A MATRIX CONVEXITY APPROACH TO SOME CELEBRATED QUANTUM INEQUALITIES

EDWARD G. EFFROS

Dedicated to Gert Pedersen, who is missed for both his brilliance and his exuberant sense of humor.

ABSTRACT. 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.

1. Introduction

In 1973, Elliott Lieb published a ground-breaking paper on operator inequalities [10]. This and a subsequent paper by Lieb and Ruskai [11] 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 particularly elegant examples are those of Nielsen and Petz [15], and Ruskai [17]), which use the analytic representations for operator convex functions. On the other hand, Frank Hansen [7] has developed a powerful theory that utilizes geometric means of positive operators. The latter noton was formulated by Pusz and Woronowicz [16], and subsequently investigated by Ando [1] (see the discussion in Section 3) and by Kubo and Ando [9].

Here 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 [5], [6]. 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 papers. The non-commutative analogues of perspectives are completely straight-forward 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

Date: January 28, 2008.

Supported by the National Science Foundation DMS-0100883.

prove Lieb's generalized $p + q \le 1$ inequality (see also the elegant proof in [7]).

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 possility is considered in Section 4.

I am grateful to Frank Hansen for alerting us to his work in this area. I also wish to thank Mary Beth Ruskai, who corrected a number of errors in my first manuscript, Jon Tyson for a host of suggestions, and Richard Kadison for his encouragement.

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 [8]). It is a simple exercise to verify that g(x,t) is a jointly convex function in the sense that if $0 \le c \le 1$, then

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

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 paper 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 [6] (see Section 5).

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

$$(2.1) f(A^*T_1A + B^*T_2B) \le A^*f(T_1)A + B^*f(T_2)B.$$

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). Using simultaneously diagonalized matrices, it is evident that we have relations such as $\log LR^{-1} = \log L - \log R$.

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

$$(2.2) \hspace{1cm} (L,R) \mapsto g(L,R) = f\left(\frac{L}{R}\right)R$$

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

(2.3)
$$g(L,R) \le cg(L_1,R_1) + (1-c)g(L_2,R_2).$$

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,

$$g(L,R)$$

$$= Rf\left(\frac{L}{R}\right)$$

$$= R^{1/2}f(R^{-1/2}LR^{-1/2})R^{1/2}$$

$$= R^{1/2}f\left(A^*\left(\frac{L_1}{R_1}\right)A + B^*\left(\frac{L_2}{R_2}\right)B\right)R^{1/2}$$

$$\leq R^{1/2}\left(A^*f\left(\frac{L_1}{R_1}\right)A + B^*f\left(\frac{L_2}{R_2}\right)B\right)R^{1/2}$$

$$= (cR_1)^{1/2}f\left(\frac{L_1}{R_1}\right)(cR_1)^{1/2} + ((1-c)R_2)^{1/2}f\left(\frac{L_2}{R_2}\right)((1-c)R_2)^{1/2}$$

$$= cg(L_1, R_1) + (1-c)g(L_2, R_2).$$

The following result is due to Lieb and Ruskai [11] (a related early discussion may be found in Lindblad [12]).

Corollary 2.3. The relative entropy function

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

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 . On the other hand the function $f(x) = x \log x$ is operator convex (see [2], p. 123), and thus

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

is jointly convex.

The following is due to Lieb [10]. It was subsequently used by Lieb and Ruskai to prove strong subadditivity for relative entropy [11]. A stronger result of Lieb is discussed in the next section.

Corollary 2.4. If 0 < s < 1, then the function

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

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

Proof. Since $f(t) = -t^s$ is operator convex (see [2] Th.5.1.9), $g(L,R) = -L^s R^{1-s}$ 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

$$(A,B)\mapsto -{\rm Trace}\, A^sK^*B^{1-s}K=\langle g(L,R)(K^*),K^*\rangle$$
 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 [13], [14]. This also has a natural matrix version. For this purpose we use the subhomogeneous form of the Hansen-Pedersen-Jensen inequality [5] (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, and $f(0) \le 0$, and that 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 [1] Theorem 6). His construction (without the extra function Φ_2 which can be incorporated with a compostion) 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, $f(0) \leq 0$ and that h is matrix concave with h > 0. Then $(L, R) \mapsto (f\Delta h)(L, R)$ is jointly convex on postive 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 $[L_i, R_i] = 0$. Then $ch(R_1) + (1 - c)h(R_2) \le h(R)$, hence

$$A = c^{1/2}h(R_1)^{1/2}h(R)^{-1/2}$$

$$B = (1-c)^{1/2}h(R_2)^{1/2}h(R)^{-1/2}$$

satisfy

$$A^*A + B^*B$$
= $h(R)^{-1/2}ch(R_1)h(R)^{1/2} + h(R)^{-1/2}(1-c)h(R_2)h(R)^{-1/2}$
< $h(R)^{-1/2}h(R)h(R)^{-1/2}I = I$.

