Elliott H. Lieb

Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture

Les rencontres physiciens-mathématiciens de Strasbourg - RCP25, 1973, tome 19
« Conférences de : P. Lelong, A. Lichnerowicz, E. Lieb et C. Stanojevic », , exp. n° 4, p. 0-35

<http://www.numdam.org/item?id=RCP25_1973__19__A4_0>

For Advances in Mathematics

CONVEX TRACE FUNCTIONS AND THE WIGNER-YANASE-DYSON CONJECTURE

Elliott H. LIEB*

Institut des Hautes Etudes Scientifiques

91440 BURES-sur-YVETTE France

:-:-:-:-:-:-:

ABSTRACT - Several convex mappings of linear operators on a Hilbert space into the real numbers are derived, an example being $A \mapsto -\text{Trace} \exp (L + \ln A)$. Some of these have applications to physics, specifically to the Wigner-Yanase-Dyson conjecture which is proved here and to the strong subadditivity of quantum mechanical entropy which will be proved elsewhere.

* On leave from the Department of Mathematics, M.I.T., Cambridge, Mass. 02139, U.S.A. Work partially supported by a Guggenheim Memorial Foundation fellowship and by U.S. National Science Foundation grant GP - 31674 X.

Janvier 1973

I.- INTRODUCTION

This paper is concerned with certain convex or concave mappings of linear operators on a Hilbert space into the reals. [$f(A)$ is convex if $f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda) f(B)$ for $0 < \lambda < 1$ and $f(A)$ is concave if $-f(A)$ is convex] . These mappings involve the trace operation which plays a central role in quantum statistical mechanics, and it is not surprising, therefore, that the mappings discussed here were motivated by considerations of physics. In particular, Theorem 1 solves affirmatively a conjecture due to Wigner, Yanase and Dyson [1] about a certain definition of information. In section III we use Theorem 1 to prove other convexity theorems when the Hilbert space is finite dimensional. One of those, Theorem 6, we extend to infinite dimensional spaces in section IV. Theorem 6 has a physical application ; it is the basis for proving that quantum mechanical entropy is strongly subadditive (cf. refs [2], [3] and [4]) . The proof of that fact will be given in a subsequent paper [5] .

From the work of Krauss and Bendat and Sherman ([6] and the references quoted therein) it is known that certain convex functions from $\mathbb{R}$ to $\mathbb{R}$ extend to operator-valued convex functions. If $f(x)$ is such a function then $A \mapsto \mathrm{Tr}Kf(A)$ (where Tr means trace) is certainly convex when $K > 0$ and fixed. Simple examples are $f(A) = A^{-p}$ and $f(A) = -A^{p}$ for $A > 0$ and $0 < p \leq 1$. However, $A \mapsto \mathrm{Tr}f(A)$ may be convex even when $f(A)$ is not convex as an operator-valued function. Examples of this are $f(A) = e^{A}$ for A selfadjoint and $f(A) = A^{-p}$ for $1 < p \leq 2$ and for $A > 0$ (cf. Theorem 8).

In this paper we shall be concerned with mappings more complicated than those just mentioned. One example, Theorem 6, is $A \mapsto -\mathrm{Tr}\exp[L + \ln A]$ for $A > 0$ and L selfadjoint.

Theorem 1 is our main theorem and Theorems 2,3,6 and 7 are derived from it. Theorems 8 and 9 are a side issue and are independent of and simpler than Theorem 1. In section V we remark briefly on the logical connection of Theorems 1,2,3,6 and 7, namely that they can all be derived simply from each other (at least for finite dimensional Hilbert spaces).

II.- THE MAIN THEOREM AND THE WIGNER-YANASE-DYSON PROBLEM.

We begin by proving our main Theorem 1 which constitutes the basis for Theorems 2,3, 6 and 7 of the next section. Theorem 1 is also the Wigner-Yanase-Dyson (WYD) conjecture [1] (actually, it is a bit stronger) and at the end of this section we shall explain the WYD problem. We also discuss another problem concerning the WYD definition of information [1] and give a partial solution of it.

Theorem 1 will be proved directly for infinite dimensional Hilbert spaces and our notation is the following :

(1) H is a separable Hilbert space with inner product (x,y) which is linear in y and conjugate linear in x.

