Contemporary Mathematics

Trace Inequalities and Quantum Entropy: An Introductory Course

Eric Carlen

Abstract. We give an elementary introduction to the subject of trace inequalities and related topics in analysis, with a special focus on results that are relevant to quantum statistical mechanics. This introductory course was presented at *Entropy and the Quantum: A school on analytic and functional inequalities with applications*, Tucson, Arizona, March 16-20, 2009.

Contents

1991 *Mathematics Subject Classification.* 47A63, 15A90.
Key words and phrases. convexity, concavity, trace inequality, entropy, operator norms.
Work partially supported by U.S. National Science Foundation grant DMS 09-01632.

1. Introduction

1.1. Basic definitions and notation. Let $\mathbf{M}_n$ denote the space of $n \times n$ matrices, which we sometimes refer to as *operators* on $\mathbb{C}^n$. The inner product on $\mathbb{C}^n$, and other Hilbert spaces, shall be denoted by $\langle \cdot, \cdot \rangle$, conjugate linear on the left, and the *Hermitian conjugate*, or simply *adjoint* of an $A \in \mathbf{M}_n$ shall be denoted by A^*.

Let $\mathbf{H}_n$ denote the $n \times n$ Hermitian matrices, i.e.; the subset of $\mathbf{M}_n$ consisting of matrices A such that $A^* = A$. There is a natural *partial order* on $\mathbf{H}_n$: A matrix $A \in \mathbf{H}_n$ is said to be *positive semi-definite* in case

$$\langle v, Av \rangle \geq 0 \qquad \text{for all } v \in \mathbb{C}^n \ , \tag{1.1}$$

in which case we write $A \geq 0$. A is said to *positive definite* in case the inequality in (1.1) is strict for all $v \neq 0$ in $\mathbb{C}^n$, in which case we write $A > 0$. Notice that in the finite dimensional case we presently consider, $A > 0$ if and only $A \geq 0$ and A is invertible.

By the Spectral Theorem, $A \geq 0$ if and only if all of the eigenvalues of A are non-negative, which is the case if and only if there is some $B \in \mathbf{M}_n$ such that $A = B^*B$.

Finally, we partially order $\mathbf{H}_n$ by defining $A \geq B$ to mean that $A - B \geq 0$. We shall also write $A > B$ to mean that $A - B > 0$. Let $\mathbf{H}_n^+$ denote the $n \times n$ *positive definite* matrices.

For $A \in \mathbf{M}_n$, the *trace* of A, $\mathrm{Tr}(A)$, is defined by

$$\mathrm{Tr}(A) = \sum_{j=1}^{n} A_{j,j} \ .$$

For any $A, B \in \mathbf{M}_n$,

$$\mathrm{Tr}(AB) = \sum_{i,j=1}^{n} A_{i,j} B_{j,i} = \sum_{i,j=1}^{n} B_{j,i} A_{i,j} = \mathrm{Tr}(BA) \ . \tag{1.2}$$

This is known as *cyclicity of the trace*. It tells us, for example that if $\{u_1, \ldots, u_n\}$ is any orthonormal basis for $\mathbb{C}^n$, then

$$\mathrm{Tr}(A) = \sum_{j=1}^{n} \langle u_j, Au_j \rangle \ . \tag{1.3}$$

Indeed if U is the unitary matrix whose jth column is u_j, $(U^*AU)_{j,j} = \langle u_j, Au_j \rangle$, and then by (1.2), $\mathrm{Tr}(U^*AU) = \mathrm{Tr}(AUU^*) = \mathrm{Tr}(A)$. Thus, $\mathrm{Tr}(A)$ is a *unitarily invariant* function of A, and as such, depends only on the eigenvalues of A. In fact, taking $\{u_1, \ldots, u_n\}$ to be an orthonormal basis of $\mathbb{C}^n$ with $Au_j = \lambda_j u_j$, $j = 1, \ldots, n$, (1.2) yields

$$\mathrm{Tr}(A) = \sum_{j=1}^{n} \lambda_j \ . \tag{1.4}$$

An $n \times n$ *density matrix* is a matrix $\rho \in \mathbf{H}_n^+$ with $\mathrm{Tr}(\rho) = 1$. The symbols ρ (and σ) are traditional for density matrices, and they are the quantum mechanical analogs of *probability densities*, and they are in one-to-one correspondence with the set of *states* of a quantum mechanical system whose observables are self adjoint operators on $\mathbb{C}^n$.

Let $\mathbf{S}_n$ denote the set of density matrices on $\mathbb{C}^n$. This is a convex set, and it is easy to see that the extreme points of $\mathbf{S}_n$ are precisely the rank one orthogonal projections on $\mathbb{C}^n$. These are called *pure states*.

Of course in many, if not most, quantum mechanical systems, the observables are operators on an infinite dimensional, but separable, Hilbert space $\mathcal{H}$. It is easy to extend the definition of the trace, and hence of density matrices, to this infinite dimensional setting. However, it is not hard to show that any positive semi-definite operator ρ on $\mathcal{H}$ with $\mathrm{Tr}(\rho) = 1$ is a compact operator, and thus it may be approximated in the operator norm by a finite rank operator. Simon's book [**31**] contains a very elegant account of all this. Here we simply note that essentially for this reason, the essential aspects of the inequalities for density matrices that we study here are contained in the finite dimensional case, to which we restrict our attention for the most part of these notes.

1.2. Trace inequalities and entropy. Much of what we discuss here is directly related to some notion of *entropy.*

1.1. DEFINITION. The *von Neuman entropy* of $\rho \in \mathbf{S}_n$, $S(\rho)$, is defined by

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) \ . \tag{1.5}$$

The operator $\rho \log \rho \in \mathbf{H}_n$ is defined using the Spectral Theorem. This says that for every self adjoint operator A on $\mathbb{C}^n$, there exists an orthonormal basis $\{u_1, \dots, u_n\}$ consisting of eigenvectors of A, meaning that for some real numbers $\{\lambda_1, \dots, \lambda_n\}$, $Au_j = \lambda_j u_j$ for $j = 1, \dots, n$. The set $\{\lambda_1, \dots, \lambda_n\}$ is the *spectrum* of A, and the numbers in it are the *eigenvalues* of A.

Given such a basis of eigenvectors, let P_j be the orthogonal projection in $\mathbb{C}^n$ onto the span of u_j. Then A can be written in the form

$$A = \sum_{j=1}^{n} \lambda_j P_j \tag{1.6}$$

This is the *spectral decomposition* of A.

For any function $f : \mathbb{R} \to \mathbb{R}$, the operator $f(A) \in \mathbf{H}_n$ is then defined by

$$f(A) = \sum_{j=1}^{m} f(\lambda_j) P_j \ . \tag{1.7}$$

Using the evident fact that $P_j P_k = \delta_{j,k} P_j$, it is easily checked that if f is a polynomial in the real variable t, say $f(t) = \sum_{j=0}^{k} a_j t^j$, then for this definition, $f(A) = \sum_{j=0}^{k} a_j A^j$. It is easily checked that in general, as in the polynomial case, $f(A)$ does not depend on the choice of the orthonormal basis $\{u_1, \dots, u_n\}$ consisting of eigenvectors of A.

A case that will be of particular focus in these notes is $f(t) = t \log(t)$. Given $\rho \in \mathbf{S}_n$, let $\{u_1, \dots, u_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of ρ: $\rho u_j = \lambda_j u_j$. Since $\rho \geq 0$, each $\lambda_j \geq 0$ for each j. Then by (1.3), $\sum_{j=1}^{n} \lambda_j = 1$, and so $\lambda_j \leq 1$ for each j. By (1.3) once more,

$$S(\rho) = -\sum_{j=1}^{n} \lambda_j \log \lambda_j \ . \tag{1.8}$$

That is, $S(\rho)$ depends on ρ only through its eigenvalues. Otherwise put, the von Neumann entropy is unitarily invariant; i.e.,

$$S(U \rho U^*) = S(\rho) \ . \tag{1.9}$$

The fact that $t \mapsto t \log(t)$ is strictly convex together with (1.8) tells us that

$$\begin{aligned} -S(\rho) = n \frac{1}{n} \sum_{j=1}^{n} \lambda_j \log \lambda_j &\leq n \left(\frac{1}{n} \sum_{j=1}^{n} \lambda_j \right) \log \left(\frac{1}{n} \sum_{j=1}^{n} \lambda_j \right) \\ &= n \left(\frac{1}{n} \right) \log \left(\frac{1}{n} \right) = -\log(n) \ , \end{aligned}$$

and there is equality if and only if each $\lambda_j = 1/n$. Thus, we have

$$0 \le S(\rho) \le \log n \tag{1.10}$$

for all $\rho \in \mathbf{S}_n$, and there is equality on the left iff ρ is a pure state, and there is equality on the right iff $\rho = (1/n)I$.

Actually, $S(\rho)$ is not only a strictly concave function of the eigenvalues of ρ, *it is strictly concave function of ρ itself.*

That is, as we shall show in the next section,

$$S((1-t)\rho_0 + t\rho_1) \ge (1-t)S(\rho_0) + tS(\rho_1) \tag{1.11}$$

for all $\rho_0, \rho_1 \in \mathbf{S}_n$, with equality iff $\rho_0 = \rho_1$. This is much stronger than concavity as a function of the eigenvalues since if ρ_0 and ρ_1 do not commute, the eigenvalues of $(1-t)\rho_0 + t\rho_1$ are not simply linear combinations of the eigenvalues of ρ_0 and ρ_1.

Since we shall be focusing on convexity and concavity of trace functions in these notes, we briefly discuss *one* reason this concavity matters, starting with the simpler fact (1.10) that we have deduced from the concavity of the entropy as a function of the eigenvalues of ρ.

In quantum statistical mechanics, *equilibrium states* are determined by maximum entropy principles, and the fact that

$$\sup_{\rho \in \mathbf{S}_n} S(\rho) = \log n \tag{1.12}$$

reflects Boltzmann's identity

$$S = k \log W$$

which is engraved on his funerary monument in Vienna.

Often however, we are not interested in the unconstrained supremum in (1.12), but instead the constrained supremum over states with a specified energy: Consider a quantum system in which the observables are self adjoint operators on $\mathbb{C}^n$, and in particular, in which the energy is represented by $H \in \mathbf{H}_n$. By the rules of quantum mechanics, the *expected value* of the energy in the state ρ is given by $\mathrm{Tr}(H\rho)$. Our constrained optimization problem is to compute

$$\sup\{\ S(\rho)\ :\ \rho \in \mathbf{S}_n\ , \mathrm{Tr}(H\rho) = E\ \}\ . \tag{1.13}$$

The key to this solving this problem is a duality formula for the entropy that is part of Theorem 2.13 which we prove in the next section. The formula in question says that for any density matrix ρ in $\mathbf{M}_n$,

$$-S(\rho) = \sup_{A \in \mathbf{H}_n} \{\mathrm{Tr}(A\rho) - \ln\left(\mathrm{Tr}e^A\right)\ \}\ , \tag{1.14}$$

and moreover

$$-S(\rho) = \mathrm{Tr}(A\rho) - \ln\left(\mathrm{Tr}e^A\right) \iff \rho = \frac{1}{\mathrm{Tr}(e^A)}e^A\ . \tag{1.15}$$

Notice that for any fixed $A \in \mathbf{H}_n$, the function

$$\rho \mapsto \mathrm{Tr}(A\rho) - \ln\left(\mathrm{Tr}e^A\right)$$

is affine, and therefore convex. It is easy to see that the pointwise supremum of any family of convex functions is again convex, and so proving the duality formula (1.14) would immediately prove the concavity of the entropy.

In fact, under mild regularity conditions, every convex function can be written in this way, as a supremum of affine functions. Such dual representations are very useful in solving variations problems involving convex functions. We illustrate this by solving the optimization problem in (1.13).

1.2. Definition (Gibbs state). Given $H \in \mathbf{H}_n$ and $\beta \in \mathbb{R}$, the *Gibbs state for Hamiltonian H at inverse temperature β* is the density matrix $\rho_{\beta,H}$ where

$$\rho_{\beta,H} := \frac{1}{\mathrm{Tr}\left[e^{-\beta H}\right]} e^{-\beta H} \ . \tag{1.16}$$

Define $\rho_{\infty,H} = \lim_{\beta\to\infty} \rho_{\beta,H}$ and $\rho_{-\infty,H} = \lim_{\beta\to-\infty} \rho_{\beta,H}$.

It may seem strange that we allow negative "temperatures" in this definition. However, it is physically natural for systems with finitely many degrees of freedom. On a mathematical level, negative "temperatures" are natural on account of the following theorem:

1.3. Theorem (Variational principle for Gibbs states). *Let $H \in \mathbf{H}_n$, and let $\lambda_1 \leq \cdots \leq \lambda_n$ be the n eigenvalues of H. Then for any E with*

$$\lambda_1 \leq E \leq \lambda_n$$

there is a unique $\beta \in [-\infty, \infty]$ such that

$$E = \mathrm{Tr}(H\rho_{\beta_E,H}) \ ,$$

and for any $\tilde{\rho} \in \mathbf{S}_n$,

$$\mathrm{Tr}(H\tilde{\rho}) = E \quad \text{and} \quad S(\tilde{\rho}) = \sup\{\ S(\rho)\ :\ \rho \in \mathbf{S}_n\ , \mathrm{Tr}(H\rho) = E\ \} \iff \tilde{\rho} = \rho_{\beta_E,H} \ .$$

Proof: We shall use (1.14) and (1.15), though these shall only be proved in Theorem 2.13. Note that by (1.14), for any $\tilde{\rho} \in \mathbf{S}_n$ with $\mathrm{Tr}(\tilde{\rho}H) = E$,

$$S(\tilde{\rho}) \leq \beta E + \ln\left(\mathrm{Tr} e^{-\beta H}\right) \ , \tag{1.17}$$

and by (1.15), there is equality in (1.17) if and only if

$$\tilde{\rho} = \frac{1}{\mathrm{Tr}\left[e^{-\beta_E H}\right]} e^{-\beta_E H} = \rho_{\beta_E,H} \ , \tag{1.18}$$

where β_E is such that $\mathrm{Tr}(\rho_{\beta_E,H}H) = E$, so that our constraint is satisfied. For $\lambda_1 \leq E \leq \lambda_n$, there is exactly one such β_E as we explain next.

Notice that

$$\mathrm{Tr}(H\rho_{\beta,H}) = \frac{1}{\mathrm{Tr}\left[e^{-\beta H}\right]}\mathrm{Tr}[He^{-\beta H}] = -\frac{\mathrm{d}}{\mathrm{d}\beta}\log\left(\mathrm{Tr}[e^{-\beta H}]\right) \ .$$

A direct calculation shows that

$$\frac{\mathrm{d}^2}{\mathrm{d}\beta^2}\log\left(\mathrm{Tr}[e^{-\beta H}]\right) = \frac{1}{\mathrm{Tr}\left[e^{-\beta H}\right]}\mathrm{Tr}[H^2 e^{-\beta H}] - \left(\frac{1}{\mathrm{Tr}\left[e^{-\beta H}\right]}\mathrm{Tr}[He^{-\beta H}]\right)^2 \ .$$

By the Schwarz inequality, this is strictly positive unless H is a multiple of the identity. If H is a multiple of the identity, necessarily EI, the optimization problem is trivial, and the optimal ρ is $(1/n)I$. So assume that H is not a multiple of the identity. Then the map $\beta \mapsto \log\left(\mathrm{Tr}[e^{-\beta H}]\right)$ is strictly convex, and hence $\beta \mapsto \mathrm{Tr}(H\rho_{\beta,H})$ is strictly monotone decreasing, and in fact,

$$\mathrm{Tr}(H\rho_{\beta,H}) = \frac{1}{\sum_{j=1}^n e^{-\beta\lambda_j}} \sum_{j=1}^n \lambda_j e^{-\beta\lambda_j} \ .$$

It follows that

$$\lim_{\beta\to\infty} \mathrm{Tr}(H\rho_{\beta,H}) = \lambda_1 \qquad \text{and} \lim_{\beta\to-\infty} \mathrm{Tr}(H\rho_{\beta,H}) = \lambda_n \ .$$

By the Spectral Theorem, $\lambda_1 \le \mathrm{Tr}(H\rho) \le \lambda_n$ for all $\rho \in \mathbf{S}_n$, and so for all $E \in [\lambda_1, \lambda_n]$, there is a unique value of $\beta \in [-\infty, \infty]$ such that $\mathrm{Tr}(H\rho_{\beta,H}) = E$. ■

The fact that the Gibbs state $\rho_{\beta_E,H}$ is characterized by the fact that

$$S(\rho_{\beta_E,H}) \ge S(\tilde\rho) \quad \text{for all} \quad \tilde\rho \in \mathbf{S}_n \quad \text{with} \quad \mathrm{Tr}(H\tilde\rho) = E$$

gives rise to the *variational principle* for Gibbs states. If one accepts the idea that *equilibrium states* should maximize the entropy given the expected values of macroscopic observables, such as energy, this leads to an identification of Gibbs states with thermodynamic equilibrium states.

There is a great deal of physics, and many open questions, that one should discuss to set this variational principle in proper physical context. Such a discussion is far beyond the scope of these notes. Here we shall simply accept the deep physical insight that *equilibrium states* should maximize the entropy given the expected values of macroscopic observables, such as energy. Then, as we have seen, this extremal property is intimately connected with the concavity of the entropy.

Further exploration of these ideas would lead one to investigate certain convexity, concavity and monotonicity properties of other trace functions. These notes will focus on the *mathematical aspects* of convexity, concavity and monotonicity properties of trace functions. This is an extremely beautiful theory, and the beauty can be deeply appreciated on a purely mathematical basis. However, one should bear in mind that many of the most profound aspects of this theory were discovered through physical enquiry.

As we have indicated, the entropy is not the only trace function that matters in statistical mechanics: Even in this very particular context of the entropy maximization problem, the duality formula (1.14) involves the trace functional $A \mapsto \log\left(\mathrm{Tr}(e^A)\right)$, and its proof makes use of what is known as the *relative entropy*:

1.4. Definition. The *relative entropy* of $\rho \in \mathbf{S}_n$ with respect to $\sigma \in \mathbf{S}_n$, $S(\rho|\sigma)$, is defined to be $+\infty$ unless the nullspace of ρ contains the nullspace of σ in which case it is defined by

$$S(\rho|\sigma) = \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\rho\log\sigma) \ . \tag{1.19}$$

As we shall see, $(\rho,\sigma) \mapsto S(\rho|\sigma)$ is *jointly convex* in the sense that for all $\rho_0, \rho_1, \sigma_0, \sigma_1 \in \mathbf{S}_n$ and any $0 \le t \le 1$,

$$S((1-t)\rho_0 + t\rho_1|(1-t)\sigma_0 + t\sigma_1) \le (1-t)S(\rho_0|\sigma_0) + tS(\rho_1|\sigma_1) \ . \tag{1.20}$$

This is a deeper fact than is needed to prove the duality formula (1.14) – which only uses the much more easily proved fact that $S(\rho|\sigma) \geq 0$ with equality if and only if $\sigma = \rho$. This is a special case of *Klein's inequality*, which we prove in Section 2, and show it to be a direct consequences of the strict concavity of the von Neuman entropy.

The joint convexity of $S(\rho|\sigma)$ is, as we have noted, much deeper than the strict concavity of the von Neuman entropy. Its proof, and its applications, shall have to wait until later. The ideas leading to the proof are closely connected with yet another type of "entropy"; i.e., the *Wigner-Yanase skew information*.

The notions of "entropy" and "information" are distinct, but closely intertwined. The *Shannon information content* $I_S(\rho)$ of a density matrix $\rho \in \mathbf{S}_n$ is defined by $I_S(\rho) = -S(\rho)$. See [**19**] for a discussion of this definition. Note that with this definition, the information content of any pure state ρ is zero.

However, for a quantum mechanical system in which the observables are self adjoint operators on $\mathbb{C}^n$, and in which the energy is the observable $H \in \mathbf{H}_n$, some states are easier to measure than others: Those that commute with H are easy to measure, and those that do not are more difficult to measure. This led Wigner and Yanase [**38, 39**] to introduce the *Wigner-Yanase skew information* $I_{WY}(\rho)$ of a density matrix ρ in a quantum mechanical system with energy operator H to be

$$I_{WY}(\rho) = -\frac{1}{2}\mathrm{Tr}\left([\sqrt{\rho}, H]^2\right) \ .$$

Note that

$$I_{WY}(\rho) = \mathrm{Tr}H^2\rho - \mathrm{Tr}\sqrt{\rho}H\sqrt{\rho}H \ , \tag{1.21}$$

which vanishes if and only if ρ commutes with H.

Wigner and Yanase [**38, 39**] proved that $\rho \mapsto I_{WY}(\rho)$ is convex on $\mathbf{S}_n$, i.e.,

$$I_{WY}((1-t)\rho_0 + t\rho_1) \leq (1-t)I_{WY}(\rho_0) + tI_{WY}(\rho_1) \ , \tag{1.22}$$

and explained the information theoretic consequences of this. From the formula (1.21), it is clear that convexity of $\rho \mapsto I_{WY}(\rho)$ amounts to concavity of $\rho \mapsto \mathrm{Tr}\sqrt{\rho}H\sqrt{\rho}H$, which they proved. Even though this is a convexity result for one variable, and not a joint convexity result, it too is much harder to prove than the concavity of the von Neuman entropy, or what is the same thing, the convexity of the Shannon information.

Wigner and Yanase left open the more general problem of concavity of

$$\rho \mapsto \mathrm{Tr}(\rho^p K \rho^{1-p} K^*) \tag{1.23}$$

for $0 < p < 1$, $p \neq 1/2$. (Dyson had raised the issue of proving this more general case.) Lieb realized that this problem was closely connected with something then known as the *strong subadditivity of quantum entropy conjecture*, which was due to Ruelle and Robinson. Lieb [**18**] proved the convexity of the function in (1.23) for all $0 < p < 1$, and this deep result is known as the Lieb Concavity Theorem. Then he and Ruskai applied it to prove [**20**] the strong subadditivity of quantum entropy conjecture. Later in these notes we shall explain what this strong subadditivity is, why it is significant, and give several proofs of it.

For now, we mention that strong subadditivity of the entropy is an example of a trace inequality for density matrices acting on a tensor product of Hilbert spaces – $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ in this case – that involves *partial traces*. The different Hilbert spaces correspond to different parts of the quantum mechanical system: Here, $\mathcal{H}_1$, $\mathcal{H}_1$ and $\mathcal{H}_3$ are the state spaces for degrees of freedom in various subsystems of the whole system, and it is often important to estimate the entropy of a density matrix ρ on the whole system in terms of the entropies of induced density matrices on the various subsystems. Later in these notes, we shall extensively develop this general topic, and inequalities for tensor products are absolutely fundamental throughout the subject. In fact, the easiest (by far) proof of the Lieb Concavity Theorem proceeds through a simple tensor product argument devised by Ando [**1**].

