Partial traces and entropy inequalities

Rajendra Bhatia

Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi 110016. India

Abstract

The partial trace operation and the strong subadditivity property of entropy in quantum mechanics are explained in linear algebra terms.

Keywords: Quantum entropy; Density matrix; Partial trace; Operator convexity; Matrix inequalities

1. Introduction

There is an interesting matrix operation called *partial trace* in the physics literature. The first goal of this short expository paper is to connect this operation to others more familiar to linear algebraists. The second goal is to use this connection to present a slightly simpler proof of an important theorem of Lieb and Ruskai [11,12] called the strong subadditivity (SSA) of quantum-mechanical entropy. Closely related to SSA, and a crucial ingredient in one of its proofs, is another theorem of Lieb [9] called Lieb's concavity theorem (solution of the Wigner–Yanase–Dyson conjecture). An interesting alternate formulation and proof of this latter theorem appeared in a paper of Ando [1] well-known to linear algebraists. Discussion of the Lieb–Ruskai theorem, however, seems to have remained confined to the physics literature. We believe there is much of interest here for others as well, particularly for those interested in linear algebra.

2. The partial trace

Let $\mathscr{H}$ be a finite-dimensional Hilbert space with inner product $\langle .\,,\,.\rangle$ and let $\mathscr{L}(\mathscr{H})$ be the space of linear operators on $\mathscr{H}$. If A is positive semi-definite $(A \geqslant O)$ and has trace one $(\operatorname{tr} A = 1)$ we say A is a *density matrix.*

Let $\mathscr{H}_1$, $\mathscr{H}_2$ be two Hilbert spaces of dimensions n, m respectively. Let $\mathscr{H}_1 \otimes \mathscr{H}_2$ be their tensor product. The *partial trace* $\operatorname{tr}_{\mathscr{H}_2}$ is a linear map from $\mathscr{L}(\mathscr{H}_1 \otimes \mathscr{H}_2)$ into $\mathscr{L}(\mathscr{H}_1)$ defined as follows. Let $e_1, \dots, e_m$ be an orthonormal basis for $\mathscr{H}_2$. If $A \in \mathscr{L}(\mathscr{H}_1 \otimes \mathscr{H}_2)$, then $\operatorname{tr}_{\mathscr{H}_2} A$ is the linear operator A_1 on $\mathscr{H}_1$ defined by the relation

$$\langle x, A_1 y\rangle = \sum_{j=1}^{m} \langle x \otimes e_j,\ A(y \otimes e_j)\rangle, \tag{1}$$

for all $x, y \in \mathscr{H}_1$.

It is easy to see that A_1 is well-defined (independent of the choice of the orthonormal basis $\{e_j\}$). The partial trace $A_2 = \operatorname{tr}_{\mathscr{H}_1} A$ is defined analogously, and is an operator on $\mathscr{H}_2$.

If A is positive, then so are its partial traces A_1, A_2; and if A is a density matrix, then so are A_1, A_2.

An operator A on $\mathscr{H}_1 \otimes \mathscr{H}_2$ is said to be *decomposable* if it can be factored as $A = A_1 \otimes A_2$ where A_1, A_2 are operators on $\mathscr{H}_1$, $\mathscr{H}_2$, respectively. If A_1, A_2 are density matrices, then so is their tensor product $A = A_1 \otimes A_2$; and in this case $A_1 = \operatorname{tr}_{\mathscr{H}_2} A$, $A_2 = \operatorname{tr}_{\mathscr{H}_1} A$. (The conditions $\operatorname{tr} A_j = 1$ are needed for this.)