(2) $\mathcal{B}(H)$ is the set of bounded linear operators from H to H ; $\mathcal{B}^s(H) \subset \mathcal{B}(H)$ are the bounded selfadjoint operators ; $\mathcal{B}^+(H) \subset \mathcal{B}^s(H)$ are the positive operators $(A \in \mathcal{B}^+(H) \Rightarrow (x,Ax) \geq 0 , \forall x)$; $\mathcal{B}^{++}(H) \subset \mathcal{B}^+(H)$ are the strictly positive operators $(A \in \mathcal{B}^{++}(H) \Rightarrow (x,Ax) > 0, \forall x)$.

(3) If $A \in \mathcal{B}^+(H)$ and $z \in \mathbb{C}$, we can use the spectral representation of A to define A^z. $z \mapsto A^z$ is entire analytic and $A^z \in \mathcal{B}^+(H)$ for $z \geq 0$.

(4) The $\mathcal{J}_q$ classes : If $A \in \mathcal{B}(H)$ we form $|A| = (A^\dagger A)^{1/2} \in \mathcal{B}^+(H)$. $A \in \mathcal{J}_q(H) \subset \mathcal{B}(H)$ $(q \geq 1)$ if $\|A\|_q \equiv (\mathrm{Tr}\ |A|^q)^{1/q} < \infty$, where Tr means trace. $\mathcal{J}_1(H)$ is the trace class and $\mathcal{J}_2(H)$ is the Hilbert-Schmidt class. $A \in \mathcal{J}_q(H)$ implies that A is compact and that

$A \in \mathcal{J}_1(H)$. $\|A\|_q = (\sum_{j=1}^{\infty} \lambda_j^{\,q})^{1/q}$ where the λ_j are the eigenvalues of $|A|$ in decreasing order, including multiplicity. If $A \in \mathcal{B}^+(H)$ but $A \notin \mathcal{J}_1(H)$, it is convenient to define $\mathrm{Tr}\, A = \infty$.

(5.) We recall that if $A \in \mathcal{B}(H)$ and if K is a linear operator (not necessarily bounded) on a dense domain, $D(K)$, in H then AK may have a bounded extension to all of H. If so, it is unique and its adjoint is a bounded extension of $K^\dagger A^\dagger$.

Theorem 1 : Let K be a linear operator (not necessarily bounded) on H, let $A,B \in \mathcal{B}^+(H)$ and let λ, $0 < \lambda < 1$, be given. Form the convex combination $C = \lambda A + (1-\lambda) B$. Let p and r be given positive real numbers with $p + r \equiv s \leq 1$. If $M \equiv C^{p/2} K C^{r/2}$ has an extension to $\mathcal{J}_2(H)$ then

(1) $A^{p/2} KA^{r/2}$ and $B^{p/2} KB^{r/2}$ have extensions to $\mathcal{J}_2(H)$ and

(2)
$$\lambda \operatorname{Tr} A^{r/2} K^\dagger A^p KA^{r/2} + (1-\lambda) \operatorname{Tr} B^{r/2} K^\dagger B^p KB^{r/2}$$
$$\leq \operatorname{Tr} C^{r/2} K^\dagger C^p KC^{r/2} \text{ , i.e.}$$
$$A \in \mathcal{B}^+(H) \mapsto \operatorname{Tr} A^{r/2} K^\dagger A^p KA^{r/2} \text{ is concave.}$$

Proof : (a) We recall the theorem [6] that the map $A \in \mathcal{B}^+(H) \mapsto A^q$ is concave on $\mathcal{B}^+(H)$ when $0 < q \leq 1$. Thus, $\lambda A^q \leq C^q$ and $\operatorname{Ker}(A^q) = \operatorname{Ker}(A) \supset \operatorname{Ker}(C) = \operatorname{Ker}(C^q)$, and similarly for B. As A,B and C are bounded, their kernels are closed subspaces and $H = \operatorname{Ker}(C) \oplus \operatorname{Ker}(C)^\perp$. The above inequalities show that for $0 < q \leq 1$, $\alpha(q) \equiv A^{q/2} \; C^{-q/2}$ and $\alpha(q)^\dagger = C^{-q/2} A^{q/2}$ can be extended to bounded operators on $\operatorname{Ker}(C)^\perp$ because $\| A^{q/2} C^{-q/2} \Psi \| \leq \lambda^{-1/2}$ for Ψ in the dense set D_C = {vectors with support away from zero in the spectral representation of C}. Similarly, we define $\beta(q) \equiv B^{q/2} C^{-q/2}$. Also, $\alpha(q)$ and $\alpha(q)^\dagger$ can be defined to be zero on $\operatorname{Ker}(C)$ and thus are defined on all of H. Clearly, $C^{q/2} \alpha(q)^\dagger = A^{q/2} = \alpha(q) C^{q/2}$. Consequently, $A^{p/2} KA^{r/2} = \alpha(p) [C^{p/2} KC^{r/2}] \alpha(r)^\dagger = \alpha(p) M\alpha(r)^\dagger \in \mathcal{J}_2(H)$, since $M \in \mathcal{J}_2(H)$. Not only is the first part of the theorem thus proved, but we also see that if $\{\Psi_i\}$ and $\{\varphi_i\}$ are orthonormal bases for $\operatorname{Ker}(C)$ and $\operatorname{Ker}(C)^\perp$ respectively, we can compute traces in the basis $\{\Psi_i\} + \{\varphi_i\}$ and all terms involving $\{\Psi_i\}$ will vanish. Thus, $\operatorname{Ker}(C)$