Before entering fully into our subject, let us close the introduction by emphasizing that in our exposition, we shall provide full mathematical detail and context, but we shall be comparatively sketchy when it comes to physical detail and context. There are many excellent accounts of the physics behind the definitions of $S(\rho)$, $S(\rho|\sigma)$ and $I_{WY}(\rho)$ and other mathematical constructions that we shall encounter here. Thirring's book [**33**] is an especially good reference for much of this. It is especially good at connecting the physics with the rigorous mathematics, but still, what we do here provides a mathematical complement to it. For example, [**33**] does not contain a full proof of the joint convexity of $S(\rho|\sigma)$. It only give the simple argument which reduces this to the Lieb Concavity Theorem, about which it says: "The proof of this rather deep proposition ... is too laborious to be repeated here". In fact, as we shall see, one can give a very simple, clean and clear proof of this. The point of view that leads to this simple proof is due to Ando [**1**], and as we shall see, it provides insight into a number of other interesting questions as well. We now turn to the systematic development of our subject – inequalities for operators and traces, with special focus on monotonicity and convexity.

2. Operator convexity and monotonicity

2.1. Some examples and the Löwner-Heinz Theorem.

2.1. Definition (Operator monotonicity and operator convexity). A function $f : (0,\infty) \to \mathbb{R}$ is said to be *operator monotone* in case whenever for all n, and all $A, B \in \mathbf{H}_n^+$,

$$A \geq B \;\Rightarrow\; f(A) \geq f(B)\ . \tag{2.1}$$

A function $f : (0,\infty) \to \mathbb{R}$ is said to be *operator convex* in case for all n and all $A, B \in \mathbf{H}_n^+$, and $0 < t < 1$,

$$f((1-t)A + tB) \leq tf(A) + (1-t)f(B)\ . \tag{2.2}$$

We say that f is *operator concave* if $-f$ is operator convex.

By considering the case of 1×1 matrices, or diagonal matrices in $\mathbf{H}_n$, one sees that if f is monotone or convex in the operator sense, then it must be monotone or convex in the usual sense as a function from $(0,\infty)$ to $\mathbb{R}$. The opposite is not true.

2.2. Example (Non-monotonicity of the square). The function $f(t) = t^2$ is monotone in the usual sense, but for $A, B \in \mathbf{H}_n^+$,

$$(A+B)^2 = A^2 + (AB + BA) + B^2 \ .$$

For any choice of A and B such that $AB + BA$ has even one strictly negative eigenvalue, $(A + tB)^2 \geq A^2$ will fail for all sufficiently small t. It is easy to find such A and B in $\mathbf{H}_n^+$. For example, take

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \ , \tag{2.3}$$

so that

$$AB + BA = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \ .$$

Thus, not even the square function is operator monotone. □

It turns out, however, that the square root function *is* operator monotone. This is an important theorem of Heinz [**13**]. The proof we give is due to Kato [**17**]; see [**9**] for its specialization to the matrix case, which we present here.

2.3. Example (Monotoncity of the square root). The square root function, $f(t) = t^{1/2}$, is operator monotone. To see this, it suffices to show that if $A, B \in \mathbf{H}_n^+$ and $A^2 \leq B^2$, then $A \leq B$.

Towards this end, consider any eigenvalue λ of the Hermitian matrix $B - A$, and let u be a unit vector that is an eigenvector with this eigenvalue. We must show that $\lambda \geq 0$. Observe that

$$(B - \lambda)u = Au \quad \Rightarrow \langle Bu, (B-\lambda)u\rangle = \langle Bu, Au\rangle \ .$$

Then by the Schwarz inequality,

$$\|Bu\|^2 - \lambda\langle u, Bu\rangle \leq \|Bu\|\|Au\| \ .$$

But since $A^2 \leq B^2$,

$$\|Bu\|\|Au\| = \langle u, B^2u\rangle^{1/2}\langle u, A^2u\rangle^{1/2} \leq \langle u, B^2u\rangle \ ,$$

we have

$$\|Bu\|^2 - \lambda\langle u, Bu\rangle \leq \|Bu\|^2 \ ,$$

and this shows that $\lambda\langle u, Bu\rangle \geq 0$, and hence $\lambda \geq 0$. □

We now give two examples pertaining to convexity:

2.4. Example (Convexity of the square). The square function is operator convex: One has the parallelogram law

$$\left(\frac{A+B}{2}\right)^2 + \left(\frac{A-B}{2}\right)^2 = \frac{1}{2}A^2 + \frac{1}{2}B^2 \ ,$$

so certainly for $f(t) = t^2$, one always has (2.2) for $t = 1/2$, which is known as *midpoint convexity*. A standard argument then gives (2.2) whenever t is a dyadic rational, and then by continuity one has it for all t, in $(0, 1)$ of course. We will often use the fact that in the presence of continuity, it suffices to check midpoint convexity. □

2.5. Example (Non-convexity of the cube). The cube function is not operator convex. To easily see this, let us deduce a consequence of (2.2) that must hold for any operator convex function f: Take $A, B \in \mathbf{H}_n^+$, and all $0 < t < 1$, and note that

$$A + tB = (1-t)A + t(A+B) \ .$$

Thus, from (2.2), $f(A+tB) \le (1-t)f(A) + tf(A+B)$ which yields

$$\frac{f(A+tB) - f(A)}{t} \le f(A+B) - f(A) \ . \tag{2.4}$$

Taking f to be the cube function, and then letting t tend to zero in (2.4), we see that convexity of the cube function would imply that for all $A, B \in \mathbf{H}_n^+$,

$$(B^3 + BAB) + (AB^2 + B^2A) \ge 0 \ .$$

This fails for A, B chosen exactly as in (2.3); indeed, note that for this choice $B^3 = B^2 = BAB = B$, so that

$$(B^3 + BAB) + (AB^2 + B^2A) = \begin{bmatrix} 4 & 1 \\ 1 & 0 \end{bmatrix} \ ,$$

which is definitely not positive semi-definite! □

After seeing these negative examples, one might suspect that the notions of operator monotonicity and operator convexity are too restrictive to be of any interest. Fortunately, this is not the case. The following result furnishes a great many positive examples.

2.6. Theorem (Löwner-Heinz Theorem). *For $-1 \le p \le 0$, the function $f(t) = -t^p$ is operator monotone and operator concave. For $0 \le p \le 1$, the function $f(t) = t^p$ is operator monotone and operator concave. For $1 \le p \le 2$, the function $f(t) = t^p$ and operator convex. Furthermore $f(t) = \log(t)$ is operator concave and operator monotone, while $f(t) = t\log(t)$ is operator convex.*

Löwner actually proved more; he gave a necessary and sufficient condition for f to be operator monotone. But what we have stated is all that we shall use.

We shall give an elementary proof of Theorem 2.6 after first proving two lemmas. The first lemma addresses the special case $f(t) = t^{-1}$.

2.7. Lemma. *The function $f(t) = t^{-1}$ is operator convex, and the function $f(t) = -t^{-1}$ is operator monotone.*

Proof: We begin with the monotonicity. Let $A, B \in \mathbf{H}_n^+$. Let $C = A^{-1/2}BA^{-1/2}$ so that

$$A^{-1} - (A+B)^{-1} = A^{-1/2}[I - (I+C)^{-1}]A^{-1/2} \ .$$

Since $C \in \mathbf{H}_n^+$, $I - (I+C)^{-1} \in \mathbf{H}_n^+$, and hence $A^{-1/2}[I - (I+C)^{-1}]A^{-1/2} > 0$. This proves the monotonicity.

Similarly, to prove midpoint convexity, we have

$$\frac{1}{2}A^{-1} + \frac{1}{2}B^{-1} - \left(\frac{A+B}{2}\right)^{-1} = A^{-1/2}\left[\frac{1}{2}I + \frac{1}{2}C^{-1} - \left(\frac{I+C}{2}\right)^{-1}\right]A^{-1/2} \ .$$

By the *arithmetic-harmonic mean inequality*, for any real numbers $a, b > 0$,

$$\frac{a+b}{2} \geq \left(\frac{a^{-1}+b^{-1}}{2}\right)^{-1} .$$

Applying this with $a = 1$ and c any eigenvalue of C^{-1}, we see from the Spectral Theorem that

$$\left[\frac{1}{2}I + \frac{1}{2}C^{-1} - \left(\frac{I+C}{2}\right)^{-1}\right] \geq 0 ,$$

from which

$$\frac{1}{2}A^{-1} + \frac{1}{2}B^{-1} - \left(\frac{A+B}{2}\right)^{-1} \geq 0$$

follows directly. Again, by continuity, the full convexity we seek follows from the midpoint convexity that we have now proved. ∎

The other ingredient to the proof of Theorem 2.6 is a set of integral representations for the functions $A \mapsto A^p$ in $\mathbf{H}_n^+$ for p in the ranges $-1 < p < 0$, $0 < p < 1$ and $1 < p < 2$:

2.8. Lemma. *For all $A \in \mathbf{H}_n^+$, one has the following integral formulas:*

$$(2.5) \quad A^p = \frac{\pi}{\sin(\pi(p+1))}\int_0^\infty t^p \frac{1}{t+A}\mathrm{d}t \qquad \text{for all } -1 < p < 0 .$$

$$(2.6) \quad A^p = \frac{\pi}{\sin(\pi p)}\int_0^\infty t^p \left(\frac{1}{t} - \frac{1}{t+A}\right)\mathrm{d}t \qquad \text{for all } 0 < p < 1 .$$

$$(2.7) \quad A^p = \frac{\pi}{\sin(\pi(p-1))}\int_0^\infty t^p \left(\frac{A}{t} + \frac{t}{t+A} - I\right)\mathrm{d}t \qquad \text{for all } 1 < p < 2 .$$

Proof: For all $a > 0$, and all $0 < p < 1$ the integral $\displaystyle\int_0^\infty t^p \left(\frac{1}{t} - \frac{1}{t+a}\right)\mathrm{d}t$ converges since the singularity at the origin is $\mathcal{O}(t^{p-1})$ and the decay at infinity is $\mathcal{O}(t^{p-2})$. Making the change of variables $t = as$, it is then easy to see that the integral is a constant multiple of a^p, where the constant depends only on p. This is all we need, but in fact a simple contour integral calculation yields the explicit result

$$a^p = \frac{\pi}{\sin(\pi p)}\int_0^\infty t^p \left(\frac{1}{t} - \frac{1}{t+a}\right)\mathrm{d}t .$$

Multiplying by a, we get

$$a^{p+1} = \frac{\pi}{\sin(\pi p)}\int_0^\infty t^p \left(\frac{a}{t} + \frac{t}{t+a} - I\right)\mathrm{d}t .$$

Dividing by a, we get

$$a^{p-1} = \frac{\pi}{\sin(\pi p)}\int_0^\infty t^{p-1} \frac{1}{t+a}\mathrm{d}t .$$

The formulas (2.5),(2.6) and (2.7) now follow by the Spectral Theorem. ∎

Proof of Theorem 2.6: Note that Lemma 2.7 yields the concavity and monotonicity of $A \mapsto A^p$ for $p = -1$. The case $p = 0$ is trivial, and we have already directly established the convexity (and the non-monotonicity) for $p = 2$. For non-integer values of p, we use Lemma 2.8 to reduce to the case $p = -1$.

By Lemma 2.7, the map $A \mapsto -(t + A)^{-1}$ is operator concave and operator monotone. Any weighted sum (with positive weights) of operator concave and operator monotone functions is again operator concave and operator monotone. For $-1 < p < 0$, (2.5) says that $-A^p$ is such a weighted sum, and so $A \mapsto A^p$ is operator convex, and $A \mapsto -A^p$ is operator monotone in this range. A very similar argument shows that for $0 < p < 1$, $A \mapsto A^p$ is operator concave and operator monotone.

The case $1 < p < 2$ is a bit different: By Lemma 2.7, the map

$$A \mapsto \frac{A}{t} + \frac{t}{t + A}$$

is a sum of operator convex functions, and hence is operator convex. However, it is a difference of operator monotone functions, and is not operator monotone. Hence all that we can conclude is that $A \mapsto A^p$ is convex in this range. (Indeed, we have seen that the monotonicity fails at $p = 2$, and so monotonicity, which is preserved under limits, cannot hold for p near 2.)

Finally, again by the Spectral Theorem,

$$\log(A) = \lim_{p\to 0} \frac{1}{p}(A^p - I) \qquad \text{and} \qquad A\log(A) = \lim_{p\to 1} \frac{A^p - A}{p - 1} \tag{2.8}$$

Since the map $A \mapsto \frac{1}{p}(A^p - I)$ has been shown to be operator monotone and operator concave for all $p \in [-1, 1]$, and since these properties are preserved under limits taken in (2.8), we see that $A \mapsto \log(A)$ is operator monotone and operator concave. Likewise, since $A \mapsto (p - 1)^{-1}(A^p - A)$ is convex for all $p \neq 1$ in the interval $[0, 2]$, we see that $A \mapsto A\log(A)$ is operator convex. ■

We close this subsection by stating Löwner's necessary and sufficient condition for $f : (0, \infty) \to \mathbb{R}$ to be operator monotone: This is the case if and only if f admits an integral representation

$$f(a) = \alpha + \beta a - \int_0^\infty \frac{1 - at}{t + a} \mathrm{d}\mu(t)$$

for some $\alpha, \beta \in \mathbb{R}$, $\beta > 0$, and some finite positive measure μ.

2.2. Convexity and monotonicity for trace functions. Given a function $f : \mathbb{R} \to \mathbb{R}$, consider the associated *trace function* on $\mathbf{H}_n$ given by

$$A \mapsto \mathrm{Tr}[f(A)] \ .$$

In this subsection we address the question: Under what conditions on f is such a trace function monotone, and under what conditions on f is it convex? We shall see that much less is required of f in this context than is required for operator monotonicity or operator convexity. Notice first of all that we are working now on $\mathbf{H}_n$ and not only $\mathbf{H}_n^+$, and with functions defined on all of $\mathbb{R}$ and not only on $(0, \infty)$.

The question concerning monotonicity is very simple. Suppose that f is continuously differentiable. Let $B, C \in \mathbf{H}_n^+$. Then by the Spectral Theorem and first order perturbation theory,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Tr}(f(B+tC))\big|_{t=0} = \mathrm{Tr}(f'(B)C) = \mathrm{Tr}(C^{1/2}f'(B)C^{1/2}) \ ,$$

where in the last step we have used cyclicity of the trace. As long as f has a positive derivative, all of the eigenvalues of $f'(B)$ will be positive, and so $f'(B)$ is positive semi-definite, and therefore so is $C^{1/2}f'(B)C^{1/2}$. It follows immediately that $\mathrm{Tr}(C^{1/2}f'(B)C^{1/2}) \geq 0$, and from here one easily sees that for $A \geq B$, and with $C = A - B$,

$$\mathrm{Tr}[f(A)] - \mathrm{Tr}[f(B)] = \int_0^1 \mathrm{Tr}(C^{1/2}f'(A+tB)C^{1/2})\mathrm{d}t \geq 0 \ .$$

Thus, $\mathrm{Tr}[f(A)] \geq \mathrm{Tr}[f(B)]$ whenever $A > B$ and f is continuously differentiable and monotone increasing. By a simple continuity argument, one may relax the requirement that f be continuously differentiable to the requirement that f be continuous.

The question concerning convexity is more interesting. Here we have the following theorem:

2.9. Theorem (Peierls Inequality). *Let $A \in \mathbf{H}_n$, and let f be any convex function on $\mathbb{R}$. Let $\{u_1, \dots, u_n\}$ be any orthonormal base of $\mathbb{C}^n$. Then*

$$\sum_{j=1}^n f(\langle u_j, Au_j\rangle) \leq \mathrm{Tr}[f(A)] \ . \tag{2.9}$$

There is equality if each u_j is an eigenvector of A, and if f is strictly convex, only in this case.

Proof: By (1.3) together with the spectral representation (1.7),

$$\begin{aligned} \mathrm{Tr}[f(A)] &= \sum_{j=1}^n \left\langle u_j \left[\sum_{k=1}^m f(\lambda_k)P_k\right] u_j \right\rangle \\ &= \sum_{j=1}^n \left(\sum_{k=1}^m f(\lambda_k)\|P_k u_j\|^2\right) \\ &\geq \sum_{j=1}^n f\left(\sum_{k=1}^m \lambda_k\|P_k u_j\|^2\right) \qquad (2.10) \\ &= \sum_{j=1}^n f(\langle u_j, Au_j\rangle) \ . \end{aligned}$$

The inequality above is simply the convexity of f, since for each j, $\sum_{k=1}^m \|P_k u_j\|^2 = \|u_j\|^2 = 1$, and thus $\sum_{k=1}^m \lambda_k\|P_k u_j\|^2$ is a weighted average of the eigenvalues of A. Note that each u_j is an eigenvector of A if and only if $\|P_k u_j\|^2 = 1$ for some k, and is 0 otherwise, in which case the inequality in (2.10) is a trivial equality. And clearly when f is strictly convex, equality can hold in (2.10) only if for each j, $\|P_k u_j\|^2 = 1$ for some k, and is 0 otherwise. ■

Now consider $A, B \in \mathbf{H}_n$, and let $f : \mathbb{R} \to \mathbb{R}$ be convex. Let $\{u_1, \dots, u_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $(A+B)/2$. Then, Theorem 2.9,

$$\begin{aligned}
\mathrm{Tr}\left[f\left(\frac{A+B}{2}\right)\right] &= \sum_{j=1}^{n} f\left(\left\langle u_j, \frac{A+B}{2} u_j\right\rangle\right) \\
&= \sum_{j=1}^{n} f\left(\frac{1}{2}\langle u_j, Au_j\rangle + \frac{1}{2}\langle u_j, Bu_j\rangle\right) \\
&\le \sum_{j=1}^{n} \left[\frac{1}{2} f(\langle u_j, Au_j\rangle) + \frac{1}{2} f(\langle u_j, Bu_j\rangle)\right] \qquad (2.11) \\
&\le \frac{1}{2}\mathrm{Tr}[f(A)] + \frac{1}{2}\mathrm{Tr}[f(B)] \ . \qquad (2.12)
\end{aligned}$$

where we have used Theorem 2.9, in the first equality, and where in (2.11) we have used the (midpoint) convexity of f, and in (2.12) we have used Theorem 2.9 again.

This shows that for every natural number n, whenever f is midpoint convex, the map $A \mapsto \mathrm{Tr}[f(A)]$ is midpoint convex on $\mathbf{H}_n$. Note that if f is strictly convex and $\mathrm{Tr}[f(A+B)/2] = \mathrm{Tr}[f(A)]/2 + \mathrm{Tr}[f(A)]/2$, we must have equality in both (2.11) and (2.12). On account of the strict convexity of f, equality in (2.11) implies that $\langle u_j, Au_j\rangle = \langle u_j, Bu_j\rangle$ for each u_j. By the conditions for equality in Peierl's inequality, equality in (2.11) implies that each u_j is an eigenvector of both A and B. Thus,

$$Au_j = \langle u_j, Au_j\rangle u_j = \langle u_j, Bu_j\rangle u_j = Bu_j \ ,$$

and so $A = B$.

A simple continuity argument now shows that if f continuous as well as convex, $A \mapsto \mathrm{Tr}[f(A)]$ is convex on $\mathbf{H}_n$, and strictly so if f is strictly convex.

Let us summarize some conclusions we have drawn so far in a theorem:

2.10. Theorem (Convexity and monotonicty of trace functions). *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous, and let n be any natural number. Then if $t \mapsto f(t)$ is monotone increasing, so is $A \mapsto \mathrm{Tr}[f(A)]$ on $\mathbf{H}_n$. Likewise, if $t \mapsto f(t)$ is convex, so is $A \mapsto \mathrm{Tr}[f(A)]$ on $\mathbf{H}_n$, and strictly so if f is strictly convex.*

2.3. Klein's Inequality and the Peierls-Bogoliubov Inequality. We close this section with three trace theorems that have significant applications in statistical quantum mechanics.

2.11. Theorem (Klein's Inequality). *For all $A, B \in \mathbf{H}_n$, and all differentiable convex functions $f : \mathbb{R} \to \mathbb{R}$, or for all $A, B \in \mathbf{H}_n^+$, and all differentiable convex functions $f : (0, \infty) \to \mathbb{R}$*

$$\mathrm{Tr}[f(A) - f(B) - (A - B)f'(B)] \ge 0 \ . \qquad (2.13)$$

In either case, if f is strictly convex, there is equality if and only if $A = B$.

Proof: Let $C = A - B$ so that for $0 < t < 1$, $B + tC = (1-t)B + tA$. Define $\varphi(t) = \mathrm{Tr}[f(B + tC)]$. By Theorem 2.10, φ is convex, and so for all $0 < t < 1$,

$$\varphi(1) = \varphi(0) \ge \frac{\varphi(t) - \varphi(0)}{t} \ ,$$

and in fact the right hand side is monotone decreasing in t. Taking the limit $t \to 0$ yields (2.13). Note that if f is strictly convex and $C \neq 0$, then φ is strictly convex. The final assertion follows from this and the fact that $\dfrac{\varphi(t)-\varphi(0)}{t}$ is monotone decreasing in t. ■

2.12. Theorem (Peierls-Bogoliubov Inequality). *For every natural number n, the map*

$$A \mapsto \log\left(\mathrm{Tr}[\exp(A)]\right)$$

is convex on $\mathbf{H}_n$

Remark The appellation "Peierls-Bogoliubov Inequality" has been attached to many inequalities by many authors. It is often used to refer to the inequality one gets as a consequence of Theorem 2.12 and the "increasing chordal slope argument" used to prove Klein's inequality.

Indeed, for any $A, B \in \mathbf{H}_n$, and any $0 < t < 1$, let $\psi(t)$ be the function

$$t \mapsto \log\left(\mathrm{Tr}[\exp(A + tB)]\right) \ .$$

By Theorem 2.12, this is convex, and hence

$$\psi(1) - \psi(0) \geq \frac{\psi(t) - \psi(0)}{t}$$

for all t. Taking the limit $t \to 0$, which exists by monotonicity, we obtain

$$\log\left(\frac{\mathrm{Tr}[e^{A+B}]}{\mathrm{Tr}[e^A]}\right) \geq \frac{\mathrm{Tr}[Be^A]}{\mathrm{Tr}[e^A]} \ . \tag{2.14}$$

Frequently this consequence of Theorem 2.12, which has many uses, is referred to as the Peierls-Bogoliubov Inequality.

