. Maréchal's Perspectives

P. Maréchal has recently introduced an interesting generalization of perspectivity for convex functions (15, 16). This also has a natural matrix version. For this purpose we use the following subhomogeneous form of the Hansen–Pedersen–Jensen inequality (9) (see Section 5). We assume that the functions f and g are defined on an interval $I \subseteq \mathbb{R}$, and that $0 \in I$.

Theorem 3.1. *If f is matrix convex,* $f(0) \le 0$*, and A and B are matrices with* $A^*A + B^*B \le I_n$*, then*

$$f(A^*T_1A + B^*T_2B) \le A^*f(T_1)A + B^*f(T_2)B.$$

Given continuous functions f and h, and commuting positive matrices L and R, we define

$$(f\Delta h)(L,R) = f\left(\frac{L}{h(R)}\right)h(R)$$

A close variation of the following result was proved for operator monotone functions f on $(0,\infty)$ by Ando (see ref. 7 theorem 6). His construction (without the extra function Φ_2 that can be incorporated with a composition) is related to Maréchal's operation $f\nabla h$ for concave functions f and h. Ando invoked the integral representation for operator monotone functions, rather than the matrix convexity argument used below.

Theorem 3.2. *Suppose that f is matrix convex,* $f(0) \le 0$*, and that h is matrix concave with* $h > 0$*. Then* $(L,R) \mapsto (f\Delta h)(L,R)$ *is jointly convex on positive commuting matrices L, R in the sense of Theorem 2.2.*

Proof: Let us suppose that $L = cL_1 + (1-c)L_2$ and $R = cR_1 + (1-c)R_2$ where $[L_j, R_j] = 0$. Then $ch(R_1) + (1-c)h(R_2) \le h(R)$, hence

$$\begin{aligned}
A &= c^{1/2}h(R_1)^{1/2}h(R)^{-1/2}\\
B &= (1-c)^{1/2}h(R_2)^{1/2}h(R)^{-1/2}
\end{aligned}$$

satisfy

$$\begin{aligned}
&A^*A + B^*B\\
&= h(R)^{-1/2}ch(R_1)h(R)^{1/2} + h(R)^{-1/2}(1-c)h(R_2)h(R)^{-1/2}\\
&\le h(R)^{-1/2}h(R)h(R)^{-1/2}I = I.
\end{aligned}$$

It follows from *Theorem 3.1* that

$$\begin{aligned}
&(f\Delta h)(L,R)\\
&= h(R)^{1/2}f(h(R)^{-1/2}Lh(R)^{-1/2})h(R)^{1/2}\\
&= h(R)^{1/2}f\left(A^*\left(\frac{L_1}{h(R_1)}\right)A + B^*\left(\frac{L_2}{h(R_2)}\right)B\right)h(R)^{1/2}\\
&\le h(R)^{1/2}A^*f\left(\frac{L_1}{h(R_1)}\right)Ah(R)^{1/2}\\
&\quad + h(R)^{1/2}B^*f\left(\frac{L_2}{h(R_2)}\right)Bh(R)^{1/2}\\
&= ch(R_1)^{1/2}f\left(\frac{L_1}{h(R_1)}\right)h(R_1)^{1/2}\\
&\quad + (1-c)h(R_2)^{1/2}f\left(\frac{L_2}{h(R_2)}\right)h(R_2)^{1/2}\\
&= c(f\Delta h)(L_1,R_1) + (1-c)(f\Delta h)(L_2,R_2).
\end{aligned}$$

□

MATHEMATICS

To illustrate this result, we reprove Lieb's extension of *Corollary 2.2* (1).

Corollary 3.1. *Suppose that* $0 < p, q$ *and that* $p + q \leq 1$. *Then the function*

$$(A, B) \mapsto \text{Trace } A^q X^* B^p X$$

is jointly concave on the positive $n \times n$ *matrices.*

Proof: Since $p + q \leq 1$, $p + q$ is a convex combination of q and 1, i.e., we may choose $0 \leq t \leq 1$ with $p + q = (1 - t)q + t1$. If we let $q = s$, then

$$p = -tq + t = (1 - q)t = (1 - s)t.$$

Thus, it suffices to show that if $0 \leq s, t \leq 1$, then

$$(A, B) \mapsto -\text{Trace } A^s X^* B^{(1-s)t} X$$

is jointly convex. The functions $f(x) = -x^s$ and $h(y) = y^t$ are operator convex and concave, respectively, and

$$(f \Delta h)(L, R) = h(R) f\left(\frac{L}{h(R)}\right) = -R^t \frac{L^s}{R^{st}} = -L^s R^{(1-s)t}.$$

If we let $L(X) = AX$ and $R(X) = XB$ for $X \in M_n$, then from *Theorem 3.2*,

$$(A, B) \mapsto -\text{Trace } A^s X^* B^{(1-s)t} X = \langle (f \Delta h)(L, R)(X^*), X^* \rangle$$

is jointly convex. □

4. Matrix Convexity

Perhaps the most intriguing aspect of Maréchal's construction is that it behaves well under the Fenchel–Legendre transform, and under iteration. Søren Winkler formulated an analogue of the Fenchel–Legendre duality for matrix convex functions (17), but the transforms are generally set-valued mappings. Further progress might result if one could reformulate his theory in terms of "left-right" commuting pairs. It should also be noted that other constructions in classical convexity theory, such as the linear fractional transformations of convex functions (see ref. 18) might also have matrix generalizations.