is an irrelevant subspace, and we shall henceforth assume that $H = \mathrm{Ker}(C)^{\perp}$, i.e. $C > 0$.

(b.) With the above definitions, part(2) is equivalent to the following : $\lambda T^A(p,r) + (1-\lambda) T^B(p,r) \le \mathrm{Tr} M^\dagger M$ for every $M \in \mathcal{J}_2(H)$, where $T^A(p,r) = \mathrm{Tr}\, \alpha(r)\ M^\dagger \alpha(p)^\dagger\ \alpha(p)\ M\alpha(r)^\dagger$ and similarly for $T^B(p,r)$.

(c.) Let $z = x + iy \in \mathbb{C}$ and consider the operator valued function $\alpha(z) \equiv A^{z/2} C^{-z/2} = A^{iy/2} A^{x/2} C^{-x/2} C^{-iy/2} = A^{iy/2} \alpha(x) C^{-iy/2}$. Since $C^{iy/2}$ is unitary and $\| A^{iy/2} \| \le 1$, $\alpha(z)$ is uniformly bounded in $S = \{z \mid 0 \le \mathrm{Re}(z) \le 1\}$. If $\bar{z} = x - iy$, $\alpha(\bar{z})^\dagger = C^{-z/2} A^{z/2}$. For $\Psi \in D_C$, $\alpha(z)\Psi$ is an entire analytic function of z because $C^{-z/2}\Psi$ is entire and $A^{z/2}$ is entire. Hence, by the boundedness of $\alpha(z)$ and a standard density argument, $\alpha(z)\Psi$ is regular on S (continuous on S and analytic in the interior of S) for all $\Psi \in H$. Since weak analyticity implies strong analyticity, we also have that $\alpha(z)$ is strongly continuous on S and is norm analytic in the interior of S . Furthermore, if $A_n \to A$ strongly and if $B \in \mathcal{J}_2(H)$, then $A_n B \to AB$ in the $\mathcal{J}_2(H)$ norm . (This is trivial if B is finite rank, but the finite rank operators are dense in the $\mathcal{J}_2(H)$ norm). Hence, $\alpha(z_1) M\alpha(\bar{z}_2)^\dagger$ is $\mathcal{J}_2(H)$ regular on $S \times S$, which means that $T^A(z_1,z_2) \equiv \mathrm{Tr}\, \alpha(z_2) M^\dagger \alpha(\bar{z}_1)^\dagger \alpha(z_1) M\alpha(\bar{z}_2)^\dagger$ is bounded and regular on $S \times S$.

(d.) We now set $z_1 = z$, $z_2 = s - z$ and consider $T^A(z) \equiv T^A(z, s-z)$ as a regular function on $\{z \mid 0 \le \mathrm{Re}(z) \le s\}$. By (b.) we need to show that $f(p) = \lambda T^A(p) + (1-\lambda) T^B(p) \le \mathrm{Tr}\, M^\dagger M$. By the maximum modulus principle for bounded regular functions on a strip, $|f(p)| \le \max \{\sup_\theta |f(i\theta)|, \ \sup_\theta |f(s+i\theta)|\}$. We shall consider only the first case, $p = i\theta$, in detail because the second case, $p = s+i\theta$, is parallel. $|f(i\theta)| \le \lambda |T^A(i\theta)| + (1-\lambda)|T^B(i\theta)|$. Using the facts that for $A \in \mathcal{B}(H)$ and $B \in \mathcal{J}_2(H)$, AB and $BA \in \mathcal{J}_2(H)$, and for $B, C \in \mathcal{J}_2(H)$, $\operatorname{Tr} BC = \operatorname{Tr} CB$ and $|\operatorname{Tr} BC| \le \frac{1}{2}\operatorname{Tr} B^\dagger B + \frac{1}{2}\operatorname{Tr} C^\dagger C$, we have that $2|T^A(i\theta)| \le \operatorname{Tr} \alpha(s-i\theta) M^\dagger \alpha(-i\theta)^\dagger \alpha(-i\theta) M \alpha(s-i\theta)^\dagger$