Proof of Theorem 2.12: We first define $\varphi : \mathbb{R}^n \to \mathbb{R}$ by $\varphi(x) = \log\left(\sum_{k=1}^n e^{x_k}\right)$. A simple computation of the Hessian matrix of φ yields

$$\frac{\partial^2}{\partial x_i \partial x_j}\varphi(x) = a_j\delta_{i,j} - a_i a_j \qquad \text{where} \qquad a_i = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}} \ .$$

Hence for any $y \in \mathbb{R}^n$,

$$\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}\varphi(x) y_i y_j = \sum_{j=1}^n a_j y_j^2 - \left(\sum_{j=1}^n a_j y_j\right)^2 \ .$$

Then by Schwarz inequality, and the fact that $\sum_{j=1}^n a_j = 1$, $\left|\sum_{j=1}^n a_j y_j\right| \leq \sum_{j=1}^n a_j y_j^2$. Thus, the Hessian matrix of φ is non-negative at each $x \in \mathbb{R}^n$, and hence φ is convex. Hence, for any $x, y \in \mathbb{R}^n$,

$$\varphi\left(\frac{x+y}{2}\right) \leq \frac{1}{2}\varphi(x) + \frac{1}{2}\varphi(y) \ . \tag{2.15}$$

To apply this, let $A, B \in \mathbf{H}_n$, and let $\{u_1, \ldots, u_n\}$ be any orthonormal basis of $\mathbb{C}^n$. For each $j = 1, \ldots, n$, let $x_j = \langle u_j, Au_j\rangle$ and $y_j = \langle u_j, Bu_j\rangle$, and let x and

y be the corresponding vectors in $\mathbb{R}^n$. Then if we take $\{u_1, \ldots, u_n\}$ to consist of eigenvectors of $(A+B)/2$, we have from Theorem 2.9 that
(2.16)

$$\log\left(\mathrm{Tr}\left[\exp\left(\frac{A+B}{2}\right)\right]\right) = \log\left(\sum_{j=1}^{n} \exp\left\langle u_j, \frac{A+B}{2}u_j\right\rangle\right) = \varphi\left(\frac{x+y}{2}\right) .$$

Now, again by Theorem 2.9, $\mathrm{Tr}[\exp(A)] \geq \sum_{j=1}^{n} \exp(\langle u_j, Au_j\rangle)$, and so by the monotonicity of the logarithm, and the definition of x and $\varphi(x)$, $\log(\mathrm{Tr}[\exp(A)]) \geq \varphi(x)$. A similar argument yields $\log(\mathrm{Tr}[\exp(B)]) \geq \varphi(y)$. Combining these last two inequalities with (2.15) and (2.16) proves the theorem. ∎

Not only are the functions $H \mapsto \log\left(\mathrm{Tr}[e^H]\right)$ and $\rho \mapsto -S(\rho)$ both convex, they are *Legendre Transforms* of one another. (See [**26**] for a full mathematical treatment of the Legendre transform.) Before proving this result, which is justifies (1.14) from the introduction, we first extend the domain of S to all of $\mathbf{H}_n$:

$$S(A) := \begin{cases} -\mathrm{Tr}(A \ln A) & A \in \mathbf{S}_n, \\ -\infty & \text{otherwise.} \end{cases} \tag{2.17}$$

2.13. THEOREM (Duality formula for the entropy). *For all $A \in \mathbf{H}_n$,*

$$-S(A) = \sup\left\{\mathrm{Tr}(AH) - \ln\left(\mathrm{Tr}\left[e^H\right]\right) \ : \ H \in \mathbf{H}_n \ \right\} . \tag{2.18}$$

The supremum is an attained maximum if and only if A is a strictly positive probability density, in which case it is attained at H if and only if $H = \ln A + cI$ for some $c \in \mathbb{R}$. Consequently, for all $H \in \mathbf{H}_n$,

$$\ln\left(\mathrm{Tr}\left[e^H\right]\right) = \sup\left\{\mathrm{Tr}(AH) + S(A) \ : \ A \in \mathbf{H}_n\right\} . \tag{2.19}$$

The supremum is a maximum at all points of the domain of $\ln\left(\mathrm{Tr}\left(e^H\right)\right)$, in which case it is attained only at the single point $A = e^H/(\mathrm{Tr}(e^H))$.

Proof: To see that the supremum is ∞ unless $0 \leq A \leq I$, let c be any real number, and let u be any unit vector. Then let H be c times the orthogonal projection onto u. For this choice of H,

$$\mathrm{Tr}(AH) - \ln\left(\mathrm{Tr}\left(e^H\right)\right) = c\langle u, Au\rangle - \ln(e^c + (n-1)) .$$

If $\langle u, Au\rangle < 0$, this tends to ∞ as c tends to $-\infty$. If $\langle u, Au\rangle > 1$, this tends to ∞ as c tends to ∞. Hence we need only consider $0 \leq A \leq I$. Next, taking $H = cI$, $c \in \mathbb{R}$,

$$\mathrm{Tr}(AH) - \ln\left(\mathrm{Tr}\left(e^H\right)\right) = c\mathrm{Tr}(A) - c - \ln(n) .$$

Unless $\mathrm{Tr}(A) = 1$, this tends to ∞ as c tends to ∞. Hence we need only consider the case that A is a density matrix ρ.

Hence, consider any $\rho \in \mathbf{S}_n$, and let H be any self-adjoint operator. In finite dimensions, necessarily $\mathrm{Tr}(e^H) < \infty$, and then we may define the density matrix σ

by

$$\sigma = \frac{e^H}{\mathrm{Tr}(e^H)} \ .$$

By Klein's inequality, $\mathrm{Tr}(\rho\ln\rho - \rho\ln\sigma) \geq 0$ with equality if and only if $\sigma = \rho$. But by the definition of σ, this reduces to

$$\mathrm{Tr}(\rho\ln\rho) \geq \mathrm{Tr}(\rho H) - \ln\left(\mathrm{Tr}\left(e^H\right)\right) \ ,$$

with equality if and only if $H = \ln\rho$. From here, there rest is very simple. ∎

As we have explained in the introduction, (1.14), which is now justified by Theorem 2.13, shows that the Gibbs states maximize the entropy given the expected value of the energy.

3. The joint convexity of certain operator functions

The route to the proof of the joint convexity of the relative entropy passes through the investigation of joint convexity for certain operator functions. This section treats three important examples.

3.1. The joint convexity of the map $(A,B) \mapsto B^*A^{-1}B^*$ **on** $\mathbf{H}_n^+ \times \mathbf{M}_n$**.** In this section we shall prove the joint convexity or joint concavity of certain operator functions. Our first example concerns the map $(A,B) \mapsto B^*A^{-1}B^*$ on $\mathbf{H}_n^+ \times \mathbf{M}_n$ which we shall show to be convex. Our next two examples concern the operator version of the harmonic and geometric means of two operators $A, B \in \mathbf{H}_n^+$. We shall show that these are jointly concave.

All three proofs follow the same pattern: In each of them, we show that the function in question has a certain maximality or minimality property, and we then easily prove the concavity or convexity as a consequence of this. All three proofs are taken from Ando's paper [**1**]. Here is the main theorem of this subsection:

3.1. Theorem. *The map* $(A,B) \mapsto B^*A^{-1}B$ *from* $\mathbf{H}_n^+ \times \mathbf{M}_n$ *to* $\mathbf{H}_n^+$ *is jointly convex. That is, for all* $(A_0,B_0),(A_1,B_1) \in \mathbf{H}_n^+ \times \mathbf{M}_n$*, and all* $0 < t < 1$*,*

(3.1)
$$[(1-t)B_0+tB_1]^*\frac{1}{(1-t)A_0+tA_1}[(1-t)B_0+tB_1] \leq (1-t)B_0^*A_0^{-1}B_0+tB_1^*A_1^{-1}B_1 \ .$$

We remark that as a special case, this yields another proof that $A \mapsto A^{-1}$ and $B \mapsto B^*B$ are convex.

The following lemma expresses a well-known minimality property of the functions $B \mapsto B^*A^{-1}B$.

3.2. Lemma. *Let* $A, C \in \mathbf{H}_n^+$ *with* A *invertible, and let* $B \in \mathbf{M}_n$*. Then the* $2n \times 2n$ *block matrix*

$$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$$

is positive semi-definite if and only if $C \geq B^*A^{-1}B$*.*

Proof: Define $D := C - B^*A^{-1}B^*$, so that

$$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix} = \begin{bmatrix} A & B \\ B^* & B^*A^{-1}B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & D \end{bmatrix} .$$

Now notice that

$$\begin{bmatrix} A & B \\ B^* & B^*A^{-1}B \end{bmatrix} = \begin{bmatrix} A^{1/2} & A^{-1/2}B \\ 0 & 0 \end{bmatrix}^* \begin{bmatrix} A^{1/2} & A^{-1/2}B \\ 0 & 0 \end{bmatrix} \geq 0 . \tag{3.2}$$

Hence, positive semi-definiteness of D is sufficient to ensure that $\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$ is positive semi-definite.

It is also evident from the factorization (3.2) that for any $v \in \mathbb{C}^n$, the vector $\begin{bmatrix} A^{-1}Bv \\ v \end{bmatrix}$ belongs to the null space of $\begin{bmatrix} A & B \\ B^* & B^*A^{-1}B \end{bmatrix}$, so that

$$\left\langle \begin{bmatrix} A^{-1}Bv \\ v \end{bmatrix}, \begin{bmatrix} A & B \\ B^* & B \end{bmatrix} \begin{bmatrix} A^{-1}Bv \\ v \end{bmatrix} \right\rangle = \langle v, Dv \rangle ,$$

and hence positive semi-definiteness of D is necessary to ensure that $\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$ is positive semi-definite. ■

Lemma 3.2 says that the set of matrices $C \in \mathbf{H}_n^+$ such that $\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$ is positive semi-definite has a *minimum*, namely $C = B^*A^{-1}B^*$. Form this, Ando draws a significant conclusion:

Proof of Theorem 3.1: By Lemma 3.2,

$$(1-t)\begin{bmatrix} A_0 & B_0 \\ B_0^* & B_0^*A_0^{-1}B_0 \end{bmatrix} + t\begin{bmatrix} A_1 & B_1 \\ B_1^* & B_1^*A_1^{-1}B_1 \end{bmatrix}$$

is a convex combination of positive semi-definite matrices, and is therefore positive semi-definite. It is also equal to

$$\begin{bmatrix} (1-t)A_0 + tA_1 & (1-t)B_0 + tB_1 \\ (1-t)B_0^* + tB_1^* & (1-t)B_0^*A_0^{-1}B_0 + tB_1^*A_1^{-1}B_1 \end{bmatrix} .$$

Now (3.1) follows by one more application of Lemma 3.2. ■

3.2. Joint concavity of the harmonic mean. Ando also uses Lemma 3.2 to prove a theorem of Anderson and Duffin on the concavity of the harmonic mean for operators.

3.3. Definition. For $A, B \in \mathbf{H}_n^+$, the *harmonic mean* of A and B, $M_{-1}(A,B)$ is defined by

$$M_{-1}(A,B) = \left(\frac{A^{-1} + B^{-1}}{2}\right)^{-1} .$$

3.4. Theorem (Joint concavity of the harmonic mean). *The map $(A,B) \mapsto M_{-1}(A,B)$ on $\mathbf{H}_n^+ \times \mathbf{H}_n^+$ is jointly concave.*

Proof: The key is the identity

$$M_{-1}(A,B) = 2B - 2B(A+B)^{-1}B \ . \tag{3.3}$$

Granted this, the jointly concavity of $M_{-1}(A,B)$ is a direct consequence of Theorem 3.1.

To prove (3.3), note that

$$\begin{aligned} B - B(A+B)^{-1}B &= (A+B)(A+B)^{-1}B - B(A+B)^{-1}B \\ &= A(A+B)^{-1}B \ , \end{aligned}$$

and also

$$(A(A+B)^{-1}B)^{-1} = B^{-1}(A+B)A^{-1} = A^{-1} + B^{-1} \ .$$

■

Here we have used the minimality property in Lemma 3.2 implicitly, but there is another way to proceed: It turns out that the harmonic mean has a certain maximality property:

3.5. Theorem (Ando's variational formula for the harmonic mean). *For all $A, B \in \mathbf{H}_n^+$, the set of all $C \in \mathbf{H}_n$ such that*

$$2\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} - \begin{bmatrix} C & C \\ C & C \end{bmatrix} \geq 0$$

has a maximal element, which is $M_{-1}(A,B)$.

We remark that this *maximum property* of the harmonic mean gives another proof of the concavity of the harmonic mean, just as the *minimum property* of $(A,B) \mapsto B^*A^{-1}B$ from Lemma 3.2 gave proof of the *convexity* of this function.

Proof of Theorem 3.5: Note that as a consequence of (3.4), (3.3) and the fact that $M_{-1}(A,B) = M_{-1}(B,A)$, we have

$$\begin{aligned} M_{-1}(B,A) &= [A - A(A+B)^{-1}A] + [B - B(A+B)^{-1}B] \\ &= A(A+B)^{-1}B + B(A+B)^{-1}A \ , \end{aligned} \tag{3.4}$$

from which it follows that

$$2M_{-1}(A,B) = (A+B) - (A-B)\frac{1}{A+B}(A-B) \ . \tag{3.5}$$

(Incidentally, we remark that from this identity one easily see the *harmonic-arithmetic mean inequality*: $M_{-1}(A,B) \leq (A+B)/2$ with equality if and only if $A = B$.)

Furthermore, by Lemma 3.2 , $\begin{bmatrix} A+B & A-B \\ A-B & A+B \end{bmatrix} - \begin{bmatrix} 2C & 0 \\ 0 & 0 \end{bmatrix} \geq 0$ if and only if

$$(A+B) - 2C \geq (A-B)(A+B)^{-1}(A-B) \ ,$$

and by (3.5), this is the case if and only if $C \leq M_{-1}(A,B)$. Finally, note that

$$\begin{bmatrix} A+B & A-B \\ A-B & A+B \end{bmatrix} - \begin{bmatrix} 2C & 0 \\ 0 & 0 \end{bmatrix} \geq 0 \iff 2\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} - \begin{bmatrix} C & C \\ C & C \end{bmatrix} \geq 0 \ .$$

■

3.3. Joint concavity of the geometric mean.

3.6. Definition. For $A, B \in \mathbf{H}_n^+$, the *geometric mean* of A and B, $M_0(A, B)$ is defined by

$$M_0(A, B) = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2} \ .$$

We note that if A and B commute, this definition reduces to $M_0(A, B) = A^{1/2}B^{1/2}$. While it is not immediately clear from the defining formula that in general one has that $M_0(A, B) = M_0(B, A)$, this is clear from the following variational formula of Ando:

3.7. Theorem (Ando's variational formula for the geometric mean). *For all $A, B \in \mathbf{H}_n^+$, the set of all $C \in \mathbf{H}_n$ such that*

$$\begin{bmatrix} A & C \\ C & B \end{bmatrix} \geq 0$$

has a maximal element, which is $M_0(A, B)$.

Proof: If $\begin{bmatrix} A & C \\ C & B \end{bmatrix} \geq 0$, then by Lemma 3.2, $B \geq CA^{-1}C$, and hence

$$A^{-1/2}BA^{-1/2} \geq A^{-1/2}CA^{-1}CA^{-1/2} = (A^{-1/2}CA^{-1/2})^2 \ .$$

By the operator monotonicity of the square root functions, which has been proved in Example 2.3 and as a special case of the Löwner-Heinz Theorem,

$$A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2} \leq C \ .$$

On the other hand, if $C = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$, then $B = CA^{-1}C$

This shows the maximality property of $M_0(A, B)$. ■

3.8. Theorem (Joint concavity of the geometric mean). *The map $(A, B) \mapsto M_0(A, B)$ on $\mathbf{H}_n^+ \times \mathbf{H}_n^+$ is jointly concave, and is symmetric in A and B. Moreover, for any non-singular matrix $D \in \mathbf{M}_n$,*

$$M_0(D^*AD, D^*BD) = D^*M_0(A, B)D \ . \tag{3.6}$$

Finally, $(A, B) \mapsto M_0(A, B)$ is monotone increasing in each variable.

Proof: The argument for the concavity is by now familiar. For (3.6), note that

$$\begin{bmatrix} A & C \\ C & B \end{bmatrix} > 0 \iff \begin{bmatrix} D^*AD & D^*CD \\ D^*CD & D^*BD \end{bmatrix} > 0 \ .$$

Finally, for fixed A, the fact that $B \mapsto A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2} = M_0(A, B)$ is monotone increasing is a direct consequence of the monotonicity of the square root function, which is contained in Theorem 2.6. By symmetry, for fixed B, $A \mapsto M_0(A, B)$ is monotone increasing. ■

3.4. The arithmetic-geometric-harmonic mean inequality.
Let $M_1(A,B)$ denote the arithmetic mean of A and B; i.e.,

$$M_1(A,B) = \frac{A+B}{2} \ .$$

3.9. THEOREM (Arithmetic-Geomemtric-Harmonic Mean Inequality). *For all $A, B \in \mathbf{H}_n^+$,*

$$M_1(A,B) \geq M_0(A,B) \geq M_{-1}(A,B) \ ,$$

with strict inequality everywhere unless $A = B$.

Proof: We first note that one can also use (3.5) to deduce that for all $A, B \in \mathbf{H}_n^+$ and nonsingular $D \in \mathbf{M}_n$,

$$M_{-1}(D^*AD, D^*BD) = D^*M_{-1}(A,B)D \tag{3.7}$$

in the same way that we deduced (3.6). However, (3.7) can also be deduced very simply from the formula that defines $M_{-1}(A,B)$.

We now show that (3.6) and (3.7) together reduce the proof to the corresponding inequality for numbers [**16**], which is quite elementary. To see this, take $D = A^{-1/2}$ and letting $L = A^{-1/2}BA^{-1/2}$, we have from the obvious companion for $M_1(A,B)$ to (3.6) and (3.7) that

$$\begin{aligned} M_1(A,B) - M_0(A,B) &= A^{1/2}[M_1(I,L) - M_0(I,L)]A^{1/2} \\ &= \frac{1}{2}A^{1/2}\left[I + L - 2L^{1/2}\right]A^{1/2} \\ &= \frac{1}{2}A^{1/2}(I - L^{1/2})^2A^{1/2} \ . \end{aligned}$$

The right hand side is evidently positive semi-definite, and even positive definite unless $L = I$, which is the case if and only if $A = B$. Likewise,

$$\begin{aligned} M_0(A,B) - M_{-1}(A,B) &= A^{1/2}[M_0(I,L) - M_{-1}(I,L)]A^{1/2} \\ &= A^{1/2}(L^{1/2} - 2(I + L^{-1})^{-1})^2A^{1/2} \ . \end{aligned}$$

The right hand side is positive semi-definite by the Spectral Theorem and the geometric-harmonic mean inequality for positive numbers, and even positive definite unless $L = I$, which is the case if and only if $A = B$. ■

4. Projections onto ∗-subalgebras and convexity inequalities

4.1. A simple example. The notion of operator convexity shall be useful to us when we encounter an operation on matrices that involves *averaging*; i.e., forming convex combinations. One is likely to encounter such operations much more frequently than one might first expect. It turns out that many natural operations on $\mathbf{M}_n$ can be written in the form

$$A \mapsto \mathcal{C}(A) := \sum_{j=1}^{N} w_j U_j A U_j^* \tag{4.1}$$

where the *weights* $w_1, \dots, w_N$ are positive numbers wtih $\sum_{j=1}^{N} w_j = 1$, and the $U_1, \dots, U_N$ are unitaries in $\mathbf{M}_n$. A basic example concerns orthogonal projections onto $*$-subalgebras, as we now explain.

A *unital $*$-subalgebra* of $\mathbf{M}_n$ is a subspace $\mathfrak{A}$ that contains the identity I, is closed under matrix multiplication and Hermitian conjugation: That is, if $A, B \in \mathfrak{A}$, then so are AB and A^*. *In what follows, all $*$-subalgebras that we consider shall be unital, and to simplify the notation, we shall simply write $*$-subalgebra in place of unital $*$-subalgebra.*

Perhaps the simplest example is

$$\mathfrak{A} = \left\{ A \ \in \mathbf{M}_2 \ : \ A = \begin{bmatrix} z & w \\ w & z \end{bmatrix} \quad w, z \in \mathbb{C} \right\} .$$

Since

$$\begin{bmatrix} x & y \\ y & x \end{bmatrix} \begin{bmatrix} \xi & \eta \\ \eta & \xi \end{bmatrix} = \begin{bmatrix} x\xi + y\eta & x\eta + y\xi \\ x\eta + y\xi & x\xi + y\eta \end{bmatrix} , \tag{4.2}$$

we see that $\mathfrak{A}$ is in fact closed under multiplication, and quite obviously it is closed under Hermitian conjugation. Moreover, one sees from (4.2) that

$$\begin{bmatrix} x & y \\ y & x \end{bmatrix} \begin{bmatrix} \xi & \eta \\ \eta & \xi \end{bmatrix} = \begin{bmatrix} \xi & \eta \\ \eta & \xi \end{bmatrix} \begin{bmatrix} x & y \\ y & x \end{bmatrix} ;$$

that is, the algebra $\mathfrak{A}$ is a *commutative* subalgebra of $\mathbf{M}_2$.

The main theme of this section is that: *Orthogonal projections onto subalgebras can be expressed in terms of averages over unitary conjugations, as in (4.1), and that this introduction of averages opens the way to the application of convexity inequalities.* By orthogonal, we mean of course orthogonal with respect to the Hilbert-Schmidt inner product. This theme was first developed by C. Davis [**8**]. Our treatment will be slightly different, and presented in a way that shall facilitate the transition to infinite dimensions.

In our simple example, it is particularly easy to see how the projection of $\mathbf{M}_2$ onto $\mathfrak{A}$ can be expressed in terms of averages. Let

$$Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} . \tag{4.3}$$

Clearly $Q = Q^*$ and $QQ^* = Q^2 = I$, so that Q is both self-adjoint and unitary. Notice that for any $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_2$,

$$\frac{1}{2}(A + QAQ^*) = \frac{1}{2}\begin{bmatrix} a+d & b+c \\ b+c & a+d \end{bmatrix} \in \mathfrak{A} .$$

Let us denote $\frac{1}{2}(A + QAQ^*)$ by $\mathbf{E}_{\mathfrak{A}}(A)$ for reasons that shall soon be explained.