Let A be any operator or $\mathscr{H}_1 \otimes \mathscr{H}_2$ with partial traces A_1, A_2 and let B be a decomposable operator of the form $B_1 \otimes I$. Then one can see that

$$\operatorname{tr} AB = \operatorname{tr} A_1 B_1. \tag{2}$$

For this choose orthonormal bases $f_1, \dots, f_n$ and $e_1, \dots, e_m$ for $\mathscr{H}_1$ and $\mathscr{H}_2$ respectively and observe that

$$\begin{aligned}
\operatorname{tr} AB &= \sum_{i,j} \langle f_i \otimes e_j, AB(f_i \otimes e_j)\rangle \\
&= \sum_{i,j} \langle f_i \otimes e_j, A(B_1 f_i \otimes e_j)\rangle \\
&= \sum_{i} \langle f_i, A_1 B_1 f_i\rangle \\
&= \operatorname{tr} A_1 B_1.
\end{aligned}$$

The relation (2) characterises the partial trace operation $A \mapsto A_1$, and may be taken as a definition instead of (1). (See the comprehensive review article by Wehrl [22, p. 242]. The definition given there restricts itself to A being a density matrix; the partial trace A_1 is then called the *reduced density matrix*.)

The third definition of partial trace that we give now seems more revealing. Let A be any operator on $\mathscr{H}_1 \otimes \mathscr{H}_2$ and write its matrix representation in a fixed orthonormal basis $f_i \otimes e_j$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$. Partition this matrix into an $n \times n$ block form

$$A = [A_{ij}] \quad 1 \leqslant i, j \leqslant n, \tag{3}$$

where each A_{ij} is an $m \times m$ matrix. Then the partial trace A_1 of A is the $n \times n$ matrix

$$A_1 = [\operatorname{tr} A_{ij}] \quad 1 \leqslant i, j \leqslant n, \tag{4}$$

i.e., the partial trace is obtained by replacing the $m \times m$ *matrix* A_{ij} in the decomposition (3) by the *number* $\operatorname{tr} A_{ij}$.

The partial trace A_2 is defined analogously: split A into an $m \times m$ block matrix with $n \times n$ blocks and then replace each block by its trace.

If B is a decomposable operator of the form $B = B_1 \otimes I$ and if $B_1 = \left[b_{ij}\right]$ is the matrix of B_1 in the basis $f_1, \ldots, f_n$, then B can be written in $n \times n$ block matrix form as

$$B = \left[b_{ij} I\right] \quad 1 \leqslant i, j \leqslant n.$$

From this one sees that if A and A_1 are as in (3) and (4), then

$$\operatorname{tr} AB = \sum_{i,j} b_{ij} \operatorname{tr} A_{ji} = \operatorname{tr} A_1 B_1.$$

Thus this definition of the operation $A \mapsto A_1$ leads to the same object as before. Next we decompose this map as a composite of three maps.

Let $\omega = \mathrm{e}^{2\pi \mathrm{i}/m}$ be the primitive mth root of unity. Let U be the $m \times m$ unitary diagonal matrix

$$U = \operatorname{diag}(1, \omega, \omega^2, \ldots, \omega^{m-1}).$$

Let T be any $m \times m$ matrix and let $\mathscr{D}(T)$, the diagonal part of T, be the matrix obtained from T by replacing all its off-diagonal entries by zeros. Then following the ideas in [4] we write

$$\mathscr{D}(T) = \frac{1}{m} \sum_{k=0}^{m-1} U^{*k} T U^k.$$

Next let

$$W = U \oplus U \oplus \cdots \oplus U \quad (n \text{ copies}).$$

Let A be any operator on $\mathscr{H}_1 \otimes \mathscr{H}_2$ and let

$$\Phi_1(A) = \frac{1}{m} \sum_{k=0}^{m-1} W^{*k} A W^k. \tag{5}$$

If the matrix of A is partitioned as in (3) then

$$\Phi_1(A) = [\mathscr{D}(A_{ij})]. \tag{6}$$

Let V be the $m \times m$ cyclic permutation matrix acting as $Ve_j = e_{j+1}$, $1 \leqslant j \leqslant m$, where $e_{m+1} = e_1$. This is the matrix

$$V = \begin{bmatrix} 0 & 0 & \cdots & & 1 \\ 1 & 0 & & & 0 \\ 0 & 1 & & & 0 \\ & & & & \\ & & & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