Until recently the theory of matrix convexity has suffered from a lack of examples and applications. With the advent of quantum information theory (QIT), this situation has dramatically changed. QIT provides a wealth of remarkable, purely nonclassical techniques that might clarify some of the conceptual problems in matrix convexity theory. However, it seems likely that matrix convexity and more generally noncommutative functional analysis will provide an appropriate framework for many of the calculations in QIT. A striking illustration of this phenomenon can be found in ref. 19.

5. A Brief Guide to the Hansen–Pedersen–Jensen Inequalities

The original proof of *Theorem 3.1* may be found in ref. 10 (*Theorem 2.1*). It is both elegant and concise. For our purposes we only need (i) implies (iii) in their proof. Winkler pointed out in ref. 17 that *Theorem 2.1* is easily derived from *Theorem 3.1*. Since our situation is slightly different, we include the argument.

We fix a point $c \in I$ and define $F(t) = f(t + c) - f(c)$. Given $T = T^* \in M_n$, we may choose a basis with respect to which $T = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Then

$$\begin{aligned} F(T) &= F\left(\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}\right) \\ &= \begin{bmatrix} f(\lambda_1 + c) - f(c) & & \\ & \ddots & \\ & & f(\lambda_n + c) - f(c) \end{bmatrix} \\ &= f(T + cI) - f(c)I \end{aligned}$$

is matrix convex and $F(0) = 0$. From *Theorem 3.1*,

$$F(A^* T_1 A + B^* T_2 B) \leq A^* F(T_1) A + B^* F(T_2) B,$$

$$\begin{aligned} & f(A^* T_1 A + B^* T_2 B) - f(c)I \\ & \leq A^* f(T_1) A - f(c) A^* A + B^* f(T_2) B - f(c) B^* B, \end{aligned}$$

and thus

$$f(A^* T_1 A + B^* T_2 B) \leq A^* f(T_1) A + B^* f(T_2) B.$$

As pointed out by Winkler (17), the result may be extended to rectangular matrices A and B. He used the case $B = 0$ to show that a real function f on an interval in $\mathbb{R}$ is a matrix convex function if and only if the supergraphs of the f_n form a matrix convex system of sets.

ACKNOWLEDGMENTS. Dedicated to Gert Pedersen, whose beautiful work with Frank Hansen on matrix convexity led to the genesis of this paper. He will be missed for both his brilliance and his exuberance, which inspired a generation of young mathematicians. I thank Frank Hansen for alerting us to his work in this area, Mary Beth Ruskai for correcting a number of errors in my first manuscript, Jon Tyson for a host of suggestions, and Richard Kadison for his encouragement.

1. Lieb E (1973) Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Adv Math* 11:267–288.
2. Lieb E, Ruskai M (1973) Proof of the strong subadditivity of quantum-mechanical entropy. With an appendix by B. *Simon J Math Phys* 14:1938–1941.
3. Nielsen M, Petz D (2005) A simple proof of the strong subadditivity inequality. *Quantum Inf Comput* 5:507–513.
4. Ruskai M (2007) Another short and elementary proof of strong subadditivity of quantum entropy. *Rep Math Phys* 60:1–12.
5. Hansen F (2006) Extensions of Lieb's Concavity Theorem. *J Stat Phys* 124:87–101.
6. Pusz W, Woronowicz S (1975) Functional calculus for sesquilinear forms and the purification map. *Rep Math Phys* 8:159–170.
7. Ando T (1979) Concavity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra Appl* 26:203–241.
8. Kubo F, Ando T (1979-1980) Means of positive linear operators. *Math Ann* 246:205–224.
9. Hansen F, Pedersen G (1981-1982) Jensen's inequality for operators and Löwner's theorem. *Math Ann* 258:229–241.
10. Hansen F, Pedersen G (2003) Jensen's operator inequality. *Bull London Math Soc* 35:553–564.
11. Petz D (1986) Quasi-entropies for finite quantum systems. *Rep Math Phys* 23:57–65.
12. Hiriart-Urruty J, Lemarchal C (2001) *Fundamentals of Convex Analysis, Grundlehren Text Ed.* (Springer, Berlin).
13. Lindblad G (1973) Entropy, information and quantum measurements. *Commun Math Phys* 33:305–322.
14. Bhatia R (1997) *Matrix Analysis.* Graduate Texts in Mathematics 169 (Springer, New York).
15. Maréchal P (2005) On a functional operation generating convex functions. I. Duality. *J Optim Theory Appl* 126:175–189.
16. Maréchal P (2005) On a functional operation generating convex functions. II. Algebraic properties. *J Optim Theory Appl* 126:357–366.
17. Winkler S (1999) The non-commutative Legendre-Fenchel transform. *Math Scand* 85:30–48.
18. Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge Univ Press, New York).
19. Devetak I, Junge M, King C, Ruskai M (2006) Multiplicativity of completely bounded *p*-norms implies a new additivity result. *Commun Math Phys* 266:37–63.