One can easily check that $\mathbf{E}_{\mathfrak{A}}(A)$ is the orthogonal projection of A onto $\mathfrak{A}$. Indeed, it suffices to check by direct computation that $\mathbf{E}_{\mathfrak{A}}(A)$ and $A - \mathbf{E}_{\mathfrak{A}}(A)$ are orthogonal in the Hilbert-Schmidt inner product.

The fact that

$$\mathbf{E}_{\mathfrak{A}}(A) = \frac{1}{2}\left(A + QAQ^*\right)$$

is an average over unitary conjugations of A means that if f is any operator convex function, then

$$\begin{aligned}\mathbf{E}_{\mathfrak{A}}[f(A)] = \frac{f(A) + Qf(A)Q^*}{2} &= \frac{f(A) + f(QAQ^*)}{2} \\ &\geq f\left(\frac{A + QAQ^*}{2}\right) = f\left(\mathbf{E}_{\mathfrak{A}}[A]\right) \ .\end{aligned}$$

Let us look ahead and consider an important application of this line of reasoning. Consider relative entropy function $A, B \mapsto H(A|B) = \mathrm{Tr}[A\log A] - \mathrm{Tr}[A\log(B)]$. In Theorem 6.3, this function shall be proved to be jointly convex on $\mathbf{H}_n^+ \times \mathbf{H}_n^+$. It is also clearly unitarily invariant; i.e., for any $n \times n$ unitary matrix U,

$$H(UAU^*|UBU^*) = H(A|B) \ .$$

It then follows that, for $n = 2$, and $\mathfrak{A}$ being the subalgebra of $\mathbf{M}_2$ defined above,

$$\begin{aligned}H(A|B) &= \frac{H(A|B) + H(QAQ^*|QBQ^*)}{2} \\ &\geq H\left(\frac{A + QAQ^*}{2} \,\middle|\, \frac{B + QBQ^*}{2}\right) \\ &= H(\mathbf{E}_{\mathfrak{A}}(A)|\mathbf{E}_{\mathfrak{A}}(B)) \ .\end{aligned}$$

It turns out that there is nothing very special about the simple example we have been discussing: if $\mathfrak{A}$ is *any* $*$-subalgebra of $\mathbf{M}_n$, the orthogonal projection onto $\mathfrak{A}$ can be expressed in terms of averages of unitary conjugations, and from this fact we shall be able to conclude a number of very useful convexity inequalities.

The notation $\mathbf{E}_{\mathfrak{A}}$ for the orthogonal projection onto a $*$-subalgebra reflects a close analogy with the operation of taking *conditional expectations* in probability theory: Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, and suppose that $\mathcal{S}$ is a sub-σ-algebra of $\mathcal{F}$. Then $L^2(\Omega, \mathcal{S}, \mu)$ will be a closed subspace of $L^2(\Omega, \mathcal{F}, \mu)$, and if X is any bounded random variable on $(\Omega, \mathcal{F}, \mu)$; i.e., any function on Ω that is measurable with respect to $\mathcal{F}$, and essentially bounded with respect to μ, the *conditional expectation of X given $\mathcal{S}$* is the orthogonal projection of X, which belongs to $L^2(\Omega, \mathcal{S}, \mu)$, onto $L^2(\Omega, \mathcal{F}, \mu)$. The bounded measurable functions on $(\Omega, \mathcal{F}, \mu)$ of course form an commutative $*$-algebra (in which the $*$ operation is pointwise complex conjugation), of which the bounded measurable functions on $(\Omega, \mathcal{S}, \mu)$ form a commutative $*$-subalgebra. The non-commutative analog of the conditional expectation that we now develop is more than an analog; it is part of a far reaching non-commutative extension of probability theory, and integration in general, due to Irving Segal [**28, 30**].

4.2. The von Neumann Double Commutant Theorem.

4.1. Definition (Commutant). Let $\mathcal{A}$ be any subset of $\mathbf{M}_n$. Then $\mathcal{A}'$, the *commutant* of $\mathcal{A}$, is given by

$$\mathcal{A}' = \{ B \in \mathbf{M}_n \ : \ BA = AB \quad \text{for all} \quad A \in \mathcal{A} \} .$$

It is easy to see that for any set $\mathcal{A}$, $\mathcal{A}'$ is a subalgebra of $\mathbf{M}_n$, and if $\mathcal{A}$ is closed under Hermitian conjugation, then $\mathcal{A}'$ is a $*$-subalgebra of $\mathbf{M}_n$.

In particular, if $\mathfrak{A}$ is a $*$-subalgebra of $\mathbf{M}_n$, then so is $\mathfrak{A}'$, the commutant of $\mathfrak{A}$. Continuing in this way, the double commutant $\mathfrak{A}''$ is also a $*$-subalgebra of $\mathbf{M}_n$, but it is nothing new:

4.2. Theorem (von Neumann Double Commutant Theorem). *For any $*$-subalgebra $\mathfrak{A}$ of $\mathbf{M}_n$,*

$$\mathfrak{A}'' = \mathfrak{A} . \tag{4.4}$$

Proof: We first show that for any $*$-subalgebra $\mathfrak{A}$, and any $B \in \mathfrak{A}''$ and any $v \in \mathbb{C}^n$, there exists an $A \in \mathfrak{A}$ such that

$$Av = Bv . \tag{4.5}$$

Suppose that this has been established. We then apply it to the $*$-subalgebra $\mathfrak{M}$ of $\mathbf{M}_{n^2}$ consisting of diagonal block matrices of the form

$$\begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ 0 & & \ddots & 0 \\ 0 & 0 & \dots & A \end{bmatrix} = A \otimes I_{n\times n} \ , \quad A \in \mathfrak{A} .$$

It is then easy to see that $\mathfrak{M}''$ consists of diagonal block matrices of the form

$$\begin{bmatrix} B & 0 & \dots & 0 \\ 0 & B & \dots & 0 \\ 0 & & \ddots & 0 \\ 0 & 0 & \dots & B \end{bmatrix} = B \otimes I_{n\times n} \ , \quad B \in \mathfrak{A}'' .$$

Now let $\{v_1, \dots, v_n\}$ be any basis of $\mathbb{C}^n$, and form the vector $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{C}^{n^2}$.

Then

$$(A \otimes I_{n\times n})v = (B \otimes I_{n\times n})v \quad \Rightarrow \quad Av_j = Bv_j \quad j = 1, \dots, n .$$

Since $\{v_1, \dots, v_n\}$ is a basis of $\mathbb{C}^n$, this means $B = A \in \mathfrak{A}$. Since B was an arbitrary element of $\mathfrak{A}''$, this shows that

$$\mathfrak{A}'' \subset \mathfrak{A} .$$

Since $\mathfrak{A} \subset \mathfrak{A}''$ is an automatic consequence of the definitions, this shall prove that $\mathfrak{A} = \mathfrak{A}''$.

Therefore, it remains to prove (4.5). Fix any $v \in \mathbb{C}^n$, and let V be the subspace of $\mathbb{C}^n$ given by

$$V = \{ \ Av \ : \ A \in \mathfrak{A} \ \} . \tag{4.6}$$

Let P be the orthogonal projection onto V in $\mathbb{C}^n$. Since, by construction, V is invariant under the action of $\mathfrak{A}$, $PAP = AP$ for all $A \in \mathfrak{A}$. Taking Hermitian conjugates, $PA^*P = PA^*$ for all $A \in \mathfrak{A}$. Since $\mathfrak{A}$ is a $*$-algebra, this imples $PA = AP$ for all $A \in \mathfrak{A}$. That is, $P \in \mathfrak{A}'$.

Thus, for any $B \in \mathfrak{A}''$, $BP = PB$, and so V is invariant under the action of $\mathfrak{A}''$. In particular, $Bv \in V$, and hence, by the definition of V, $Bv = Av$ for some $A \in \mathfrak{A}$. ■

4.3. Remark. *von Neumann proved his double commutant theorem* [**37**] *for operators on an infinite dimensional Hilbert space, but all of the essential aspects are present in the proof of the finite dimensional specialization presented here. The relevant difference between finite and infinite dimensions is, of course, that in finite dimensional all subspaces are closed, while this is not the case in infinite dimensions.*

Thus in the infinite dimensional case, we would have to replace (4.6) by

$$V = \overline{\{\ Av\ :\ A \in \mathfrak{A}\ \}}\ , \tag{4.7}$$

taking the closure of $\{\ Av\ :\ A \in \mathfrak{A}\ \}$. *The same proof would then lead to the conclusion that for all* $B \in \mathfrak{A}''$, Bv *lies in the closure of* $\{\ Av\ :\ A \in \mathfrak{A}\ \}$. *Thus one concludes that* $\mathfrak{A}'' = \mathfrak{A}$ *if and only if* $\mathfrak{A}$ *is closed in the weak operator topology, which is the usual formulation in the infinite dimensional context.*

4.4. Example. The $*$-subalgebra $\mathfrak{A}$ of $\mathbf{M}_2$ from subsection 4.1 is spanned by $I_{2\times 2}$ and Q, where Q is given by (4.3). This $*$-subalgebra happens to be commutative, and so it is certainly the case that $\mathfrak{A} \subset \mathfrak{A}'$ – a feature that is special to the commutative case. In fact, one can easily check that $AQ = QA$ if and only if $A \in \mathfrak{A}$, and so $\mathfrak{A} = \mathfrak{A}'$. It is then obviously the case that $\mathfrak{A} = \mathfrak{A}''$, as von Neumann's theorem tells us. □

4.5. Lemma. *Let* $\mathfrak{A}$ *be a* $*$*-subalgebra of* $\mathbf{M}_n$. *For any self-adjoint* $A \in \mathfrak{A}$ *let* $A = \sum_{j=1}^m \lambda_j P_j$ *be the spectral decomposition of* A. *Then each of the spectral projections* P_j, $j = 1, \dots, m$ *belongs to* $\mathfrak{A}$. *Moreover, each* $A \in \mathfrak{A}$ *can be written as a linear combination of at most* 4 *unitary matrices, each of which belongs to* $\mathfrak{A}$.

Proof: It is easy to see that

$$P_j = \prod_{i\in\{1,\dots,n\}\backslash\{j\}} \frac{1}{\lambda_i - \lambda_j}(\lambda_i - A)\ . \tag{4.8}$$

As a polynomial in A, this belongs to $\mathfrak{A}$.

If furthermore A is a self adjoint contraction, then each λ_j lies in $[-1, 1]$, and hence $\lambda_j = \cos(\theta_j)$ for some $\theta \in [0, \pi]$. In this case we may write

$$A = \sum_{j=1}^m \lambda_j P_j = \frac{1}{2}\left(\sum_{j=1}^m e^{i\theta_j}P_j + \sum_{j=1}^m e^{-i\theta_j}P_j\right)\ .$$

Note that $U = \sum_{j=1}^m e^{i\theta_j}P_j$ is unitary, and since it is a linear combination of elements of $\mathfrak{A}$, $U \in \mathfrak{A}$. Thus we have seen that every self-adjoint contraction in $\mathfrak{A}$

is of the form

$$A = \frac{1}{2}(U + U^*) \qquad U \in \mathfrak{A} \ .$$

Now for general $A \in \mathfrak{A}$, write

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2i}i(A - A^*) \ .$$

which expresses A as a linear combination of self-adjoint elements of $\mathfrak{A}$. From here, the rest easily follows. ■

4.6. Lemma. *Let $\mathfrak{A}$ be a $*$-subalgebra of $\mathbf{M}_n$. Then for any $A \in \mathbf{M}_n$*

$$A \in \mathfrak{A} \quad \Longleftrightarrow \quad A = UAU^* \quad \text{for all} \quad U \in \mathfrak{A}' \ . \tag{4.9}$$

Proof: Since for unitary U, $A = UAU^*$ if and only if $UA = AU$, the condition that $A = UAU^*$ for all $U \in \mathfrak{A}'$ amounts to the condition that A commutes with every unitary matrix in $\mathfrak{A}'$. But by the previous lemma, commuting with every unitary matrix in $\mathfrak{A}'$ is the same thing as commuting with every matrix in $\mathfrak{A}'$. Thus

$$A = UAU^* \quad \text{for all} \quad U \in \mathfrak{A}' \quad \Longleftrightarrow \quad A \in \mathfrak{A}'' \ .$$

Then by the von Neumann Double Commutant Theorem, (4.9) follows. ■

The fact that all $*$-subalgebras of $\mathbf{M}_n$ contain "plenty of projections, and plenty of unitaries", as expressed by Lemma 4.5, is often useful. As we shall see, there is another important sense in which $*$-subalgebra of $\mathbf{M}_n$ are rich in unitaries. We first recall the *polar factorization* of a matrix $A \in \mathbf{M}_n$.

4.7. Lemma. *For any matrix $A \in \mathbf{M}_n$, let $|A| = (A^*A)^{1/2}$. Then there is a unique partial isometry $U \in \mathbf{M}_n$ such that $A = U|A|$, U is an isometry from the range of A^* onto the range of A, and U is zero on the nullspace of A. If A is invertible, U is unitary, and in any case, for all $v \in \mathbb{C}^n$,*

$$Uv = \lim_{\epsilon \to 0} A(A^*A + \epsilon I)^{-1/2} v \ .$$

We leave the easy proof to the reader. Now let $p_n(t)$ be a sequence of polynomials such that $\lim_{n\to\infty} p_n(t) = \sqrt{t}$, uniformly on an interval containing the spectrum of A^*A. Then

$$|A| = \lim_{n\to\infty} p_n(A^*A) \ .$$

Now, if $\mathfrak{A}$ is any $*$-subalgebra of $\mathbf{M}_n$, and A any matrix in $\mathfrak{A}$, then for each n, $p_n(A^*A) \in \mathfrak{A}$, and hence $|A| \in \mathfrak{A}$. The same argument shows that for each $\epsilon > 0$, $(A^*A + \epsilon I)^{1/2} \in \mathfrak{A}$. We now claim that $(A^*A + \epsilon I)^{-1/2} \in \mathfrak{A}$ as well. Indeed:

4.8. Lemma. *Let $\mathfrak{A}$ be any $*$-subalgebra of $\mathbf{M}_n$, and let $B \in \mathbf{H}_n^+$ belong to $\mathfrak{A}$. Then the inverse of B also belongs to $\mathfrak{A}$.*

Proof: The spectrum of $\|B\|^{-1}B$ lies in the interval $(0, 1]$, and hence $\|I - \|B\|^{-1}B\| < 1$. Thus,

$$(\|B\|B)^{-1} = [I - (I - \|B\|B)]^{-1} = \sum_{n=0}^{\infty}(I - \|B\|B)^n \ ,$$

and by the above, each term in the convergent power series on the right belongs to $\mathfrak{A}$. Thus $(\|B\|B)^{-1}$ belongs to $\mathfrak{A}$, and hence so does B^{-1}. ■

Thus for each $\epsilon > 0$,

$$A \in \mathfrak{A} \quad \Rightarrow \quad A(A^*A + \epsilon I)^{-1/2} \in \mathfrak{A} \ .$$

Taking limits, we see that if $A = U|A|$ is the polar factorization of A, then both U and $|A|$ belong to $\mathfrak{A}$. We can now improve Lemma 4.8:

4.9. THEOREM. *Let $\mathfrak{A}$ be any $*$-subalgebra of $\mathbf{M}_n$. Then for all $A \in \mathfrak{A}$ such that A is invertible in $\mathbf{M}_n$, A is invertible in $\mathfrak{A}$; i.e., the inverse of A belongs to $\mathfrak{A}$.*

Proof: Let A be invertible in $\mathbf{M}_n$, and let $A = U|A|$ be the polar factorization of A. Then $A^{-1} = |A|^{-1}U^*$. Since $U \in \mathfrak{A}$, which is a $*$-subalgebra, $U^* \in \mathfrak{A}$ as well. Since A is invertible, so is $|A|$, and we have seen that $|A| \in \mathfrak{A}$. Then by Lemma 4.8, $|A|^{-1} \in \mathfrak{A}$. Altogether, we have that

$$A^{-1} = |A|^{-1}U^* \in \mathfrak{A} \ .$$

■

4.3. Properties of the conditional expectation.

4.10. DEFINITION. For any $*$-subalgebra $\mathfrak{A}$ of $\mathbf{M}_n$, let $\mathbf{E}_{\mathfrak{A}}$ be the orthogonal projection, with respect to the Hilbert-Schmidt inner product of $\mathbf{M}_n$ onto $\mathfrak{A}$ (which is a closed subspace of $\mathbf{M}_n$.) We refer to $\mathbf{E}_{\mathfrak{A}}$ as the *conditional expectation given* $\mathfrak{A}$.

4.11. EXAMPLE. Let $\{u_1, \dots, u_n\}$ be any orthonormal basis of $\mathbb{C}^n$. Let $\mathfrak{A}$ be the subalgebra of $\mathbf{M}_n$ consisting of matrices that are diagonal in this basis; i.e., $A \in \mathfrak{A}$ if and only if $A = \sum_{j=1}^n a_j u_j u_j^*$ for some $a_1, \dots, a_n \in \mathbb{C}$. Here, we are writing $u_j u_j^*$ to denote the orthogonal projection onto the span of the unit vector u_j. In the usual physics notation, this is denoted by $|u_j\rangle\langle u_j|$. We shall use these notations interchangeably.

For any $B \in \mathbf{M}_n$, the matrix $\widetilde{B} := \sum_{j=1}^n \langle u_j, Bu_j\rangle |u_j\rangle\langle u_j| \in \mathfrak{A}$, and moreover, for any $A \in \mathfrak{A}$,

$$\mathrm{Tr}[A(B - \widetilde{B})] = \sum_{j=1}^n \langle u_j, A(B - \widetilde{B})u_j\rangle = 0 \ ,$$

and so $\widetilde{B}$ is the orthogonal projection of B onto $\mathfrak{A}$. Thus, we have the formula

$$\mathbf{E}_{\mathfrak{A}}(B) = \sum_{j=1}^n \langle u_j, Bu_j\rangle |u_j\rangle\langle u_j| \tag{4.10}$$

for all $B \in \mathbf{M}_n$, where in the usual physical notation, $\rangle\langle u_j|$ denotes the orthogonal projection onto the span of u_j. □

The next result is based on the projection lemma, which is very useful also in finite dimensions! For the sake of completeness, we give the simple proof:

4.12. THEOREM (Projection Lemma). *Let K be a closed convex set in a Hilbert space. Then K contains a unique element of minimal norm. That is, there exists $v \in K$ such that $\|v\| < \|w\|$ for all $w \in K$, $w \neq v$.*

Proof: Let $D := \inf\{\|w\| \ : \ w \in K \ \}$. If $D = 0$, then $0 \in K$ since K is closed, and this is the unique element of minimal norm. Hence we may suppose that $D > 0$. Let $\{w_n\}_{n\in\mathbb{N}}$ be a sequence in K such that $\lim_{n\to\infty} \|w_n\| = D$. By the parallelogram identity

$$\left\|\frac{w_m + w_n}{2}\right\|^2 + \left\|\frac{w_m - w_n}{2}\right\|^2 = \frac{\|w_m\|^2 + \|w_n\|^2}{2} \ .$$

By the convexity of K, and the definition of D, $\left\|\dfrac{w_m + w_n}{2}\right\|^2 \geq D^2$ and so

$$\left\|\frac{w_m - w_n}{2}\right\|^2 \leq \frac{\left(\|w_m\|^2 - D^2\right) + \left(\|w_n\|^2 - D^2\right)}{2} \ .$$

By construction, the right side tends to zero, and so $\{w_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. Then, by the completeness that is a defining property of Hilbert spaces, $\{w_n\}_{n\in\mathbb{N}}$ is a convergent sequence. Let v denote the limit. By the continuity of the norm, $\|v\| = \lim_{n\to\infty} \|w_n\| = D$. Finally, if u is any other vector in K with $\|u\| = D$, $(u + v)/2 \in K$, so that $\|(u + v)/2\| \geq D$. Then by the parallelogram identity once more $\|(u - v)/2\| = 0$, and so $u = v$. This proves the uniqueness. ∎

4.13. THEOREM. *For any $A \in \mathbf{M}_n$, and any $*$-subalgebra $\mathfrak{A}$ of $\mathbf{M}_n$, let K_A denote the closed convex hull of the operators UAU^*, $U \in \mathfrak{A}'$. Then $\mathbf{E}_{\mathfrak{A}}(A)$ is the unique element of minimal (Hilbert-Schmidt) norm in K_A. Furthermore,*

$$\mathrm{Tr}[\mathbf{E}_{\mathfrak{A}}(A)] = \mathrm{Tr}[A] \ , \tag{4.11}$$

$$A > 0 \quad \Rightarrow \quad \mathbf{E}_{\mathfrak{A}}(A) > 0 \ , \tag{4.12}$$

and for each $B \in \mathfrak{A}$,

$$\mathbf{E}_{\mathfrak{A}}(BA) = B\mathbf{E}_{\mathfrak{A}}(A) \qquad \text{and} \qquad \mathbf{E}_{\mathfrak{A}}(AB) = \mathbf{E}_{\mathfrak{A}}(A)B \ . \tag{4.13}$$

Proof: We apply the Projection Lemma in $\mathbf{M}_n$ equipped with the Hilbert-Schmidt inner product, so that it becomes a Hilbert space, and we may then apply the Projection Lemma. For each $A \in \mathbf{M}_n$, let $\widetilde{A}$ denote the unique element of minimal norm in K_A, and let $U \in \mathfrak{A}'$ be unitary. Then by the parallelogram law,

$$\left\|\frac{\widetilde{A} + U\widetilde{A}U^*}{2}\right\|^2 + \left\|\frac{\widetilde{A} - U\widetilde{A}U^*}{2}\right\|^2 = \frac{\|\widetilde{A}\|^2 + \|U\widetilde{A}U^*\|^2}{2} = \|\widetilde{A}\|^2 \ .$$

Since $(\widetilde{A} + U\widetilde{A}U^*)/2 \in K_A$, $\|(\widetilde{A} + U\widetilde{A}U^*)/2\| \geq \|\widetilde{A}\|$, the minimal norm in K_A, and hence

$$\left\|\frac{\widetilde{A} - U\widetilde{A}U^*}{2}\right\|^2 = 0 \ .$$

This means that $\widetilde{A} = U\widetilde{A}U^*$ for all unitary $U \in \mathfrak{A}'$. By Lemma 4.6, this means that $\widetilde{A} \in \mathfrak{A}$.

Next we claim that

$$\langle B, A - \widetilde{A}\rangle = 0 \qquad \text{for all} \quad B \in \mathfrak{A}\ , \tag{4.14}$$

which, together with the fact that $\widetilde{A} \in \mathfrak{A}$ identifies $\widetilde{A}$ as the orthogonal projection of A onto $\mathfrak{A}$.