If D is an $m \times m$ diagonal matrix, then

$$\frac{1}{m}\sum_{k=0}^{m-1} V^{*k} D V^k = \frac{1}{m}\operatorname{diag}(\operatorname{tr} D, \ldots, \operatorname{tr} D),$$

a diagonal matrix with all its diagonal entries equal. Let

$$X = V \oplus V \oplus \cdots \oplus V \qquad (n \text{ copies}),$$

and for $A \in \mathscr{L}(\mathscr{H}_1 \otimes \mathscr{H}_2)$ let

$$\Phi_2(A) = \frac{1}{m}\sum_{k=0}^{m-1} X^{*k} A X^k. \tag{7}$$

If A is partitioned as in (3), we have

$$\Phi_2 \circ \Phi_1(A) = \left[\frac{1}{m}(\operatorname{tr} A_{ij}) I_m\right], \tag{8}$$

where I_m is the $m \times m$ identity matrix.

Now let $T^{(1,1)}$ denote the (1,1) entry of a matrix T. If A is an $nm \times nm$ partitioned matrix as in (3), let $\Phi_3(A)$ be the $n \times n$ matrix defined as

$$\Phi_3(A) = m[A_{ij}^{(1,1)}]. \tag{9}$$

We have then

$$\operatorname{tr}_{\mathscr{H}_2} A = \Phi_3 \circ \Phi_2 \circ \Phi_1(A). \tag{10}$$

The expressions (5) and (7) for Φ_1, Φ_2 clearly display them as "averaging operations". The map Φ_3 is, upto a constant multiple, picking out a principal submatrix. This too has an interpretation as an averaging operation [3,6]. All three are completely positive maps [5], Φ_1, Φ_2 and $\Phi_3 \circ \Phi_2 \circ \Phi_1$ are trace-preserving.

The partial trace (third definition) and its generalisations have been studied in the matrix literature, though not under this name. For example, de Pillis [7] has shown that the partial trace of a positive (semi-definite) matrix is positive. Generalisations, where the blocks are replaced not by their traces but by other functions may be found in [7,15,16] and in references given therein.

3. Entropy inequalities

Let A be a density matrix. The (von Neumann) *entropy* of A is the non-negative number

$$S(A) = -\mathrm{tr}(A \log A). \tag{11}$$

One of its basic properties is *concavity* as a function of A:

$$S\left(\frac{A+B}{2}\right) \geqslant \frac{S(A)+S(B)}{2}. \tag{12}$$

It is not difficult to prove this; see e.g., [2, Problem IX.8.14]. In fact, a much stronger statement follows from Loewner's theory of operator monotone functions. The function $f(t) = -t \log t$ is operator concave on $(0, \infty)$ (see [2, Exercise V.2.13]).

Some deeper properties of entropy were proved by Lieb and Ruskai in 1973. Of these a few have found their way into the matrix theory literature [1,2]. We will touch upon these briefly on way to the Lieb–Ruskai theorem on SSA.

Let A be a density matrix and K any self-adjoint operator. For $0 < t < 1$ let

$$S_t(A, K) = \frac{1}{2} \mathrm{tr}\, [A^t, K][A^{1-t}, K], \tag{13}$$

where $[X, Y]$ stands for the commutator $XY - YX$. The quantity (13) is a measure of non-commutativity of A and K, and is called *skew-entropy*. This too is a concave function of A, a consequence of a more general theorem of Lieb [9]:

Lieb's Concavity Theorem. *The function*

$$f(A, B) = \mathrm{tr}\, X^* A^t X B^{1-t} \tag{14}$$

of positive matrices A, B is jointly concave for each matrix X and for $0 \leqslant t \leqslant 1$.

Using the familiar identification of $\mathscr{L}(\mathscr{H})$ with $\mathscr{H} \otimes \mathscr{H}^*$, this statement can be reformulated as: the function

$$g(A, B) = A^t \otimes B^{1-t} \tag{15}$$

of positive matrices A, B is jointly concave for $0 \leqslant t \leqslant 1$.

This formulation, and a proof of it, were given by Ando [1]. Other proofs of Lieb's theorem that predate Ando's include ones by Epstein [8], Uhlmann [21] and Simon [20].