$+ \operatorname{Tr} \alpha(s+i\theta) M^\dagger \alpha(i\theta)^\dagger \alpha(i\theta) M \alpha(s+i\theta)^\dagger$.

However, $\|\alpha(-i\theta)^\dagger \alpha(-i\theta)\| \le 1$, so the first term is at most $\operatorname{Tr} \alpha(s-i\theta) M^\dagger M \alpha(s-i\theta)^\dagger = \operatorname{Tr} M \alpha(s-i\theta)^\dagger \alpha(s-i\theta) M^\dagger$
$= \operatorname{Tr} M C^{-i\theta/2} \alpha(s)^\dagger \alpha(s) C^{i\theta/2} M^\dagger = \operatorname{Tr} \alpha(s)^\dagger \alpha(s) C^{i\theta/2} M^\dagger M C^{-i\theta/2}$.
Likewise, the second term is at most $\operatorname{Tr} \alpha(s)^\dagger \alpha(s) C^{-i\theta/2} M^\dagger M C^{i\theta/2}$.
If we add to these the corresponding two terms for $|T^B(i\theta)|$ we obtain

$$\lambda|T^A(i\theta)| + (1-\lambda)|T^B(i\theta)| \le \tfrac{1}{2} \operatorname{Tr}[\lambda\alpha^\dagger(s)\alpha(s) + (1-\lambda)\beta^\dagger(s)\beta(s)]\, P, \qquad (2.1)$$

where $P = C^{i\theta/2} M^\dagger M C^{-i\theta/2} + C^{-i\theta/2} M^\dagger M C^{i\theta/2} \in \mathcal{B}^+(H)$. As we remarked before, $\lambda A^s + (1-\lambda) B^s \le C^s$, whence
$\lambda\alpha(s)^\dagger \alpha(s) + (1-\lambda)\beta(s)^\dagger \beta(s) = C^{-s/2}[\lambda A^s + (1-\lambda) B^s] C^{-s/2} \le \mathbb{1}$.
Substituting this in (2.1) proves the theorem. Q.E.D.

<u>REMARK</u> : If $C^{p/2}K$ has an extension to $\mathcal{J}_1(H)$ then so does $A^{p/2}K$ and $B^{p/2}K$ since $\|C^{p/2}K\| \ge \lambda^{1/2} \|A^{p/2}K\|$. In this case $\operatorname{Tr} C^{r/2} K^\dagger C^p K C^{r/2} = \operatorname{Tr} C^r K^\dagger C^p K = \operatorname{Tr} K^\dagger C^p K C^r$ and similarly for A and B.

<u>COROLLARY 1.1</u> : With p and r as in Theorem 1, the function from $\mathcal{B}^+(H) \times \mathcal{B}^+(H) \times \mathcal{B}(H)$ to the nonnegative reals defined by $(A,B,K) \mapsto F(A,B,K) = \mathrm{Tr}\, A^{r/2} K^\dagger B^p KA^{r/2}$

(1) is jointly concave in (A,B)

(2) is convex in K .

<u>Proof</u> : Consider the Hilbert space $H' \equiv H \oplus H$ and define the following operators in $\mathcal{B}(H')$:

$$k : (x,y) \mapsto (0,Kx)$$
$$k^\dagger : (x,y) \mapsto (K^\dagger y,0)$$
$$a : (x,y) \mapsto (Ax,By) \ , \ a \in \mathcal{B}^+(H') \quad .$$

Applying Theorem 1 to $\mathrm{Tr}\, a^{r/2} k^\dagger a^p . k\, a^{r/2}$ proves the first part. The second part follows from a Schwartz inequality type of argument since $F(A,B,K)$ is nonnegative and quadratic in K . Q.E.D.