To prove (4.14) note that for $B \in \mathfrak{A}$ and $U_j \in \mathfrak{A}'$, $U(AB)U^* = UAU^*B$ Hence if $\sum_{j=1}^N w_jU_j(AB)U_j^*$ is any convex combination of unitary conjugations of AB with each $U_j \in \mathfrak{A}'$,

$$\sum_{j=1}^N w_jU_j(AB)U_j^* = \left(\sum_{j=1}^N w_jU_jAU_j^*\right) B\ .$$

It readily follows that

$$\widetilde{AB} = \widetilde{A}B\ . \tag{4.15}$$

Now observe that since unitary conjugation preserves the trace, each element of K_A has the same trace, namely the trace of A. In particular, for all $A \in \mathbf{M}_n$,

$$\mathrm{Tr}[\widetilde{A}] = \mathrm{Tr}[A]\ . \tag{4.16}$$

Combining this with (4.15) yields

$$\begin{aligned} 0 &= \mathrm{Tr}[\widetilde{AB}] - \mathrm{Tr}[AB] \\ &= \mathrm{Tr}[\widetilde{A}B] - \mathrm{Tr}[AB] \\ &= \mathrm{Tr}[(\widetilde{A} - A)B]\ . \end{aligned}$$

Since B is an arbitrary element of $\mathfrak{A}$, this proves (4.14). Thus, $\widetilde{A}$ is the orthogonal projection of A onto $\mathfrak{A}$. Now that we know $\widetilde{A} = \mathbf{E}_{\mathfrak{A}}(A)$, (4.11) follows from (4.16), and the identity on the right in (4.13) now follows from (4.15), and then the identity on the left follows by Hermitian conjugation. Finally, if $A > 0$, so is each UAU^*, and hence so is each member of K_A, including the element of minimal norm, $\mathbf{E}_{\mathfrak{A}}(A)$. ■

4.14. Remark. *Theorem 4.13 says that for all* $A \in \mathbf{M}_n$*, all* $*$*-subalgebras* $\mathfrak{A}$ *of* $\mathbf{M}_n$ *and all* $\epsilon > 0$*, there exists some set of* N *unitaries* $U_1, \dots, U_N \in \mathfrak{A}'$ *and some set of* N *weights,* $w_1, \dots, w_N$*, non-negative and summing to one, such that*

$$\|\mathrm{E}_{\mathfrak{A}}(A) - \sum_{j=1}^N w_jU_jAU_j^*\| \le \epsilon\ . \tag{4.17}$$

In fact, in finite dimensional settings, one can often avoid the limiting procedure and simply write

$$\mathrm{E}_{\mathfrak{A}}(A) = \sum_{j=1}^N w_jU_jAU_j^*\ , \tag{4.18}$$

as an exact equality. An important instance is provided in the next example. However, while the statement of Theorem 4.13 is couched in finite dimensional terms, this is only for simplicity of exposition: The proof makes no reference to the finite *dimension of $\mathbf{M}_n$, and the approximation of conditional expectations provided by (4.17) is valid – and useful – in infinite dimensions as well.*

4.15. EXAMPLE. As in Example 4.11, let $\{u_1, \dots, u_n\}$ be any orthonormal basis of $\mathbb{C}^n$, and let $\mathfrak{A}$ be the subalgebra of $\mathbf{M}_n$ consisting of matrices that are diagonal in this basis. There we derived an explicit formula (4.10) for $\mathbf{E}_\mathfrak{A}$. Theorem 4.13 says that there exists an alternate expression of $\mathbf{E}_\mathfrak{A}$ as a limit of averages of unitary conjugations. In fact, it can be expressed as an average over n unitary conjugations, and no limit is needed in this case.

To see this, for $k = 1, \dots, n$ define the unitary matrix U_k by

$$U_k = \sum_{\ell=1}^{n} e^{i2\pi \ell k/n} |u_\ell\rangle\langle u_\ell| \ . \tag{4.19}$$

Then, for any $B \in \mathbf{M}_n$, $U_k B U_k^* = \sum_{\ell,m=1}^{n} \langle u_m B u_\ell\rangle e^{i2\pi(m-\ell)k/n} |u_m\rangle\langle u_\ell|$. Therefore, averaging over k, and then swapping orders of summation,

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} U_k B U_k^* &= \sum_{\ell,m=1}^{n} \langle u_m B u_\ell\rangle \left(\frac{1}{n}\sum_{k=1}^{n} e^{i2\pi(m-\ell)k/n}\right) |u_m\rangle\langle u_\ell| \\ &= \sum_{m=1}^{n} \langle u_m B u_m\rangle |u_m\rangle\langle u_m| = \mathbf{E}_\mathfrak{A}(B) \ .\end{aligned}$$

In summary, with $U_1, \dots, U_n$ defined by (4.19), and $\mathfrak{A}$ being the $*$-subalgebra of $\mathbf{M}_n$ consisting of matrices that are diagonalized by $\{u_1, \dots, u_n\}$,

$$\mathbf{E}_\mathfrak{A}(B) = \frac{1}{n}\sum_{k=1}^{n} U_k B U_k^* \ . \tag{4.20}$$

In other words, the "diagonal part" of B is an average over n unitary conjugations of B. □

Theorem 4.13 is the source of many convexity inequalities for trace functions. Here is one of the most important:

4.16. THEOREM. *Let f be any operator convex function. Then for any $*$-subalgebra $\mathfrak{A}$ of $\mathbf{M}_n$, and any self-adjoint operator $A \in \mathbf{M}_n$,*

$$f(\mathbf{E}_\mathfrak{A}(A)) \leq \mathbf{E}_\mathfrak{A}(f(A)) \ .$$

Proof: Since $f(\mathbf{E}_\mathfrak{A}(A)) - \mathbf{E}_\mathfrak{A}(f(A))$ is a self-adjoint operator in $\mathfrak{A}$, all of its spectral projections are in $\mathfrak{A}$, and thus, it suffices to show that for every orthogonal projection P in $\mathfrak{A}$,

$$\mathrm{Tr}\,[P f(\mathbf{E}_\mathfrak{A}(A))] \leq \mathrm{Tr}\,[P\mathbf{E}_\mathfrak{A}(f(A))] \ . \tag{4.21}$$

But, since $P \in \mathfrak{A}$,

$$\mathrm{Tr}\,[P\mathbf{E}_\mathfrak{A}(f(A))] = \mathrm{Tr}\,[P f(A)] \ . \tag{4.22}$$

Next, by Theorem 4.13, for any $A \in \mathbf{M}_n$, $\mathbf{E}_{\mathfrak{A}}(A)$ is a limit of averages of unitary conjugates of A. That is $\mathbf{E}_{\mathfrak{A}}(A) = \lim_{k\to\infty} \mathcal{C}_k(A)$, where each $\mathcal{C}_k(A)$ has the form

$$\mathcal{C}_k(A) = \sum_{j=1}^{N_k} p_{k,j} U_{n,j} A U_{k,j}^* \tag{4.23}$$

and where for each k, j $U_{k,j}$ is a unitary in $\mathfrak{A}'$, $p_{k,j} > 0$, and $\sum_{j=1}^{N_k} p_{k,j} = 1$. Then, by the operator convexity of f,

$$f(\mathcal{C}_k(A)) \leq \left(\sum_{j=1}^{N_k} p_{k,j} U_{k,j} f(A) U_{k,j}^* \right) \ ,$$

and then since $P \in \mathfrak{A}$ and $U_{k,j} \in \mathfrak{A}'$ for each k, j,

$$\begin{aligned} \mathrm{Tr}[Pf(\mathcal{C}_k(A))] &\leq \sum_{j=1}^{N_k} p_{k,j} \mathrm{Tr}\left[U_{k,j} P f(A) U_{k,j}^*\right] \\ &= \sum_{j=1}^{N_k} p_{k,j} \mathrm{Tr}\left[Pf\,(A)\right] \\ &= \mathrm{Tr}[Pf(A)] \ . \end{aligned}$$

Therefore,

$$\mathrm{Tr}\left[Pf(\mathbf{E}_{\mathfrak{A}}(A))\right] = \lim_{k\to\infty} \mathrm{Tr}[Pf(\mathcal{C}_k(A))] \leq \mathrm{Tr}[Pf(A)] \ .$$

Combining this with (4.22) proves (4.21). ■

Let us apply this to the von Neumann entropy. First of all, note that since $\mathbf{E}_{\mathfrak{A}}$ is the orthogonal projection onto $\mathfrak{A}$, it is continuous. Thus, $\mathbf{E}_{\mathfrak{A}}$ not only preserves the class of positive definite operators, as asserted in (4.12), it also preserves the class of positive semidefinite operators:

$$A \geq 0 \quad \Rightarrow \quad \mathbf{E}_{\mathfrak{A}}(A) \geq 0 \ . \tag{4.24}$$

This, together with (4.11) implies that if $\rho \in \mathbf{M}_n$ is a density matrix, then so is $\mathbf{E}_{\mathfrak{A}}(\rho)$.

Now we may apply Theorem 4.16 to see that for any $*$-subalgebra of $\mathbf{M}_n$, and any $\rho \in \mathbf{S}_n$, the von Neumann entropy of $\mathbf{E}_{\mathfrak{A}}(\rho)$, $S(\mathbf{E}_{\mathfrak{A}}(\rho))$, is no less than the von Neumann entropy of ρ, $S(\rho)$:

$$S(\mathbf{E}_{\mathfrak{A}}(\rho)) \geq S(\rho) \ .$$

In later sections, where we shall encounter *jointly convex* functions on $\mathbf{M}_n$, we shall apply the same sort of reasoning. Towards that end, the following simple observation is useful: For any A, B in $\mathbf{M}_n$, there is a *single* sequence $\{\mathcal{C}_k\}_{k\in\mathbb{N}}$ of operators of the form (4.23) such that *both*

$$\mathbf{E}_{\mathfrak{A}}(A) = \lim_{k\to\infty} \mathcal{C}_k(A) \qquad \text{and} \qquad \mathbf{E}_{\mathfrak{A}}(B) = \lim_{k\to\infty} \mathcal{C}_k(B) \ .$$

There is a very simple way to see that this is possible: Consider the $*$-subalgebra $\mathbf{M}_2(\mathfrak{A})$ of $\mathbf{M}_{2n}$ consisting of block matrices of the form

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad , \quad A, B, C, D \in \mathfrak{A} \ .$$

Then the same computations which show that the only matrices in $\mathbf{M}_2$ that commute with all other matrices in $\mathbf{M}_2$ are multiples of the identity show that $(\mathbf{M}_2(\mathfrak{A}))'$ consists of matrices of the form

$$\begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} \quad , \quad X \in \mathfrak{A}' \ ,$$

and hence the unitaries in $(\mathbf{M}_2(\mathfrak{A}))'$ have the form

$$\begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} \quad , \quad U \in \mathfrak{A}' \ , UU^* = I \ ,$$

One readily computes that for any $A, B, C, D \in \mathbf{M}_n$ ($\mathbf{M}_n$ now, not only $\mathfrak{A}$), and any $U \in \mathfrak{A}'$,

$$\begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^* = \begin{bmatrix} UAU^* & UBU^* \\ UCU^* & UDU^* \end{bmatrix} \ .$$

and moreover,

$$\left\| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\|_{\mathrm{HS}}^2 = \|A\|_{\mathrm{HS}}^2 + \|B\|_{\mathrm{HS}}^2 + \|C\|_{\mathrm{HS}}^2 + \|D\|_{\mathrm{HS}}^2 \ .$$

From this and Theorem 4.13, one readily concludes that

$$\mathbf{E}_{\mathbf{M}_2(\mathfrak{A})}\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix} \right) = \begin{bmatrix} \mathbf{E}_{\mathfrak{A}}(A) & \mathbf{E}_{\mathfrak{A}}(B) \\ \mathbf{E}_{\mathfrak{A}}(C) & \mathbf{E}_{\mathfrak{A}}(D) \end{bmatrix} \ ,$$

and that there exists a sequence $\{\mathcal{C}_k\}_{k\in\mathbb{N}}$ of operators of the form (4.23) such that

$$\begin{bmatrix} \mathbf{E}_{\mathfrak{A}}(A) & \mathbf{E}_{\mathfrak{A}}(B) \\ \mathbf{E}_{\mathfrak{A}}(C) & \mathbf{E}_{\mathfrak{A}}(D) \end{bmatrix} = \lim_{k\to\infty} \begin{bmatrix} \mathcal{C}_k(A) & \mathcal{C}_k(B) \\ \mathcal{C}_k(C) & \mathcal{C}_k(D) \end{bmatrix} \ .$$

The same argument clearly applies to the larger block-matrix algebras $\mathbf{M}_m(\mathfrak{A})$, $m \geq 2$, and we draw the following conclusion:

4.17. Lemma. *For any m matrices $A_1, \dots, A_m \in \mathbf{M}_n$, and any $*$-subalgebra $\mathfrak{A}$ of $\mathbf{M}_n$, there exists a sequence $\{\mathcal{C}_k\}_{k\in\mathbb{N}}$ of operators of the form (4.23) such that ,*

$$\mathbf{E}_{\mathfrak{A}}(A_j) = \lim_{k\to\infty} \mathcal{C}_k(A_j) \qquad \text{for each} \quad j = 1, \dots, m \ .$$

The block matrix construction that has led to the proof of Lemma 4.17 provides a powerful perspective on a great many problems, and it will turn out to be important for far more than the proof of this lemma. In the meantime however, let us turn to some concrete examples of conditional expectations.

4.4. Pinching, conditional expectations, and the Operator Jensen Inequality.

4.18. Example (Pinching). Let $A \in \mathbf{H}_n$ have the spectral representation $A = \sum_{j=1}^k \lambda_j P_j$, with $\sum_{j=1}^k P_j = I$. (That is, we include the zero eigenvalue in the sum if zero is an eigenvalue.) Let $\mathfrak{A}_A$ denote the commutant of A, which, as we have observed is a $*$-subalgebra of $\mathbf{M}_n$. For simplicity of notation, let E_A denote the conditional expectation given $\mathfrak{A}_A$; i.e.,

$$\mathrm{E}_A := \mathrm{E}_{\mathfrak{A}_A} \ .$$

We now claim that for any $B \in \mathbf{M}_n$,

$$\mathrm{E}_A(B) = \sum_{j=1}^k P_j B P_j \ . \tag{4.25}$$

To prove this, note that $P_j A = AP_j = \lambda_j P_j$, so that

$$\left(\sum_{j=1}^k P_j B P_j\right) A = \sum_{j=1}^k \lambda_j P_j B P_j = A\left(\sum_{j=1}^k P_j B P_j\right) \ ,$$

and thus the right hand side of (4.25) belongs to $\mathfrak{A}_A$.

Next, ss we have seen in (4.8), each of the spectral projections P_j can be written as a polynomial in A, and hence belongs to $\mathfrak{A}_A$. Furthermore, for all $C \in \mathfrak{A}_A$, and each $j = 1\dots,k$, $CP_j = P_jC$.

Therefore, for such C,

$$\left(\sum_{j=1}^k P_j B P_j\right) C = \sum_{j=1}^k P_j BC P_j ,$$

so that

$$\mathrm{Tr}\left[\left(B - \left(\sum_{j=1}^k P_j B P_j\right)\right) C\right] = \mathrm{Tr}\left[BC - \left(\sum_{j=1}^k P_j BC P_j\right)\right] = 0$$

since $\sum_{j=1}^k P_j = I$. This shows that the right hand side of (4.25) is in fact the orthogonal projection of B onto $\mathfrak{A}_A$, and proves (4.25).

Davis [**8**] refers to the operation $B \mapsto \sum_{j=1}^k P_j B P_j$ for a set of orthogonal projections $P_1, \dots, P_k$ satisfying $\sum_{j=1}^k P_j = I$ as a *pinching operation.* The calculation we have just made shows that *pinching* is a conditional expectation: Indeed, given the orthogonal projections $P_1, \dots, P_k$ satisfying $\sum_{j=1}^k P_j = I$, define the self-adjoint operator A by $A = \sum_{j=1}^k jP_j = I$. With this definition of A, (4.25) is true using A on the left and $P_1, \dots, P_k$ on the right.

It now follows from Theorem 4.16 that for any operator convex function f, and any set of orthogonal projections $P_1, \dots, P_k$ satisfying $\sum_{j=1}^k P_j = I$,

$$f\left(\sum_{j=1}^k P_j B P_j\right) \le f(B) \tag{4.26}$$

for all B in $\mathbf{H}_n$.

The fact that "pinching" is actually a conditional expectation has several useful consequences, with which we close this section. □

4.19. THEOREM (Sherman-Davis Inequality). *For all operator convex functions f, and all orthogonal projectios $P \in \mathbf{M}_n$,*

$$Pf(PAP)P \leq Pf(A)P \tag{4.27}$$

for all $A \in \mathbf{H}_n$.

Proof: We take $P = P_1$ and let $P_2 = I - P_1$. Then using (4.26) for $P_1 + P_2 = I$ and

$$P_1 f\left(\sum_{j=1}^{2} P_j B P_j\right) P_1 = P_1 f(P_1 B P_1) P,$$

we obtain (4.27). ■

Theorem 4.19 is due to Davis and Sherman [**7, 8**]. Note that if $f(0) = 0$, (4.27) may be shortened to

$$f(PAP) \leq Pf(A)P \ ,$$

but one case that comes up in applications is $f(s) = s^{-1}$ where (4.27) must be used as is.

The next inequality is a variant of Theorem 4.19 due to Davis [**8**].

4.20. THEOREM (The Operator Jensen Inequality). *Let $V_1, \ldots, V_k \in \mathbf{M}_n$ satisfy*

$$\sum_{j=1}^{k} V_j^* V_j = I \tag{4.28}$$

Then for any operator convex function f, and any $B_1, \ldots, B_k \in \mathbf{H}_n^+$,

$$f\left(\sum_{j=1}^{k} V_j^* B_j V_j\right) \leq \sum_{j=1}^{k} V_j^* f(B_j) V_j \ . \tag{4.29}$$

Proof: Let $\mathcal{U}$ be any $kn \times kn$ unitary matrix which, when viewed as a $k \times k$ block matrix with $n \times n$ blocks $\mathcal{U}_{i,j}$ has

$$\mathcal{U}_{i,n} = V_i \qquad i = 1, \ldots, n \ . \tag{4.30}$$

Since (4.28) is satisfied, there are many ways to construct such a matrix $\mathcal{U}$: (4.30) specifies the final n columns, which are unit vectors in $\mathbb{C}^{kn}$ by (4.28), and then the remaining columns can be filled in by extending these n unit vectors to an orthonormal basis of $\mathbb{C}^{kn}$.

Next, let $\mathcal{B}$ be the $kn \times kn$ matrix, again viewed as an $k \times k$ block matrix, which has B_j for its jth diagonal block, and zero for all off-diagonal blocks. Finally, let $\mathcal{P}$ by the $kn \times kn$ orthogonal projection with $I_{n \times n}$ in the upper left block, and zeros elsewhere.

Note that $f(\mathcal{B})$ has $f(B_j)$ as its jth diagonal block, and zeros elsewhere. A simple calculation now shows that $\mathcal{U}^*\mathcal{B}\mathcal{U}$ has $\sum_{j=1}^k V_j^* B_j V_j$ as its upper left $n \times n$ block, and

$$f\left(\mathcal{U}^*\mathcal{B}\mathcal{U}\right) = \mathcal{U}^* f(\mathcal{B})\mathcal{U}$$

has $\sum_{j=1}^k V_j^* f(B_j) V_j$ as its upper left $n \times n$ block.

By (4.27),

$$\mathcal{P} f\left(\mathcal{P}\mathcal{U}^*\mathcal{B}\mathcal{U}\mathcal{P}\right)\mathcal{P} \leq \mathcal{P} f\left(\mathcal{U}^*\mathcal{B}\mathcal{U}\right)\mathcal{P} \ ,$$

which, by the calculation we have just made, and by the definition of $\mathcal{P}$, is equivalent to (4.29). ■

4.21. Remark. *It is clear, upon taking each V_j to be a positive multiple of the identity, that the operator convexity of f is not only a sufficient condition for (4.29) to hold whenever $V_1, \ldots, V_k \in \mathbf{M}_n$ satisfy (4.28); it is also necessary. It is remarkable that the class of functions f with the operator convexity property of Theorem 4.20, in which the convex combination is taken using operator valued weights, is not a proper subclass of the class of operator convex functions we have already defined using scalar valued weights.*

5. Tensor products

5.1. Basic definitions and elementary properties of tensor products. If V and W are two finite dimensional vector spaces, their tensor product is the space of all bilinear forms on $V^* \times W^*$, where V^* and W^* are the dual spaces of V and W respectively. That is, V^* consists of the linear functionals f on V, with the usual vector space structure ascribed to spaces of functions, and similarly for W.

Of course, in any inner product space, we have an identification of V^* with V provided by the inner product. However, certain formulas for the tensor product that we shall make much use of will be most clear if we introduce the tensor product in its purest form as a vector space construct, without reference to any inner product.

The next few paragraphs recall some elementary facts about dual spaces and matrix representations. Many readers may prefer to skip ahead to the formal definition of the tensor product, but we include the material to ensure that our notation is absolutely unambiguous.

If $\{v_1, \ldots, v_m\}$ is any basis of V, let $\{f_1, \ldots, f_m\}$ denote the corresponding *dual basis* of V^*. That is, for any $v \in V$, write

$$v = \sum_{j=1}^m a_j v_j \quad , \quad a_1, \ldots, a_m \in \mathbb{C} \ .$$

Since the coefficients $a_1, \ldots, a_j$ are uniquely determined by v, the map

$$f_j : v \mapsto a_j$$

is well defined and is clearly a linear transformation from V to $\mathbb{C}$; i.e., an element of V^*. It is easy to see that $\{f_1, \ldots, f_m\}$ spans V^*, and also that

$$f_i(v_j) = \delta_{i,j} \quad , \quad 1 \leq i, j \leq m$$

from which linear independence easily follows. Thus, $\{f_1, \dots, f_m\}$ is a basis of V^*, and is, by definition, the basis dual to $\{v_1, \dots, v_m\}$.