Another quantity of interest is the *relative entropy*

$$S(A|B) = \mathrm{tr}\, A(\log A - \log B) \tag{16}$$

associated with a pair of density matrices A, B.

If f is any convex function on the real line, then for Hermitian matrices A, B

$$\mathrm{tr}\, [f(A) - f(B)] \geqslant \mathrm{tr}\, [(A - B) f'(B)].$$

This inequality (called Klein's inequality) applied to the function $f(t) = t \log t$ on the positive half-line shows that for positive matrices A, B,

$$\operatorname{tr} A(\log A - \log B) \geqslant \operatorname{tr}(A - B).$$

If A, B are density matrices, then $\operatorname{tr}(A - B) = 0$, and hence

$$S(A|B) \geqslant 0. \tag{17}$$

It is an easy corollary of Lieb's concavity theorem that

$$S(A|B) \text{ is jointly convex in } A, B. \tag{18}$$

(Choose $X = I$ and differentiate the function (14) at $t = 0$.)

Finally, we come to the properties of $S(A)$ and $S(A|B)$ related to partial traces.

Let A_1, A_2 be density matrices. It is easy to see that S is *additive* over tensor products:

$$S(A_1 \otimes A_2) = S(A_1) + S(A_2). \tag{19}$$

Now let A be a density matrix on $\mathscr{H}_1 \otimes \mathscr{H}_2$ and let A_1, A_2 be its partial traces. The *subadditivity* property of S says that

$$S(A) \leqslant S(A_1) + S(A_2). \tag{20}$$

This can be proved as follows. From (17) we have

$$\begin{aligned} 0 &\leqslant S(A|A_1 \otimes A_2) \\ &= \operatorname{tr} A(\log A - \log(A_1 \otimes A_2)) \\ &= \operatorname{tr} A(\log A - \log(A_1 \otimes I) - \log(I \otimes A_2)) \\ &= \operatorname{tr}(A \log A - A_1 \log A_1 - A_2 \log A_2) \quad \text{using (2)} \\ &= -S(A) + S(A_1) + S(A_2). \end{aligned}$$

From the definition (16) it is obvious that

$$S(UAU^*|UBU^*) = S(A|B)$$

for every unitary matrix U. Hence, from the representations (5) and (7) we see that

$$S(\Phi_2 \circ \Phi_1(A)|\Phi_2 \circ \Phi_1(B)) \leqslant S(\Phi_1(A)|\Phi_1(B)) \leqslant S(A|B).$$

Note that $\Phi_2 \circ \Phi_1(A)$ is a matrix of the special form (8). It is easy to see that

$$S(\Phi_3 \circ \Phi_2 \circ \Phi_1(A)|\Phi_3 \circ \Phi_2 \circ \Phi_1(B)) = S(\Phi_2 \circ \Phi_1(A)|\Phi_2 \circ \Phi_1(B)).$$

Thus we have

$$S(A_1|B_1) \leqslant S(A|B). \tag{21}$$

A more general result is known [13,14] and can be proved using our techniques: if Φ is any completely positive, trace-preserving map, then

$$S(\Phi(A)|\Phi(B)) \leqslant S(A|B).$$

Now consider a tensor product $\mathscr{H}_1 \otimes \mathscr{H}_2 \otimes \mathscr{H}_3$ of three Hilbert spaces. For simplicity of notation let us use the notation A_{123} for an operator on this Hilbert space, and drop the index j when we take a partial trace $\operatorname{tr}_{\mathscr{H}_j}$. Thus $\operatorname{tr}_{\mathscr{H}_3} A_{123} = A_{12}$, $\operatorname{tr}_{\mathscr{H}_1} A_{12} = A_2$, etc.

SSA of entropy is the following statement.