<u>COROLLARY 1.2</u> : With p and r as in Theorem 1, $p + r \equiv s \leq 1$, the functions from $\mathcal{B}^+(H) \times \mathcal{B}^+(H) \times \mathcal{B}(H)$ to the nonnegative reals defined by

$$(A,B,K) \mapsto F_q(A,B,K) = \{\mathrm{Tr}\, A^{r/2} K^\dagger B^p KA^{r/2}\}^q$$

(1) are jointly concave in (A,B) when $0 < q \leq 1/s$

(2) are jointly convex in (A,B) when $q < 0$

(3) are convex in K when $q \geq 1/2$.

Proof : The proof is a standard one for homogeneous concave (or convex functions [7] . Let $x = (x_1, x_2) \in \mathbb{R}_+^2 = \{(x_1, x_2) | x_1 \geq 0, x_2 \geq 0\}$ and define $f(x) = F_1(x_1 A + x_2 A', x_1 B + x_2 B', K)$ for an arbitrary , but henceforth fixed, choice of $A , A' , B , B' \in \mathfrak{B}^+(H)$.
Parts (1) and (2) are equivalent to showing that for all such choices, $f(x)^q$ is concave (or convex). By Corollary 1.1 $f(x)$ is nonnegative, concave and homogeneous of order s, i.e. $f(\lambda x) = \lambda^s f(x)$ for $\lambda \geq 0$.
For each $\alpha \geq 0$, define $G_\alpha = \{ x | f(x) \geq \alpha , x \in \mathbb{R}_+^2 \}$. It is easily seen from the properties of $f(x)$ that G_α is a convex subset of $\mathbb{R}_+^2$ and $G_\alpha = \alpha^{1/s} G_1$ for $\alpha > 0$. Define
$k(x) = \sup \{ \mu \geq 0 \mid x \in G_{\mu^s} \}$ for $x \in \mathbb{R}_+^2$. As $x \in G_{f(x)}$, $k(x)$ is everywhere defined. In fact, since $f(x) = \sup \{\alpha \geq 0 | x \in G_\alpha\}$, $f(x) = k(x)^s$. Obviously, $k(x)$ is nonnegative and homogeneous of order one and, since $k(x) = \sup \{ \mu > 0 \mid x \in \mu G_1\}$ when $f(x) \neq 0$, it is easy to check that $k(x)$ is a concave function. For a nonnegative concave function, $k(x)$, $k(x)^p$ is concave when $0 < p \leq 1$ and $k(x)^p$ is convex when $p < 0$. This proves parts (1) and (2). For part (3) we define $f(x) = F_1(A,B, x_1 K + x_2 K')$, with $K, K' \in \mathfrak{B}(H)$.
$f(x)$ is nonnegative, convex and homogeneous of order 2. We define :
$G_\alpha = \{x | f(x) \leq \alpha , x \in \mathbb{R}_+^2 \}$ which is convex ; $k(x) = \inf\{\mu \geq 0 | x \in G_{\mu^2}\}$.
Then $k(x)$ is nonnegative, convex and homogeneous of order one and $f(x) = k(x)^2$. For any nonnegative convex function, $k(x)$, $k(x)^p$ is convex when $p \geq 1$. Q.E.D.

The setting for the next corollary is the following :
Let H^1 and H^2 be two separable Hilbert spaces and $H^{12} \equiv H^1 \otimes H^2$ their tensor product. If $A_{12} \in \mathfrak{B}^+(H^{12})$ and $A_{12} \in \mathcal{J}_1(H^{12})$ we can define $A_1 \in \mathfrak{B}^+(H^1)$ by means of the partial trace, i.e.

[8.] The idea of representing A_1 in this manner is due to A. Uhlmann, Endlich Dimensionale Dichtematrizen II, preprint, submitted to Wissen. Z. Karl-Marx-Univ., Leipzig (1972).

[9.] P. Bautaann, and R. Jost, Remarks on a Conjecture of Robinson and Ruelle Concerning the Quantum Mechanical Entropy, in "Problems of Theoretical Physics ; Essays Devoted to N.N. Bogoliubov", (1969), 285-293, **MOBCOW**, Nauka.

[10.] F. Baumann, Bemerkungen Ueber Quantenmechanische Entropie Ungleichungen, Helv. Phys. Acta 44 (1971), 95-100.

[ll.] S. Golden, Lower Bounds for the Helmhotz Function , Phys. Rev. 137 (1965) B 1127-1128.

[12.] C.J. Thompson, Inequality with Applications in Statistical Mechanics, J. Math. Phys. 6(1965), 1812-1813.

[13.] M.B. Ruskai, Inequalities for Traces on Von Neumann Algebras, Commun. Math. Phys. 26 (1972), 280-289.