The *coordinate maps*

$$v \mapsto (f_1(v), \dots, f_m(v)) \in \mathbb{C}^m$$

and

$$f \mapsto (f(v_1), \dots, f(v_m)) \in \mathbb{C}^m$$

are the isomorphisms of V and V^* respectively with $\mathbb{C}^m$ that are induced by the dual bases $\{v_1, \dots, v_m\}$ and $\{f_1, \dots, f_m\}$, and ultimately by the basis $\{v_1, \dots, v_m\}$, since this determines its dual basis. In particular, for any $v \in V$ and any $f \in V^*$,

$$v = \sum_{j=1}^{m} f_j(v) v_j \qquad \text{and} \qquad f = \sum_{j=1}^{m} f(v_j) f_j \ . \tag{5.1}$$

The dual basis is useful for many purposes. One is writing down matrix representations of linear transformations. If $T : V \to V$ is any linear transformation of V, let $[T] \in \mathbf{M}_m$ be defined by

$$[T]_{i,j} = f_i(T(v_j)) \ ,$$

For any fixed basis, the matrix $[T]$ gives the action of T on coordinate vectors for that basis: If a vector $v \in V$ has jth coordinate a_j; i.e., $a_j = f_j(v)$, $j = 1, \dots, m$, then Tv has ith coordinate $\sum_{j=1}^{m} [T]_{i,j} a_j$.

5.1. Definition (Tensor product of two finite dimensional vector spaces). For two finite dimensional vector spaces V and W, their *tensor product space* $V \otimes W$ is the vector space consisting of all bilinear forms K on $V^* \times W^*$, equipped with the usual vector space structure ascribed to spaces of functions.

Given $v \in V$ and $w \in W$, $v \otimes w$ denote the bilinear from on $V^* \times W^*$ given by

$$v \otimes w(f, g) = f(v) g(w) \qquad \text{for all} \quad f \in V^* \ , g \in W^* \ . \tag{5.2}$$

If $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ are bases of V and W respectively, let $\{f_1, \dots, f_m\}$ and $\{g_1, \dots, g_n\}$ denote the corresponding dual bases. By (5.1) and the definition of $v_i \otimes w_j$, for any bilinear form K on $V^* \times W^*$,

$$\begin{aligned} K(f, g) &= \sum_{i,j} K(f(v_i) f_i, g(w_j) g_j) = \sum_{i,j} K(f_i, g_j) f(v_i) g(w_j) \\ &= \sum_{i,j} K(f_i, g_j) [v_i \otimes w_j](f, g) \ . \end{aligned}$$

That is,

$$K = \sum_{i,j} K(f_i, g_j) [v_i \otimes w_j] \ . \tag{5.3}$$

Thus,

$$\{v_i \otimes w_j \ : \ 1 \le i \le m \ , \ 1 \le j \le n \ \} \tag{5.4}$$

spans $V \otimes W$. It is also linearly independent, and is therefore a basis of $V \otimes W$.

To see this linear independence, suppose that for some numbers $b_{i,j}$, $1 \le i \le m$, $1 \le j \le n$, $\sum_{i,j} b_{i,j} v_i \otimes w_j = 0$; i.e., $\sum_{i,j} b_{i,j} v_i \otimes w_j$ is the bilinear map on

$V \times W$ sending everything to zero. But then applying $\sum_{i,j} b_{i,j} v_i \otimes w_j$ to (f_k, g_ℓ) we see

$$0 = \left(\sum_{i,j} b_{i,j} v_i \otimes w_j\right)(f_k, g_\ell) = \sum_{i,j} b_{i,j} f_k(v_i) g_\ell(w_j) = b_{k,\ell} \ ,$$

which shows the linear independence. We are now in a position to define a key isomorphism:

5.2. Definition (Matrix isomorphism). Given any two bases $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ of V and W respectively, and hence the corresponding dual bases $\{f_1, \dots, f_m\}$ and $\{g_1, \dots, g_n\}$ of V^* and W^* respectively, the *matrix isomorphism* is the identification of $V \otimes W$ with the space $\mathbf{M}_{m\times n}$ of $m \times n$ matrices given by

$$V \otimes W \ni K \mapsto [K] \in \mathbf{M}_{m\times n} \tag{5.5}$$

where

$$[K]_{i,j} = K(f_i, g_j) \ . \tag{5.6}$$

The fact that (5.5) is an isomorphism follows directly from (5.3) and the fact that (5.4) is a basis of $V \otimes W$. Of course, this isomorphism depends on the choice of bases, but that shall not diminish its utility.

5.3. Example. For any $v \in V$ and $w \in W$, what is the matrix corresponding to $v \otimes w$? (Of course we assume that the bases $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ of V and W, and their corresponding dual bases are specified.) Since $v \otimes w(f_i, g_j) = f_i(v) g_j(w)$, we have

$$[v \otimes w]_{i,j} = [v \otimes w](f_i, g_j) = f_i(v) g_j(w) = [v]_i [w]_j \tag{5.7}$$

where $[v]_i := f_i(v)$ is the ith coordinate of v, while $[w]_j := g_j(w)$ is the jth coordinate of w. In other words, the matrix corresponding to $v \otimes w$, for this choice of bases, is the rank one matrix with entries $[v]_i [w]_j$.

Every rank one matrix arises this way, and thus the matrix isomorphism identifies the set of *product vectors* in $V \otimes W$ with the set of rank one matrices in $\mathbf{M}_{m\times n}$. Since we know how to compute the rank of matrices by row reduction, this gives us a means to determine whether or not any given element of $V \otimes W$ is a product vector on not. □

5.4. Definition (Entanglement). The *Schmidt rank* of a vector K in $V \otimes W$ is the rank of the corresponding matrix $[K] \in \mathbf{M}_{m\times n}$. (Note that this rank is independent of the choice of bases used to determine $[K]$.) If the Schmidt rank of K is greater than one, then K is an *entangled* vector, and otherwise, if the Schmidt rank of K equals one, K is a *product* vector, in which case one may say K is *unentangled*. As noted above, the matrix isomorphism provides an effective means to determine whether a given $K \in V \otimes W$ is entangled or not.

Now let $T : V \to V$ and $S : W \to W$ be linear transformations. This pair, (T, S), induces a linear transformation $T \otimes S : V \otimes W \to V \otimes W$ by

$$[T \otimes S(K)](f, g) = K(f \circ S, g \circ T) \ ,$$

where of course $f \circ T \in V^*$ is given by $f \circ T(v) = f(T(v))$, and similarly for $g \circ S$. By (5.1),

$$f_i \circ T = \sum_{k=1}^{m} ((f_i \circ T)(v_k) f_k = \sum_{k=1}^{m} f_i(T(v_k)) f_k = \sum_{k=1}^{m} [T]_{i,k} f_k \ ,$$

and likewise

$$g_j \circ S = \sum_{\ell=1}^{n} [S]_{j,\ell} g_\ell \ .$$

Therefore,

$$(T \otimes S)K(f_i, g_j) = \sum_{k,\ell} [T]_{i,k}[S]_{j,\ell} K(f_k, g_\ell) \ , \tag{5.8}$$

which means that

$$[(T \otimes S)K]_{i,j} = \sum_{k,\ell} [T]_{i,k}[S]_{j,\ell}[K]_{k,\ell} \ . \tag{5.9}$$

In other words, under the isometry $K \mapsto [K]$ of $V \otimes W$ with the space of $m \times n$ matrices, the action of $T \otimes S$ on $V \otimes W$ has a very simple matrix expression: The matrix $[T]$ of T acts on the left index of $[K]$, and the matrix $[S]$ of S acts on the right index of K.

5.2. Tensor products and inner products. Now suppose that V and W are inner product spaces; i.e., finite dimensional Hilbert spaces. We denote the inner product on either space by $\langle \cdot, \cdot \rangle$. At this stage we may as well identify V with $\mathbb{C}^m$ and W with $\mathbb{C}^n$, so let us suppose that $V = \mathbb{C}^m$ and $W = \mathbb{C}^n$, both equipped with their standard Euclidean inner products.

Now let $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ be *orthonormal* bases for V and W respectively. We can now express the dual basis elements in terms of the inner product: For any $v \in V$ and $w \in W$, $f_i(v) = \langle v_i, v \rangle$, $i = 1, \dots, m$ and $g_j(w) = \langle w_j, w \rangle$, $j = 1, \dots, n$. In particular, from (5.7) we have that

$$[v \otimes w]_{i,j} = \langle v_i, v \rangle \langle w_j, w \rangle \qquad 1 \le i \le m \ , \ 1 \le j \le n \ . \tag{5.10}$$

As above, for $K \in V \otimes W$, let $[K]$ denote the $m \times n$ matrix corresponding to K under the matrix isomorphism that is induced by our choice of orthonormal bases. Now use the Hilbert-Schmidt inner product on the space of $m \times n$ matrices to induce an inner product on $V \otimes W$. Define, for $B, C \in V \otimes W$,

$$\langle B, C \rangle = \mathrm{Tr}\,([B]^*[C]) = \sum_{i,j} \overline{[B]_{i,j}} [C]_{j,i} \ . \tag{5.11}$$

Combining (5.10) and (5.11), we see that for any $v, v' \in V$ and any $w, w' \in W$,

$$\begin{aligned} \langle v \otimes w, v' \otimes w' \rangle &= \sum_{i,j} \overline{\langle v_i, v \rangle \langle w_i, w \rangle} \langle v_i, v' \rangle \langle w_i, v' \rangle \\ &= \left(\sum_{i=1}^{m} \langle v, v_i \rangle \langle v_i, v' \rangle \right) \left(\sum_{i=1}^{m} \langle w, w_j \rangle \langle w_j, w' \rangle \right) \\ &= \langle v, v' \rangle \langle w, w' \rangle \ . \end{aligned} \tag{5.12}$$

Notice that the right hand side does not depend on our choices of orthonormal bases. Thus, while our inner product on $V \otimes W$ defined by (5.11) might at first sight seem to depend on the choice of the orthonormal bases used to identify $V \otimes W$ with the space of $m \times n$ matrices, we see that this is not the case.

There is one more important conclusion to be drawn from (5.12): For any orthonormal bases $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ of V and W respectively,

$$\{v_i \otimes w_j \ : \ 1 \le i \le m \ , \ 1 \le j \le n \ \}$$

is an orthonormal basis of $V \otimes W$.

When V and W are inner product spaces, we can quantify the degree of entanglement of vectors in $V \otimes W$ in a meaningful way, independent of the choice of bases. The Schmidt rank gives one such quantification, but as rank is not a continuous function on $\mathbf{M}_{m\times n}$, it has limited use beyond its fundamental role in defining entanglement.

Recall that any $K \in \mathbf{M}_{m\times n}$ has a *singular value decomposition*

$$K = U\Sigma V^* \tag{5.13}$$

where Σ is an $r \times r$ diagonal matrix with *strictly* positive entries $\sigma_1 \ge \cdots \ge \sigma_r$ known as the singular values of K, and where U and V are isometries from $\mathbb{C}^r$ into $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively. That is $U = [u_1, \dots, u_r]$ where each u_j is a unit vector in $\mathbb{C}^m$, and $V = [v_1, \dots, v_r]$ where each v_j is a unit vector in $\mathbb{C}^n$. In other words,

$$U^*U = V^*V = I_{r\times r} \ . \tag{5.14}$$

Evidently r is the rank of K.

While the matrices U and V are "essentially uniquely" determined by K, what is important to us here is that the matrix Σ is absolutely uniquely determined by K: It makes sense to speak of *the* singular values of K.

By the definition of the inner product on $V \otimes W$, if K is any unit vector in $V \otimes W$, then

$$\mathrm{Tr}[K^*K] = \mathrm{Tr}[KK^*] = 1 \ , \tag{5.15}$$

and so both K^*K and KK^* are density matrices, on $\mathbb{C}^n$ and $\mathbb{C}^m$ respectively. Notice that by (5.14)

$$K^*K = V\Sigma^2V^* \qquad \text{and} \qquad KK^* = U\Sigma^2U^*$$

so the squares of the singular values of K are the non-zero eigenvalues of these two density matrices, and in particular $\sum_{j=1}^r \sigma_j^2 = 1$. Computing the von Neumann entropies of these two density matrices, we find

$$S(K^*K) = S(KK^*) = -\sum_{j=1}^r \sigma_j^2 \log(\sigma_j^2) \ .$$

Thus, we come to the conclusion that K is a product state if and only if $S(K^*K) = 0$, and otherwise, if $S(K^*K) > 0$, K is entangled. Since $S(K^*K)$ depends continuously on K, this provides us with a useful measure of the degree of entanglement.

5.3. Tensor products of matrices. In this subsection, we focus on the tensor product of $\mathbb{C}^m$ and $\mathbb{C}^n$, each equipped with their usual Euclidean inner products.

Given matrices $A \in \mathbf{M}_m$ and $B \in \mathbf{M}_n$, identify these matrices with the linear transformations that they induce on $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively. It then follows from (5.7) and (5.9) that the map

$$v \otimes w \mapsto Av \otimes Bw \ , \tag{5.16}$$

extends to a linear transformation on $C^m \otimes \mathbb{C}^n$. This linear transformation is denoted by $A \otimes B$. Expressing it more generally and concretely, it follows from (5.9) that for any $K \in \mathbf{M}_{m\times n}$ regarded as a vector in $\mathbb{C}^m \otimes \mathbb{C}^n$,

$$[(A \otimes B)K]_{i,j} = \sum_{k,\ell} A_{i,k}B_{j,\ell}K_{k,\ell} \ . \tag{5.17}$$

Thus, for all $A, C \in \mathbf{M}_m$ and $B, D \in \mathbf{M}_n$,

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD) \ . \tag{5.18}$$

In particular, if A and B are invertible, then so is $A \otimes B$, and

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \ .$$

It follows from (5.12) that for all $A \in \mathbf{M}_m$, $B \in \mathbf{M}_n$, $v_1, v_2 \in \mathbb{C}^m$ and $w_1, w_2 \in \mathbb{C}^n$,

$$\begin{aligned}\langle v_1 \otimes w_1, (A \otimes B)v_2 \otimes w_2\rangle &= \langle v_1, Av_2\rangle\langle w_1, Bw_2\rangle = \langle A^*v_1, v_2\rangle\langle B^*w_1, w_2\rangle \\ &= \langle (A^* \otimes B^*)v_1 \otimes w_1, v_2 \otimes w_2\rangle \ .\end{aligned}$$

That is,

$$(A \otimes B)^* = A^* \otimes B^* \ . \tag{5.19}$$

Consequently, suppose $A \in \mathbf{H}_m^+$ and $B \in \mathbf{H}_n^+$. Then we can write $A = C^*C$ and $B = D^*D$ for $C \in \mathbf{M}_m$ and $D \in \mathbf{M}_n$. Then $A \otimes B = (C \otimes D)^*(C \otimes D)$, so $(A \otimes B)$ is at least positive semi-definite. Since A and B are both invertible, so is $A \otimes B$, and hence $(A \otimes B)$ is positive definite. That is, whenever $A \in \mathbf{H}_m^+$ and $B \in \mathbf{H}_n^+$, then $A \otimes B$ is positive definite.

The equation (5.17) provides one useful way to represent the action of the operator $A \otimes B$, but there is another that is also often useful: a representation of the operator $A \otimes B$ in terms of block matrices. If $K = [v_1, \ldots, v_n]$ is the $m \times n$ matrix whose jth column is $v_j \in \mathbb{C}^m$, let us "vertically stack" K as a vector in $\mathbb{C}^{mn}$:

$$K_{\text{vec}} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} . \tag{5.20}$$

Then $A \otimes B$ is represented by the block matrix

$$\begin{bmatrix} B_{1,1}A & \cdots & B_{1,n}A \\ \vdots & \ddots & \vdots \\ B_{n,1}A & \cdots & B_{n,n}A \end{bmatrix} . \tag{5.21}$$

5.4. The partial trace. Let D be any operator on $\mathbb{C}^m \otimes \mathbb{C}^n$, regarded as an mn dimensional inner product space, as described above. Let Tr denote the trace on $\mathbb{C}^m \otimes \mathbb{C}^n$, and consider the linear functional

$$A \mapsto \mathrm{Tr}[DA]$$

defined on $\mathfrak{A} := \mathcal{L}(\mathbb{C}^m \otimes \mathbb{C}^n)$, the operators on $\mathbb{C}^m \otimes \mathbb{C}^n$.

Let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be the subalgebras of $\mathfrak{A}$ consisting of operators of the form $B \otimes I_{n\times n}$ and $I_{m\times m} \otimes C$ respectively. (In what follows, we shall usually simply write I in place of $I_{m\times m}$ or $I_{n\times n}$ where the meaning is clear.)

The maps

$$B \mapsto \mathrm{Tr}[D(B \otimes I)] \qquad \text{and} \qquad C \mapsto \mathrm{Tr}[D(I \otimes C)]$$

are then linear functionals on $\mathbf{M}_m$ and $\mathbf{M}_n$ respectively.

Since $\mathbf{M}_n$ is an inner product space with the inner product $\langle X, Y\rangle = \mathrm{Tr}[X^*Y]$, for every linear functional φ on $\mathbf{M}_n$, there is a unique $X_\varphi \in \mathbf{M}_n$ such that

$$\varphi(Y) = \langle X_\varphi, Y\rangle = \mathrm{Tr}[X_\varphi Y] \quad \text{for all} \quad Y \in \mathbf{M}_n \ .$$

This justifies the following definition:

5.5. Definition. For any operator D on $\mathbb{C}^m \otimes \mathbb{C}^n$, $\mathrm{Tr}_1[D]$ is the unique element of $\mathbf{M}_n$ such that

$$\mathrm{Tr}[D(I \otimes C)] = \langle (I \otimes C^*), D\rangle = \langle C^*, \mathrm{Tr}_1[D]\rangle = \mathrm{Tr}\left[\mathrm{Tr}_1[D]C\right] \ , \tag{5.22}$$

where the trace on the left hand side of (5.22) is taken on $\mathbb{C}^m \otimes \mathbb{C}^n$, and the trace on the right is taken on $\mathbb{C}^n$. We refer to $\mathrm{Tr}_1[D]$ as the *partial trace* of D onto $\mathbf{M}_n$. In the same way, we define $\mathrm{Tr}_2[D]$ so that

$$\mathrm{Tr}[(\mathrm{Tr}_2 D)B] = \mathrm{Tr}[D(B \otimes I)] \tag{5.23}$$

for all $B \in \mathbf{M}_m$.

If we represent $K \in \mathbb{C}^m \otimes \mathbb{C}^n$ as a vector in $\mathbb{C}^{mn}$ as in (5.20) then D can be represented as a block matrix with n^2 blocks

$$\begin{bmatrix} D_{(1,1)} & \cdots & D_{(1,n)} \\ \vdots & \ddots & \vdots \\ D_{(n,1)} & \cdots & D_{(n,n)} \end{bmatrix} , \tag{5.24}$$

where each $D_{(i,j)} \in \mathbf{M}_m$. Then by (5.20),

$$D(B \otimes I) = \begin{bmatrix} D_{(1,1)}B & \cdots & D_{(1,n)}B \\ \vdots & \ddots & \vdots \\ D_{(n,1)}B & \cdots & D_{(n,n)}B \end{bmatrix} , \tag{5.25}$$

and therefore

$$\mathrm{Tr}[D(B \otimes I)] = \sum_{j=1}^{n} \mathrm{Tr}[D_{(j,j)}B] \ ,$$

where the trace on the left is taken in $\mathbb{C}^m \otimes \mathbb{C}^n$, and on the right in $\mathbb{C}^m$. Thus we see

$$\mathrm{Tr}_2[D] = \sum_{j=1}^{n} D_{(j,j)} \ .$$

That is, if D is written in block matrix form, then the partial trace is simply the sum of the diagonal blocks.

The partial trace has an important physical interpretation. In quantum mechanics, the density matrices ρ on $\mathbb{C}^m \otimes \mathbb{C}^n$ represent the possible *states* of a system whose observables are operators A on $\mathbb{C}^m \otimes \mathbb{C}^n$. Then the value $\mathrm{Tr}[\rho A]$ represents the expected value of a measurement of the observable A, at least in the case that S is self-adjoint, in which case a well-defined measurement procedure is supposed to exist. Let $\mathfrak{A}$ denote the algebra of observables on the whole system, i.e., $\mathfrak{A}$ denotes the linear transformations from $\mathbb{C}^m \otimes \mathbb{C}^n$ into itself.

The tensor product structure of our (finite dimensional) Hilbert space $\mathbb{C}^m \otimes \mathbb{C}^n$ arises whenever our quantum mechanical system is composed of two subsystems: The first may consist of some degrees of freedom that we are trying to measure; i.e., that are coupled to some experimental apparatus, and the second may be the "environment", a heat bath of some sort, or just some other degrees of freedom that are not *directly* coupled to our measurement apparatus.

In this circumstance, the subalgebra $\mathfrak{A}_1$ of observables of the form $B \otimes I$; i.e., observables on the first subsystem is of obvious interest. And clearly, it is of obvious interest to *restrict* the linear functional

$$A \mapsto \mathrm{Tr}[\rho A] \ ,$$

which gives expected values, to the subalgebra $\mathfrak{A}_1$ of observables that our apparatus might measure. This restriction is evidently given by

$$(B \otimes I) \mapsto \mathrm{Tr}[\rho(B \otimes I)] \ .$$

The partial trace allows us to express this restriction in terms of a density matrix on the subsystem. By the definition of the partial trace,

$$\mathrm{Tr}[\rho(B \otimes I)] = \mathrm{Tr}\left[\mathrm{Tr}_2[\rho] B\right] \ .$$

The fact that $\mathrm{Tr}_2[\rho]$ is a density matrix on $\mathbb{C}^m$ whenever ρ is a density matrix on $\mathbb{C}^m \otimes \mathbb{C}^n$ is clear from the fact that

$$B \geq 0 \Rightarrow B \otimes I \geq 0 \to \mathrm{Tr}[\rho(B \otimes I)] \geq 0 \Rightarrow \mathrm{Tr}\left[\mathrm{Tr}_2[\rho] B\right] \geq 0 \ ,$$

so that $\mathrm{Tr}_2[\rho] \geq 0$, and taking $B = I$, we see that $\mathrm{Tr}[\mathrm{Tr}_2[\rho]] = 1$. In summary:

5.6. Theorem (The partial traces preserve positivity and traces). *For all operators D on $\mathbb{C}^m \otimes \mathbb{C}^n$, the map $D \mapsto \mathrm{Tr}_j(D)$, $j = 1, 2$ satisfies*

$$\mathrm{Tr}[\mathrm{Tr}_j(D)] = \mathrm{Tr}[D] \ , \tag{5.26}$$

and

$$D \geq 0 \quad \Rightarrow \quad \mathrm{Tr}_j(D) \geq 0 \ . \tag{5.27}$$

That is, $D \mapsto \mathrm{Tr}_j(D)$, $j = 1, 2$, is trace preserving *and* positivity preserving.