Theorem (Lieb–Ruskai). *Let A_{123} be a density matrix on $\mathscr{H}_1 \otimes \mathscr{H}_2 \otimes \mathscr{H}_3$. Then*

$$S(A_{123}) + S(A_2) \leqslant S(A_{12}) + S(A_{23}). \tag{22}$$

Proof. Using the diminishing property (21) with respect to the partial trace $\operatorname{tr}_{\mathscr{H}_3}$ we see that

$$S(A_{12}|A_1 \otimes A_2) \leqslant S(A_{123}|A_1 \otimes A_{23}).$$

We have seen while proving (20) that $S(T|T_1 \otimes T_2) = -S(T) + S(T_1) + S(T_2)$. So the above inequality can be rewritten as

$$-S(A_{12}) + S(A_1) + S(A_2) \leqslant -S(A_{123}) + S(A_1) + S(A_{23}),$$

and on rearranging terms, as (22). □

The reader should note that the inequality (22) is similar in form to others that look like

$$S(E \cup F) + S(E \cap F) \leqslant S(E) + S(F).$$

As we said in the introduction, our emphasis here has been on linear algebra and matrix inequalities. To understand the importance of these results in physics the reader should turn to the original papers of Lieb and Ruskai, and to the review articles by Lieb [10] and by Wehrl [22,23] where references to other works (including alternate proofs and extensions) may be found. See also the books [17,18] and the recent article [19].

Acknowledgements

The author thanks Centro de Estruturas Lineares e Combinatórias, Universidade de Lisboa, for its hospitality in June 2001 when this work was done. He thanks J.A. Dias da Silva, R. Horn, E.H. Lieb and M.B. Ruskai for comments and references to the literature.

References

[1] T. Ando, Concavity of certain maps on positive definite matrices and applications to Hadamard products, Linear Algebra Appl. 26 (1979) 203–241.
[2] R. Bhatia, Matrix Analysis, Springer, 1997.

[3] R. Bhatia, Pinching, trimming, truncating, and averaging of matrices, Amer. Math. Monthly 107 (2000) 602–608.
[4] R. Bhatia, C. Davis, M.-D. Choi, Comparing a matrix to its off-diagonal part, Oper. Theory Adv. Appl. 40 (1989) 151–164.
[5] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. 10 (1975) 285–290.
[6] C. Davis, Various averaging operations onto subalgebras, Illinois J. Math. 3 (1959) 538–553.
[7] J. de Pillis, Transformations on partitioned matrices, Duke Math. J. 36 (1969) 511–515.
[8] H. Epstein, Remarks on two theorems of Lieb, Commun. Math. Phys. 31 (1973) 317–325.
[9] E.H. Lieb, Convex trace functions and the Wigner–Yanase–Dyson conjecture, Adv. in Math. 11 (1973) 267–288.
[10] E.H. Lieb, Some convexity and subadditivity properties of entropy, Bull. Amer. Math. Soc. 81 (1975) 1–13.
[11] E.H. Lieb, M.B. Ruskai, A fundamental property of quantum-mechanical entropy, Phys. Rev. Lett. 30 (1973) 434–436.
[12] E.H. Lieb, M.B. Ruskai, Proof of the strong subadditivity of quantum-mechanical entropy, J. Math. Phys. 14 (1973) 1938–1941.
[13] G. Lindblad, Expectations and entropy inequalities for finite quantum systems, Commun. Math. Phys. 39 (1974) 111–119.
[14] G. Lindblad, Completely positive maps and entropy inequalities, Commun. Math. Phys. 40 (1975) 147–151.
[15] M. Marcus, W. Watkins, Partitioned Hermitian matrices, Duke Math. J. 38 (1971) 237–249.
[16] R. Merris, Trace functions I, Duke Math. J. 38 (1971) 527–530.
[17] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2000.
[18] M. Ohya, D. Petz, Quantum Entropy and Its Use, Springer, 1993.
[19] M.B. Ruskai, Inequalities for quantum entropy: a review with conditions for equality, preprint, 2002.
[20] B. Simon, Trace Ideals and Their Applications, Cambridge University Press, 1979.
[21] A. Uhlmann, Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory, Commun. Math. Phys. 54 (1977) 21–32.
[22] A. Wehrl, General properties of entropy, Rev. Modern Phys. 50 (1978) 221–260.
[23] A. Wehrl, The many facets of entropy, Rep. Math. Phys. 30 (1991) 119–129.