It follows from Theorem 3.1 that

$$(f\Delta h)(L,R)$$

$$= h(R)^{1/2} f(h(R)^{-1/2} L h(R)^{-1/2}) h(R)^{1/2}$$

$$= h(R)^{1/2} f\left(A^* \left(\frac{L_1}{h(R_1)}\right) A + B^* \left(\frac{L_2}{h(R_2)}\right) B\right) h(R)^{1/2}$$

$$\leq h(R)^{1/2} A^* f\left(\frac{L_1}{h(R_1)}\right) A h(R)^{1/2} + h(R)^{1/2} B^* f\left(\frac{L_2}{h(R_2)}\right) B h(R)^{1/2}$$

$$= ch(R_1)^{1/2} f\left(\frac{L_1}{h(R_1)}\right) h(R_1)^{1/2} + (1-c)h(R_2)^{1/2} f\left(\frac{L_2}{h(R_2)}\right) h(R_2)^{1/2}$$

$$= c(f\Delta h)(L_1, R_1) + (1-c)(f\Delta h)(L_2, R_2).$$

To illustrate this result, we reprove Lieb's extension of Corollary 2.4 [10].

Corollary 3.3. Suppose that 0 < p, q and that $p+q \le 1$. Then the function

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

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

Proof. Since $p + q \le 1$, p + q is a convex combination of q and 1, i.e., we may choose $0 \le t \le 1$ 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 $0 \le s, t \le 1$, then

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

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,

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

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 [18], 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 [3]) 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 non-classical techniques that might clarify some of the conceptual problems in matrix convexity theory. On the other hand, it seems likely that matrix convexity and more generally non-commutative functional analysis will provide an appropriate framework for many of the calculations in QIT. A striking illustration of this phenomenon can be found in [4].

5. A BRIEF GUIDE TO THE HANSEN-PEDERSEN-JENSEN INEQUALITIES

The original proof of Theorem 3.1 may be found in [5] (Theorem 2.1). It is both elegant and concise. For our purposes we only need (i) implies (iii) in their proof. On the other hand, Winkler pointed out in [18] 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 $T = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then

$$F(T) = F\left(\begin{bmatrix} \lambda_1 \\ \ddots \\ \lambda_n \end{bmatrix}\right)$$

$$= \begin{bmatrix} f(\lambda_1 + c) - f(c) \\ & \ddots \\ & f(\lambda_n + c) - f(c) \end{bmatrix}$$

$$= f(T + cI) - f(c)I$$

is matrix convex and F(0) = 0. From Theorem 2.1,

$$F(A^*T_1A + B^*T_2B) \le A^*F(T_1)A + B^*F(T_2)B,$$

$$f(A^*T_1A + B^*T_2B) - f(c)I$$

$$\leq A^*f(T_1)A - f(c)A^*A + B^*f(T_2)B - f(c)B^*B,$$

and thus

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

As pointed out by Winkler [18], 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.

References

- [1] Ando, T. Concavity of certain maps on positive definite matrices and applications to Hadamard products. Linear Algebra Appl. 26 (1979), 203–241.
- [2] Bhatia, R. Matrix analysis. Graduate Texts in Mathematics, 169. Springer-Verlag, New York, 1997. xii+347 pp.
- [3] Boyd, S.; Vandenberghe, L. Convex Optimization, to appear.
- [4] Devetak, I.; Junge, M.; King, C.; Ruskai, M. Multiplicativity of completely bounded p-norms implies a new additivity result. Comm. Math. Phys. 266 (2006), no. 1, 37–63.
- [5] Hansen, F.; Pedersen, G. Jensen's inequality for operators and Löwner's theorem. Math. Ann. 258 (1981/82), no. 3, 229–241.
- [6] Hansen, F.; Pedersen, G. Jensen's operator inequality. Bull. London Math. Soc. 35 (2003), no. 4, 553–564.
- [7] Hansen, F. Extensions of Lieb's Concavty Theorem, J. Stat. Physics, 124 (2006), no. 1, 87-101.
- [8] Hiriart-Urruty, J.B., Lemarchal, C. (1993) Convex Analysis and Minimization Algorithms, I and II, Springer Verlag, Berlin, Germany
- [9] Kubo, Fumio; Ando, Tsuyoshi Means of positive linear operators. Math. Ann. 246 (1979/80), no. 3, 205–224.
- [10] Lieb, E. Convex trace functions and the Wigner-Yanase-Dyson conjecture. Advances in Math. 11 (1973), 267–288.
- [11] Lieb, E.; Ruskai, M. Proof of the strong subadditivity of quantum-mechanical entropy. With an appendix by B. Simon. J. Mathematical Phys. 14 (1973), 1938–1941.
- [12] Lindblad, G. Entropy, information and quantum measurements. Comm. Math. Phys. 33 (1973), 305–322.

- [13] Maréchal, P. On a functional operation generating convex functions. I. Duality. J. Optim. Theory Appl. 126 (2005), no. 1, 175–189.
- [14] Maréchal, P. On a functional operation generating convex functions. II. Algebraic properties. J. Optim. Theory Appl. 126 (2005), no. 2, 357–366.
- [15] Nielsen, M.; Petz, D. A simple proof of the strong subadditivity inequality. Quantum Inf. Comput. 5 (2005), no. 6, 507–513.
- [16] Pusz, W.; Woronowicz, S. L. Functional calculus for sesquilinear forms and the purification map. Rep. Mathematical Phys. 8 (1975), no. 2, 159–170.
- [17] Ruskai, M. Another short and elementary proof of strong subadditivity of quantum entropy. Rep. Math. Phys. 60 (2007), no. 1, 1–12.
- [18] Winkler, S. The non-commutative Legendre-Fenchel transform. Math. Scand. 85 (1999), no. 1, 30–48.

DEPARTMENT OF MATHEMATICS, UCLA, Los Angeles, CA 90095-1555 $\emph{E-mail address},$ Edward G. Effros: ege@math.ucla.edu