We now make an observation that may already have occurred to the reader: The partial trace is nothing but a special case of the conditional expectation: Using the notation introduced above, consider $\mathrm{E}_{\mathfrak{A}_2}$, the conditional expectation that is the orthogonal projection onto the $*$-subalgebra $\mathfrak{A}_2$ consisting of operators in $\mathfrak{A}$ of the form $I \otimes C$, $C \in \mathbf{M}_n$. Then, by the definition of the conditional expectation as an orthogonal projection, for any $D \in \mathfrak{A}$, and any $C \in \mathbf{M}_n$,

$$\langle I \otimes C^*, D\rangle = \langle I \otimes C^*, \mathrm{E}_{\mathfrak{A}_2}(D)\rangle \ . \tag{5.28}$$

By definition, $\mathrm{E}_{\mathfrak{A}_2}(D)$ has the form $I \otimes \widetilde{D}$ for some $\widetilde{D} \in \mathbf{M}_n$. Thus, we can rewrite (5.28) as

$$\mathrm{Tr}_{\mathbb{C}^m\otimes\mathbb{C}^n}[(I \otimes C)D] \ = \ \mathrm{Tr}_{\mathbb{C}^m\otimes\mathbb{C}^n}[(I \otimes C)(I \otimes \widetilde{D})] = m\mathrm{Tr}_{\mathbb{C}^n}[C\widetilde{D}] \ ,$$

where the subscripts indicate the different Hilbert spaces over which the traces are taken. Comparing this with (5.22), we see that $\widetilde{D} = \mathrm{Tr}_1[D]$. That is,

$$\frac{1}{m} I \otimes \mathrm{Tr}_1[D] \ = \ \mathrm{E}_{\mathfrak{A}_2}(D) \ .$$

This result, combined with Theorem 4.13 allows us to express partial traces as averages over unitary conjugations, and this will be useful in many applications of convexity. Therefore, we summarize in a theorem:

5.7. Theorem. *Let $\mathfrak{A} := \mathcal{L}(\mathbb{C}^m \otimes \mathbb{C}^n)$, the $*$-algebra of linear transformations from $\mathbb{C}^m \otimes \mathbb{C}^n$ into itself. Let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be the $*$-subalgebras of $\mathfrak{A}$ consisting of operators of the form $B \otimes I_{n\times n}$ and $I_{m\times m} \otimes C$ respectively, with $B \in \mathbf{M}_m$ and $C \in \mathbf{M}_n$. Then, for any $D \in \mathfrak{A}$,*

$$\frac{1}{m} I_{m\times m} \otimes \mathrm{Tr}_1[D] = \mathrm{E}_{\mathfrak{A}_2}(D) \qquad \text{and} \qquad \frac{1}{n}\mathrm{Tr}_2[D] \otimes I_{n\times n} = \mathrm{E}_{\mathfrak{A}_1}(D) \ . \tag{5.29}$$

Continuing with the notation of Theorem 5.7, we observe that $\mathfrak{A}_1' = \mathfrak{A}_2$ and $\mathfrak{A}_2' = \mathfrak{A}_1$. In particular, the unitaries in $\mathfrak{A}_2'$ are the unitaries in $\mathfrak{A}_1$, which means they have the form $U \otimes I$, where U is a unitary in $\mathbf{M}_m$.

We also mention at this point that the maps $D \mapsto \mathrm{Tr}_j(D)$, $j = 1, 2$ are not only positivity preserving, but that they also have a stronger property known as *complete positivity*. This is physically very significant, and we shall explain this later. In the meantime, let us make one more definition, and then turn to examples and digest what has been introduced so far.

5.8. Definition. The map $B \mapsto \overline{B}$ on $\mathbf{M}_{m\times n}$ is defined so that each entry of $\overline{B}$ is the complex conjugate of the corresponding entry of B. This map is an antilinear isometry from $\mathbf{M}_{m\times n}$ to itself.

The map $B \mapsto \overline{B}$ preserves positivity: From the spectral representation written in the form $B = \sum_{j=1}^n \lambda_j u_j u_j^*$, one sees that $\overline{B}$ is unitarily equivalent to B under the unitary transformation that takes each u_j to its complex conjugate $\overline{u}_j$. In particular, if $B \in \mathbf{H}_n^+$, then $\overline{B} \in \mathbf{H}_n^+$ as well, and for any $f : \mathbb{R} \to \mathbb{R}$,

$$\overline{f(B)} = f(\overline{B}) \ .$$

5.9. EXAMPLE ($\mathrm{Tr}_1(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|)$ and $\mathrm{Tr}_2(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|)$). For $K \in \mathbf{M}_{m\times n}$ with $\mathrm{Tr}[K^*K] = 1$, considered as a unit vector K_{vec} in $\mathbb{C}^m \otimes \mathbb{C}^n$, let $|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|$ denote the rank one projection onto the span of K_{vec} in $\mathbb{C}^m \otimes \mathbb{C}^n$. In the language and notation of quantum statistical mechanics, this projection is a pure state density matrix on $\mathbb{C}^m \otimes \mathbb{C}^n$. .

By definition, for all $A \in \mathbf{M}_m$,

$$\begin{aligned}
\mathrm{Tr}[\mathrm{Tr}_2(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|)A] &= \mathrm{Tr}[|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|(A \otimes I_{n\times n})] \\
&= \langle K_{\mathrm{vec}}, (A \otimes I_{n\times n})K_{\mathrm{vec}}\rangle \\
&= \sum_{i,j} K^*_{j,i} \sum_k A_{i,k}K_{k,j} \\
&= \mathrm{Tr}[K^*AK] = \mathrm{Tr}[(KK^*)A] \ .
\end{aligned}$$

Thus, we have the useful identity $\mathrm{Tr}_2(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|) = KK^*$. Likewise, for all $B \in \mathbf{M}_n$,

$$\begin{aligned}
\mathrm{Tr}[\mathrm{Tr}_1(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|)B] &= \langle K_{\mathrm{vec}}, (I_{m\times m} \otimes B)K_{\mathrm{vec}}\rangle \\
&= \sum_{i,j} K^*_{j,i} \sum_\ell B_{j,\ell}K_{i,\ell} = \sum_{i,j}\sum_\ell K^*_{j,i}K_{i,\ell}B_{j,\ell} \\
&= \mathrm{Tr}[(\overline{K^*K})B] \ ,
\end{aligned}$$

and hence $\mathrm{Tr}_1(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|) = \overline{K^*K}$. □

This computation that

$$\mathrm{Tr}_1(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|) = \overline{K^*K} \quad \text{and} \quad \mathrm{Tr}_2(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|) = KK^*$$

has a significant consequence:

5.10. THEOREM. *For any pure state ρ on $\mathbb{C}^m \otimes \mathbb{C}^n$, the two restricted density matrices $\mathrm{Tr}_1\rho$ and $\mathrm{Tr}_2\rho$ have the same non-zero spectrum, including multiplicities. In particular, $S(\mathrm{Tr}_1\rho) = S(\mathrm{Tr}_2\rho)$.*

Proof: For any $m \times n$ matrix K, KK^*, K^*K and $\overline{K^*K}$ have the same non-zero spectrum, including multiplicities, and the statement about entropy follows directly from this. ■

5.11. EXAMPLE (Mixed states as partial traces of pure states). Consider any $\rho \in \mathbf{S}_n$. Let $\{u_1, \dots, u_n\}$ be any orthonormal basis for $\mathbb{C}^n$. Let $K_{\mathrm{vec}} \in \mathbb{C}^n \otimes \mathbb{C}^n$ by given by

$$K_{\mathrm{vec}} = \sum_{j=1}^n u_j \otimes \rho^{1/2}u_j \ .$$

We now claim that

$$\mathrm{Tr}_1(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|) = \rho \ .$$

This shows that every mixed state; i.e., density matrix on $\mathbb{C}^n$ is the partial trace of a pure state on $\mathbb{C}^n \otimes \mathbb{C}^n$.

To verify the claim, consider any $B \in \mathbf{M}_n$. Then

$$\begin{aligned}
\mathrm{Tr}[\mathrm{Tr}_1(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|)B] &= \mathrm{Tr}[(|K_{\mathrm{vec}}\rangle\langle K_{\mathrm{vec}}|(I_{m\times m}\otimes B)] \\
&= \langle K_{\mathrm{vec}}, (I_{m\times m}\otimes B)K_{\mathrm{vec}}\rangle \\
&= \left\langle \sum_{i=1}^n u_i\otimes \rho^{1/2}u_i\ ,\ \sum_{j=1}^n u_j\otimes B\rho^{1/2}u_j \right\rangle \\
&= \sum_{i,j=1}^n \langle u_i, u_j\rangle\langle \rho^{1/2}u_i, B\rho^{1/2}u_j\rangle = \mathrm{Tr}[B\rho]\ .
\end{aligned} \tag{5.30}$$

□

5.5. Ando's identity. The next lemma records an important observation of Ando.

5.12. Lemma (Ando's identity). *Let $A \in \mathbf{H}_m^+$, $B \in \mathbf{H}_n^+$ and let K be any $m\times n$ matrix considered as a vector in $\mathbb{C}^m \otimes \mathbb{C}^n$. Then*

$$\langle K, (A\otimes B)K\rangle = \mathrm{Tr}(K^*AK\overline{B})\ . \tag{5.31}$$

Proof: $(A\otimes B)K$, considered as an $m\times n$ matrix, has the entries

$$[(A\otimes B)K]_{i,j} = \sum_{k,\ell} A_{i,k}B_{j,\ell}K_{k,\ell}\ .$$

Since $B \in \mathbf{H}_n$, $B_{j,\ell} = \overline{B}_{\ell,j}$, and so

$$[(A\otimes B)K]_{i,j} = \sum_{k,\ell} A_{i,k}K_{k,\ell}\overline{B}_{\ell,j} = [AK\overline{B}]_{i,j}\ .$$

Then since $\langle K, (A\otimes B)K\rangle = \mathrm{Tr}(K^*[(A\otimes B)K])$, the result is proved. ■

One easy consequence of this identity is the following: For $A \in \mathbf{H}_m^+$, $B \in \mathbf{H}_n^+$, by cyclicity of the trace,

$$\langle K, (A\otimes B)K\rangle = \mathrm{Tr}(\overline{B}^{1/2}K^*AK\overline{B}^{1/2}) = \mathrm{Tr}(A^{1/2}K\overline{B}K^*A^{1/2})\ ,$$

and so the map

$$(A,B) \mapsto (A\otimes B)$$

on $\mathbf{H}_m^+ \times \mathbf{H}_n^+$ is monotone in each argument.

6. Lieb's Concavity Theorem and related results

6.1. Lieb's Concavity Theorem. In this section, we prove the following fundamental theorem of Lieb [**18**]:

6.1. Theorem (Lieb's Concavity Theorem). *For all $m\times n$ matrices K, and all $0 \le q, r \le 1$, with $q + r \le 1$ the real valued map on $\mathbf{H}_m^+ \times \mathbf{H}_n^+$ given by*

$$(A,B) \mapsto \mathrm{Tr}(K^*A^qKB^r) \tag{6.1}$$

is concave.

The following proof is due to Ando [**1**]

Proof of Theorem 6.1: Since the map $B \mapsto \overline{B}$ is linear over $\mathbb{R}$, Theorem 5.12 shows that an equivalent formulation of Lieb's Concavity Theorem is that for $0 \leq q, r \leq 1$,

$$(A, B) \mapsto A^q \otimes B^r \tag{6.2}$$

is concave from $\mathbf{H}_m^+ \times \mathbf{H}_n^+$ to $\mathbf{H}_{mn}^+$

Let Ω be the subset of $(0, \infty) \times (0, \infty)$ consisting of points (q, r) such that $(A, B) \mapsto A^q \otimes B^r$ is concave. Obviously, $(0, 1)$, $(1, 0)$ and $(0, 0)$ all belong to Ω, and hence it suffices to show that Ω is convex. By continuity, it suffices to show that if $(q_1, r_1), (q_2, r_2) \in \Omega$, then so is

$$(q, r) := \left(\frac{q_1 + q_2}{2}, \frac{r_1 + r_2}{2}\right) .$$

The key to this is to use the joint concavity properties of the em operator geometric mean M_0 that we have studied in Section 3 of these notes: Observe that by (5.18), for such (p, q), (p_1, q_1) and (p_2, q_2),

$$A^q \otimes B^r = M_0(A^{q_1} \otimes B^{r_1}, A^{q_2} \otimes B^{r_2}) .$$

Since $(q_1, r_1), (q_2, r_2) \in \Omega$,

$$\left(\frac{A+C}{2}\right)^{q_j} \otimes \left(\frac{B+D}{2}\right)^{r_j} \geq \frac{A^{q_j} \otimes B^{r_j} + C^{q_j} \otimes D^{r_j}}{2} \qquad j = 1, 2 .$$

Then by the monotonicity and concavity of the operator geometric mean,

$$\begin{aligned}
\left(\frac{A+C}{2}\right)^{q} \otimes \left(\frac{B+D}{2}\right)^{r} &= M_0\left(\left(\frac{A+C}{2}\right)^{q_1} \otimes \left(\frac{B+D}{2}\right)^{r_1}, \left(\frac{A+C}{2}\right)^{q_2} \otimes \left(\frac{B+D}{2}\right)^{r_2}\right) \\
&\geq M_0\left(\frac{A^{q_1} \otimes B^{r_1} + C^{q_1} \otimes D^{r_1}}{2}, \frac{A^{q_2} \otimes B^{r_2} + C^{q_2} \otimes D^{r_2}}{2}\right) \\
&\geq \frac{1}{2} M_0(A^{q_1} \otimes B^{r_1}, A^{q_2} \otimes B^{r_2}) + \frac{1}{2} M_0(C^{q_1} \otimes D^{r_1}, C^{q_2} \otimes D^{r_2}) \\
&= \frac{1}{2} A^q \otimes B^r + \frac{1}{2} C^q \otimes D^r .
\end{aligned}$$

This proves the midpoint concavity of $(A, B) \mapsto A^q \otimes B^r$, and now the full concavity follows by continuity. Thus, $(q, r) \in \Omega$, as was to be shown. ■

6.2. Ando's Convexity Theorem. Ando's proof [**1**] of Lieb's Concavity Theorem leads to the following significant complement to it:

6.2. Theorem (Ando's Convexity Theorem). *For all $m \times n$ matrices K, and all $1 \leq q \leq 2$ and $0 \leq r \leq 1$ with $q - r \geq 1$, the real valued map on $\mathbf{H}_m^+ \times \mathbf{H}_n^+$ given by*

$$(A, B) \mapsto \mathrm{Tr}(K^* A^q K B^{-r}) \tag{6.3}$$

is convex.

Proof of Theorem 6.2: First note that

$$A^q \otimes B^{-r} = A \otimes I \frac{1}{A^{2-q} \otimes B^r} A \otimes I \ . \tag{6.4}$$

Next, for $1 \le q \le 2$, $0 \le r \le 1$ and $q - r \ge 1$, we have $0 \le 2 - q \le 1$ and $0 \le (2-q) + r \le 1$. Therefore, by Theorem 6.1, $(A,B) \mapsto A^{2-q} \otimes B^r$ is concave, so that for all $A, C \in \mathbf{H}_m^+$ and all $B, D \in \mathbf{H}_n^+$,

$$\left(\frac{A+C}{2}\right)^{2-q} \otimes \left(\frac{B+D}{2}\right)^r \ge \frac{A^{2-q} \otimes B^r + C^{2-q} \otimes D^r}{2} \ . \tag{6.5}$$

Thus, by the obvious monotonicity of $X \mapsto Y^* X^{-1} Y$,

$$\begin{aligned}\left(\frac{A+C}{2}\right)^q \otimes \left(\frac{A+C}{2}\right)^{-r} &= \left[\left(\frac{A+C}{2}\right) \otimes I\right] \left[\left(\frac{A+C}{2}\right)^{2-q} \otimes \left(\frac{B+D}{2}\right)^r\right]^{-1} \left[\left(\frac{A+C}{2}\right) \otimes I\right] \\ &\le \left[\left(\frac{A+C}{2}\right) \otimes I\right] \left[\frac{A^{2-q} \otimes B^r + C^{2-q} \otimes D^r}{2}\right]^{-1} \left[\left(\frac{A+C}{2}\right) \otimes I\right]\end{aligned}$$

Finally, by Theorem 3.1, which asserts the joint convexity of $(X,Y) \mapsto Y^* X Y$, and then (6.4) once more, we have

$$\left(\frac{A+C}{2}\right)^q \otimes \left(\frac{B+D}{2}\right)^{-r} \le \frac{A^q \otimes B^{-r} + C^q \otimes D^{-r}}{2} \ ,$$

which is the midpoint version of the desired convexity statement. The general case follows by continuity. ■

6.3. Lieb's Concavity Theorem and joint convexity of the relative entropy. Consider the map

$$(A,B) \mapsto \mathrm{Tr}[A \log A] - \mathrm{Tr}[A \log(B)] := H(A|B) \tag{6.6}$$

on $\mathbf{H}_n^+ \times \mathbf{H}_n^+$. In particular, for density matrices ρ and σ, $H(\rho|\sigma) = S(\rho|\sigma)$, the relative entropy of ρ with respect to σ. We shall prove:

6.3. THEOREM. *The map $(A,B) \mapsto \mathrm{Tr}[A \log A] - \mathrm{Tr}[A \log(B)]$ from $\mathbf{H}_n^+ \times \mathbf{H}_n^+$ to $\mathbb{R}$ is jointly convex.*

Proof: For all $0 < p < 1$, $(A,B) \mapsto \mathrm{Tr}(B^{1-p} A^p)$ is jointly concave, by Lieb's Concavity Theorem, and thus

$$(A,B) \mapsto \frac{1}{p-1}\left(\mathrm{Tr}(B^{1-p} A^p) - \mathrm{Tr}(A)\right)$$

is convex. But

$$\lim_{p \to 1} \frac{1}{p-1}\left(\mathrm{Tr}(B^{1-p} A^p) - \mathrm{Tr}(A)\right) = H(A|B) \ ,$$

and convexity is preserved in the limit. ■

6.4. Monotonicity of the relative entropy.

6.4. THEOREM. *Let $\mathfrak{A}$ be any $*$-subalgebra of $\mathbf{M}_n$. Then for any two density matrices $\rho, \sigma \in \mathbf{S}_n$,*

$$S(\rho|\sigma) \geq S(\mathbf{E}_{\mathfrak{A}}(\rho)|\mathbf{E}_{\mathfrak{A}}(\sigma)) \ . \tag{6.7}$$

Proof: We suppose first that ρ and σ are both positive definite, so that $(\widetilde{\rho}, \widetilde{\sigma}) \mapsto S(\widetilde{\rho}|\widetilde{\sigma})$ is continuous in a neighborhood of (ρ, σ). By Lemma 4.17, there is a sequence $\{\mathcal{C}_k\}_{n\in\mathbb{N}}$ of operators of the form (4.23) such that

$$\mathbf{E}_\rho(\sigma) = \lim_{k\to\infty} \mathcal{C}_k(\rho) \qquad \text{and} \qquad \mathbf{E}_\sigma(\sigma) = \lim_{k\to\infty} \mathcal{C}_k(\sigma) \ .$$

Then by the joint convexity of the relative entropy from Theorem 6.3, and the unitary invariance of the relative entropy; i.e., $S(U\rho U^*|U\sigma U^*) = S(\rho|\sigma)$, and the specific form (4.23) of $\mathcal{C}_k$, we have that for each k,

$$S(\mathcal{C}_k(\rho)|\mathcal{C}_k(\sigma)) \leq S(\rho|\sigma) \ .$$

Now taking k to infinity, we obtain the result for positive definite ρ and σ.

To pass to the general case, first note that unless the nullspace of σ is contained in the nullspace of ρ, $S(\rho|\sigma) = \infty$, and there is nothing to prove.

Hence we assume that the nullspace of σ is contained in the nullspace of ρ. This implies that the nullspace of $\mathbf{E}_{\mathfrak{A}}(\sigma)$ is contained in the nullspace of $\mathbf{E}_{\mathfrak{A}}(\rho)$. To see this, let P denote the orthogonal projection onto the nullspace of $\mathbf{E}_{\mathfrak{A}}(\sigma)$. Then $P \in \mathfrak{A}$, and so

$$0 = \mathrm{Tr}(P\mathbf{E}_{\mathfrak{A}}(\sigma)) = \mathrm{Tr}(P\sigma) \ ,$$

and then since the nullspace of σ is contained in the nullspace of ρ,

$$0 = \mathrm{Tr}(P\rho) = \mathrm{Tr}(P\mathbf{E}_{\mathfrak{A}}(\rho)) \ .$$

Now replace ρ and σ by $\rho_\epsilon := (1-\epsilon)\rho + (\epsilon/n)I$ and $\sigma_\epsilon := (1-\epsilon)\sigma + (\epsilon/n)I$ respectively with $1 > \epsilon > 0$. Note that since $I \in \mathfrak{A}$,

$$\mathbf{E}_{\mathfrak{A}}((1-\epsilon)\rho + (\epsilon/n)I) = (1-\epsilon)\mathbf{E}_{\mathfrak{A}}(\rho) + (\epsilon/n)I = (\mathbf{E}_{\mathfrak{A}}(\rho))_\epsilon \ ,$$

and likewise, $\mathbf{E}_{\mathfrak{A}}(\sigma_\epsilon) = (\mathbf{E}_{\mathfrak{A}}(\sigma))_\epsilon$. Therefore

$$\begin{aligned} S(\rho_\epsilon|\sigma_\epsilon) &\geq S(\mathbf{E}_{\mathfrak{A}}(\rho_\epsilon)|\mathbf{E}_{\mathfrak{A}}(\sigma_\epsilon)) \\ &= S((\mathbf{E}_{\mathfrak{A}}(\rho))_\epsilon|(\mathbf{E}_{\mathfrak{A}}(\sigma))_\epsilon) \ . \end{aligned} \tag{6.8}$$

It is easy to show that when the nullspace of σ is contained in the nullspace of ρ, $S(\rho|\sigma) = \lim_{\epsilon\to 0} S(\rho_\epsilon|\sigma_\epsilon)$. By what we have shown above, we then also have $S(\mathbf{E}_{\mathfrak{A}}(\rho)|\mathbf{E}_{\mathfrak{A}}(\sigma)) = \lim_{\epsilon\to 0} S((\mathbf{E}_{\mathfrak{A}}(\rho))_\epsilon|(\mathbf{E}_{\mathfrak{A}}(\sigma))_\epsilon)$. This together with (6.8) proves the result in the general case. ■

6.5. Subadditivity and strong subadditivity of the entropy. Consider a density matrix ρ on the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ of two finite dimensional Hilbert spaces. To be concrete, we may as well suppose that for some m and n, $\mathcal{H}_1 = \mathbb{C}^m$ and $\mathcal{H}_2 = \mathbb{C}^n$. Let $\rho_1 = \mathrm{Tr}_2\rho$ be the density matrix on $\mathcal{H}_1$ obtained by taking the partial trace over $\mathcal{H}_2$ of ρ, and let ρ_2 be defined in the analogous way.

Then $\rho_1 \otimes \rho_2$ is a density matrix on $\mathcal{H}_1 \otimes \mathcal{H}_2$, and by Klein's inequality,

$$S(\rho|\rho_1 \otimes \rho_2) \geq 0$$

with equality if and only if $\rho_1 \otimes \rho_2 = \rho$. Let us assume that ρ is strictly positive, so that ρ_1 and ρ_2 are also strictly positive, and compute the left hand side.

Then since $\log(\rho_1) \otimes I_{\mathcal{H}_2}$ and $I_{\mathcal{H}_1} \otimes \log(\rho_2)$ commute, it is clear that

(6.9)
$$\exp(\log(\rho_1) \otimes I_{\mathcal{H}_2} + I_{\mathcal{H}_1} \otimes \log(\rho_2)) = \exp(\log(\rho_1) \otimes I_{\mathcal{H}_2}) \exp(I_{\mathcal{H}_1} \otimes \log(\rho_2)) = (\rho_1 \otimes I_{\mathcal{H}_2})(I_{\mathcal{H}_1} \otimes \rho_2) = \rho_1 \otimes \rho_2 \ .$$

It follows that

$$\log(\rho_1 \otimes \rho_2) = \log(\rho_1) \otimes I_{\mathcal{H}_2} + I_{\mathcal{H}_1} \otimes \log(\rho_2) \ ,$$

and hence that

$$\begin{aligned} S(\rho|\rho_1 \otimes \rho_2) &= -S(\rho) - \mathrm{Tr}\left[\rho\left(\log(\rho_1) \otimes I_{\mathcal{H}_2} + I_{\mathcal{H}_1} \otimes \log(\rho_2)\right)\right] \\ &= -S(\rho) + S(\rho_1) + S(\rho_2) \end{aligned}$$

where we have used the definition of the partial trace in the second equality. Since the left hand side is non-negative by Klein's inequality, we conclude that $S(\rho) \leq S(\rho_1)+S(\rho_2)$. This inequality is known as the *subadditivity of the quantum entropy.* We summarize our conclusions in the following theorem:

6.5. Theorem (Subadditivity of quantum entropy). *Let ρ be a density matrix on the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ of two finite dimensional Hilbert spaces. For $j = 1, 2$, let ρ_j denote the density matrix on $\mathcal{H}_j$ obtained by taking the partial trace of ρ over the other Hilbert space. Then*

$$S(\rho) \leq S(\rho_1) + S(\rho_2) \ , \tag{6.10}$$

and there is equality if and only if $\rho_1 \otimes \rho_2 = \rho$.

Note that the dimension does not really enter our considerations, and so this inequaity is easily generalized to the infinite dimensional case. In the spirit of these notes, we leave this to the reader.

There is a much deeper subadditivity inequality for density matrices on a tensor product of three Hilbert spaces $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$. Let ρ be a density matrix. By taking the various partial traces of ρ, we obtain various density matrices from ρ. We shall use the following notation for these:

$$\rho_{123} := \rho \quad \rho_{23} := \mathrm{Tr}_1\rho \ , \quad \rho_3 := \mathrm{Tr}_{12}\rho$$

and so forth, where Tr_1 denotes the partial trace over $\mathcal{H}_1$, Tr_{12} denotes the partial trace over $\mathcal{H}_1 \otimes \mathcal{H}_2$ and so forth. (That is, the subscripts indicate the spaces "remaining" after the traces.)

6.6. Theorem (Strong subadditivity of quantum entropy). *Let ρ be a density matrix on the tensor product $\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3$ of three finite dimensional Hilbert spaces. Then, using the notation introduced above*

$$S(\rho_{13}) + S(\rho_{23}) \geq S(\rho_{123}) + S(\rho_3)\ . \tag{6.11}$$

This theorem was conjectured by Lanford, Robinson and Ruelle [**25**], and was proved by Lieb and Ruskai [**20**].

Proof of Theorem 6.6: As in the proof of Theorem 6.5, we compute that

$$S(\rho_{123}|\rho_{12}\otimes\rho_3) = -S(\rho_{123}) + S(\rho_{12}) + S(\rho_3)\ . \tag{6.12}$$

Now let $\mathfrak{A}$ be the $*$-subalgebra of operators on $\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3$ of the form $I_{\mathcal{H}_1}\otimes A$ where A is an operator on $\mathcal{H}_2\otimes\mathcal{H}_3$. Then by the monotonicity of the relative entropy, Thoerem 6.4,

$$S(\rho_{123}|\rho_{12}\otimes\rho_3) \geq S(\mathrm{E}_{\mathfrak{A}}(\rho_{123})|\mathrm{E}_{\mathfrak{A}}(\rho_{12}\otimes\rho_3))\ . \tag{6.13}$$

But by Theorem 6.4,
(6.14)

$$\mathrm{E}_{\mathfrak{A}}(\rho_{123}) = \frac{1}{\dim(\mathcal{H}_1)}I_{\mathcal{H}_1}\otimes\rho_{23} \qquad \text{and} \qquad \mathrm{E}_{\mathfrak{A}}(\rho_{12}\otimes\rho_3) = \frac{1}{\dim(\mathcal{H}_1)}I_{\mathcal{H}_1}\otimes(\rho_2\otimes\rho_3)\ .$$

Therefore, by Theorem 6.5,

$$S(\mathrm{E}_{\mathfrak{A}}(\rho_{123})|\mathrm{E}_{\mathfrak{A}}(\rho_{12}\otimes\rho_3)) = -S(\rho_{23}) + S(\rho_2) + S(\rho_3)\ .$$

Combining this with (6.13) and (6.12) yields (6.11). ■

7. L^p norms for matrices and entropy inequalities

In this section, we shall prove various L^p norm inequalities for matrices that have a connection with quantum entropy. The basic idea is this: Let $\rho \in \mathbf{S}_n$. Then the map

$$p \mapsto \mathrm{Tr}[\rho^p]$$

is differentiable at $p = 1$, and

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathrm{Tr}[\rho^p]\bigg|_{p=1} = \mathrm{Tr}[\rho\log(\rho)] = -S(\rho).$$

The inequalities we obtain in this section will be of interest in their own right, but shall also lead to new proofs of entropy inequalities such as strong subadditivitiy of quantum entropy. We begin with an elementary introduction to the matricial analogs of the L^p norms.

7.1. The matricial analogs of the L^p norms. Let $\mathbf{M}_n$ denote the set of $n\times n$ matrices with complex entries, and let A^* denote the Hermitian conjugate of $A \in \mathbf{M}_n$. For $0 < q < \infty$, and $A \in \mathbf{M}_n$, define

$$\|A\|_q = (\mathrm{Tr}[(A^*A)^{q/2}])^{1/q}\ . \tag{7.1}$$

For $q = \infty$, we define $\|A\|_\infty$ to be the operator norm of A.

For $q \geq 1$, (7.1) defines a norm on $\mathbf{M}_n$, but not for $q < 1$. Nonetheless, it will be convenient here to use this notation for all $q > 0$.

We shall now show that $\|\cdot\|_q$ is in fact a norm for $1\ leqq < \infty$. Let $|A|$ denote $(A^*A)^{1/2}$, and let $\{u_1,\dots,u_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $|A|$ with $|A|u_j = \lambda_j u_j$. Then

$$\|A\|_q = (\mathrm{Tr}[(A^*A)^{q/2}])^{1/q} = (\mathrm{Tr}[|A|^q])^{1/q} = \left(\sum_{j=1}^n \lambda_j^q\right)^{1/q} .$$

The eigenvalues of $|A|$ are the *singular values* of A. Thus, $\|A\|_q$ is the ℓ_q norm of the sequence of singular values of A.

7.1. Theorem (Duality formula for $\|A\|_q$). *For all $q \geq 1$, define p by $1/q + 1/p = 1$. Then for all A in $\mathbf{M}_n$,*

$$\|A\|_q = \sup_{B\in\mathbf{M}_n} \{\mathrm{Tr}[B^*A] \ :\ \|B\|_p = 1\ \} .$$

Proof: For any invertible $A, B \in \mathbf{M}_n$ let $A = U|A|$ and $B = V|B|$ be their polar decompositions, and let $W = V^*U$. Let $\{u_1,\dots,u_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $|B|$ with $|B|u_j = \lambda_j u_j$. Then

$$\mathrm{Tr}(B^*A) = \sum_{j=1}^n \langle u_j, |B|W|A|u_j\rangle = \sum_{j=1}^n \lambda_j\langle u_j, W|A|u_j\rangle \ . \tag{7.2}$$

Now let us suppose that $q > 1$. By Hölder's inequality, for any $q > 1$ and $p = q/(q-1)$,

$$\begin{aligned} \left|\sum_{j=1}^n \lambda_j\langle u_j W|A|u_j\rangle\right| &\leq \left(\sum_{j=1}^n \lambda_j^p\right)^{1/p}\left(\sum_{j=1}^n |\langle u_j, W|A|u_j\rangle|^q\right)^{1/q} \\ &= \|B\|_p\left(\sum_{j=1}^n |\langle u_j, W|A|u_j\rangle|^q\right)^{1/q} . \end{aligned} \tag{7.3}$$

Now define $v_j = W^*u_j$. Then by the Schwarz inequality twice , and then Peierl's inequality,

$$\begin{aligned} \sum_{j=1}^n |\langle u_j, W|A|u_j\rangle| &\leq \sum_{j=1}^n \langle v_j, |A|v_j\rangle^{q/2}\langle u_j, |A|u_j\rangle q/2 \\ &\leq \left(\sum_{j=1}^n \langle v_j, |A|v_j\rangle^q\right)^{1/2}\left(\sum_{j=1}^n \langle u_j, |A|u_j\rangle^q\right)^{1/2} \\ &\leq (\mathrm{Tr}[|A|^q])^{1/2}(\mathrm{Tr}[|A|^q])^{1/2} = \|A\|_q^q . \end{aligned} \tag{7.4}$$

Combining (7.2), (7.3) and (7.4), we have

$$|\mathrm{Tr}(B^*A)| \leq \|B\|_p\|A\|_q \ , \tag{7.5}$$

which is the tracial version of *Hölder's inequality*, and we note that if $B = \|A\|_q^{1-q} U|A|^{q-1}$, then $\|B\|_p = 1$ and

$$\mathrm{Tr}(B^*A) = \|A\|_q^{1-q}\mathrm{Tr}\left[|A|^{q-1}U^*U|A|\right] = \|A\|_q^{1-q}\mathrm{Tr}\left[|A|^q\right] = \|A\|_q \ .$$

Combining this with (7.5) yields the result for $q > 1$. The easy extension to $q = 1$ is left to the reader. ■

Starting from Theorem 7.1, the proof of the Minkowski inequality for $\|\cdot\|_q$ proceeds exactly as it does for the L^p norms: Given $A, C \in \mathbf{M}_n$,

$$\begin{aligned} \|A+C\|_q &= \sup_{B\in\mathbf{M}_n} \{|\mathrm{Tr}[B^*(A+C)]| \ : \ \|B\|_p = 1\ \} \\ &\le \sup_{B\in\mathbf{M}_n} \{|\mathrm{Tr}[B^*A]| \ : \ \|B\|_p = 1\ \} + \sup_{B\in\mathbf{M}_n} \{|\mathrm{Tr}[B^*C]| \ : \ \|B\|_p = 1\ \} \\ &= \|A\|_q + \|C\|_q \ . \end{aligned}$$

7.2. Convexity of $A \mapsto \mathrm{Tr}\left[(B^*A^pB)^{q/p}\right]$ and certain of its applications. For any fixed $B \in \mathbf{M}_n$ and any numbers $p, q > 0$, define $\Upsilon_{p,q}$ on $\mathbf{H}_n^+$ by

$$\Upsilon_{p,q}(A) = \mathrm{Tr}\left[(B^*A^pB)^{q/p}\right] \ . \tag{7.6}$$

7.2. Theorem. *For all $1 \le p \le 2$, and for all $q \ge 1$, $\Upsilon_{p,q}$ is convex on $\mathbf{H}_n^+$. For $0 \le p \le q \le 1$, $\Upsilon_{p,q}$ is concave on $\mathbf{H}_n^+$. For $p > 2$, there exist B such that $\Upsilon_{p,q}$ is not convex or concave for any values of $q \ne p$.*

7.3. Remark. *The function $\Upsilon_{p,q}^{1/q}$ has the same convexity and concavity properties as $\Upsilon_{p,q}$. To see this, note that $\Upsilon_{p,q}$ is homogeneous of degree $q \ge 1$. Recall that a function f that is homogeneous of degree one is convex if and only if the level set $\{x \ : \ f(x) \le 1\}$ is convex, while it is concave if and only if the level set $\{x \ : \ f(x) \ge 1\}$ is convex. Hence, if $g(x)$ is homogeneous of degree q, and convex, so that $\{x \ : \ g(x) \le 1\}$ is convex, $g^{1/q}$ is convex, and similarly for concavity.*

The concavity of $\Upsilon_{p,1}$ for $0 < p < 1$ was proved by Epstein [**11**]. The convexity of $\Upsilon_{p,1}$ was conjectured for $1 < p < 2$ and proved for $p = 2$ in [**3**], where it was also proved that neither concavity nor convexity held for $p > 2$. Finally the convexity $1 < p < 2$ was proved in [**4**], where the generalization to $q \ne 1$ was also treated.

Before giving the proof of Theorem 7.2, we give several applications.

7.4. Theorem (Lieb-Thirring trace inequality). *For all $A, B \in \mathbf{H}_n^+$ and all $t \ge 1$,*

$$\mathrm{Tr}\left[(B^{1/2}AB^{1/2})^t\right] \le \mathrm{Tr}\left[B^{t/2}A^tB^{t/2}\right] \ . \tag{7.7}$$

Proof: Define $C = A^t$ and $p = 1/t \le 1$, so that $A = C^p$. Then

$$\mathrm{Tr}\left[(B^{1/2}AB^{1/2})^t\right] - \mathrm{Tr}\left[B^{t/2}A^tB^{t/2}\right] = \mathrm{Tr}\left[(B^{1/2}C^pB^{1/2})^{1/p}\right] - \mathrm{Tr}\left[CB^{1/p}\right] \ ,$$

and by Epstein's part of Theorem 7.2 the right hand side is a concave function of C. Now we apply Example 4.15: Choose an orthonormal basis $\{u_1, \dots, u_n\}$ diagonalizing B. Let $\mathfrak{A}$ be the $*$-subalgebra of $\mathbf{M}_n$ consisting of matrices that are diagonal in this basis. Then as shown in Example 4.15, $\mathrm{E}_{\mathfrak{A}}(C)$ is an average over unitary conjugates of C, by unitaries that commute with B. It follows that

$$\mathrm{Tr}\left[(B^{1/2}C^pB^{1/2})^{1/p}\right] - \mathrm{Tr}\left[CB^{1/p}\right] \geq \mathrm{Tr}\left[(B^{1/2}(\mathrm{E}_{\mathfrak{A}}(C))^pB^{1/2})^{1/p}\right] - \mathrm{Tr}\left[(\mathrm{E}_{\mathfrak{A}}(C))B^{1/p}\right] .$$

However, since $\mathrm{E}_{\mathfrak{A}}(C)$ and B commute, the right hand side is zero. ■

Theorem 7.4 was first proved in [**22**], and has had many applications since then. The next application is taken from [**3, 4**]. We first define another trace function:

For any numbers $p, q > 0$, and any positive integer m, define $\Phi_{p,q}$ on $(\mathbf{H}_n^+)^m$, the m-fold cartesian product of $\mathbf{H}_n^+$ with itself by

$$\Phi_{p,q}(A_1, \ldots, A_m) = \| \ (\textstyle\sum_{j=1}^m A_j^p)^{1/p} \ \|_q \ . \tag{7.8}$$

7.5. Theorem. *For all $1 \leq p \leq 2$, and for all $q \geq 1$, $\Phi_{p,q}$ is jointly convex on $(\mathbf{H}_n^+)^m$, while for $0 \leq p \leq q \leq 1$, $\Phi_{p,q}$ is jointly concave on $(\mathbf{H}_n^+)^m$. For $p > 2$, $\Phi_{p,q}$ is not convex or concave, even separately, for any value of $q \neq p$.*

Proof : Define the $mn \times mn$ matrices $\mathcal{A} = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix}$ and $\mathcal{B} = \begin{bmatrix} I & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ I & 0 & \ldots & 0 \end{bmatrix}$. ($\mathcal{A}$ is block diagonal with A_j as the jth diagonal block, and $\mathcal{B}$ has $n \times n$ identities in each block in the first column, and zeros elsewhere.) Then $\mathcal{B}\mathcal{A}^p\mathcal{B}$ is the block matrix with $\sum_{j=1}^m A_j^p$ in its upper left block, and zeros elsewhere. Thus,

$$\left(\mathrm{Tr}\left[\left(\textstyle\sum_{j=1}^m A_j^p\right)^{q/p}\right]\right)^{1/q} = \left(\mathrm{Tr}\left[(\mathcal{B}\mathcal{A}^p\mathcal{B})^{q/p}\right]\right)^{1/q} .$$

By Theorem 7.2 and Remark 7.3, the right hand side is convex in $\mathcal{A}$ for all $1 \leq p \leq 2$, $q \geq 1$, and concave in $\mathcal{A}$ for $0 \leq p \leq q \leq 1$.

We now show, by means of a Taylor expansion, that both convexity and concavity fail for $p > 2$ and any $q \neq p$. By simple differentiation one finds that for any $A, B \in \mathbf{H}_n^+$,

$$\Phi_{p,q}(tA, B) = \|B\|_q + \frac{t^p}{p}\|B\|_q^{1-q}\,\mathrm{Tr}A^pB^{q-p} + O(t^{2p}) \ . \tag{7.9}$$

Keeping B fixed, but replacing A by A_1, A_2 and $(A_1 + A_2)/2$, we find

$$\frac{1}{2}\Phi_{p,q}(tA_1, B) + \frac{1}{2}\Phi_{p,q}(tA_2, B) - \Phi_{p,q}\left(t\frac{A_1 + A_2}{2},\, B\right) =$$
$$\frac{t^p}{p}\,\|B\|_q^{1-q}\left[\frac{1}{2}\mathrm{Tr}\left(A_1^pB^{q-p}\right) + \frac{1}{2}\mathrm{Tr}\left(A_2^pB^{q-p}\right) - \mathrm{Tr}\left(\left(\frac{A_1 + A_2}{2}\right)^p B^{q-p}\right)\right] + O(t^{2p}).$$

Now if $p > 2$, $A \mapsto A^p$ is not operator convex, and so we can find A_1 and A_2 in $\mathbf{H}_n^+$ and a unit vector v in $\mathbb{C}^n$ such that

$$\frac{1}{2}\langle v, A_1^p v\rangle + \frac{1}{2}\langle v, A_2^p v\rangle - \left\langle v, \left(\frac{A_1 + A_2}{2}\right)^p v\right\rangle < 0 \ , \tag{7.10}$$

and of course since $A \mapsto \mathrm{Tr}(A^p)$ is convex for $p > 2$, we can find v so that the left hand side in (7.10) is positive. For $q \neq p$, take B^{q-p} to be (a close approximation of) the rank one projection onto v. ∎

7.6. Remark. *We showed in the proof of Theorem 7.5 that $\Phi_{p,q}$ is convex or concave for given p and q whenever $\Upsilon_{p,q}$ is. Thus our proof that $\Phi_{p,q}$ is not convex or concave for $p > 2$ and $q \neq p$ implies this part of Theorem 7.2.*

We now define one more trace function, and make a simple reformulation of Theorem 7.5 that will lead to another proof of the strong subadditivity of quantum entropy.

For any numbers $p, q > 0$, and any positive integers m and n, define $\Psi_{p,q}$ on $\mathbf{H}_{mn}^+$, viewed as the space of linear operators on $\mathbb{C}^m \otimes \mathbb{C}^n$ by

$$\Psi_{p,q}(A) = \| \left(\mathrm{Tr}_1 A^p\right)^{1/p} \|_q \ . \tag{7.11}$$

7.7. Theorem. *For $1 \le p \le 2$ and $q \ge 1$, $\Psi_{p,q}$ is convex on $\mathbf{H}_{mn}^+$, while for $0 \le p \le q \le 1$, $\Psi_{p,q}$ is concave on $\mathbf{H}_{mn}^+$.*

Proof: We shall apply Theorem 5.7. Let $\mathfrak{A}$ be the subalgebra of $\mathbf{M}_{mn}$, identified with $\mathbb{C}^m \otimes \mathbb{C}^n$, consisting of operators of the form $I_{m\times m} \otimes B$, $B \in \mathbf{M}_n$. By Theorem 5.7,

$$\frac{1}{m} I_{m\times m} \otimes \mathrm{Tr}_1 A^p = \mathrm{E}_{\mathfrak{A}}(A^p) \ ,$$

and so

$$\Psi_{p,q}(A^p) = m^{1/p} \| \left(\mathrm{E}_{\mathfrak{A}}(A^p)\right)^{1/p} \|_q \ . \tag{7.12}$$

The factor of $n^{1/p}$ does not affect the convexity properties of $\Psi_{p,q}$, and so it suffices to consider the convexity properties of $A \mapsto \| \left(\mathrm{E}_{\mathfrak{A}}(A^p)\right)^{1/p} \|_q$.

By Theorem 4.13 and the remark following its proof, $\mathrm{E}_{\mathfrak{A}}(A^p)$ is a limit of operators $\mathcal{C}_k(A^p)$ of the form

$$\mathcal{C}_k(A^p) = \sum_{j=1}^{N_k} p_{k,j} I_{m\times m} \otimes W_{k,j} A^p I_{m\times m} \otimes W_{k,j}^* \ . \tag{7.13}$$

where each $W_{k,j}$ is a unitary matrix in $\mathfrak{A}'$ (which means that it has the form $I_{m\times m}\otimes U$ where U is unitary in $\mathbf{M}_n$). But then

$$\begin{aligned} n^{-1/p}\Psi_{p,q}(A) &= \lim_{k\to\infty}\left\|\left(\sum_{j=1}^{N_k} p_{k,j}W_{k,j}A^pW_{k,j}^*\right)^{1/p}\right\|_q \\ &= \lim_{k\to\infty}\left\|\left(\sum_{j=1}^{N_k} p_{k,j}\left(p_{k,j}^{1/p}W_{k,j}AW_{k,j}^*\right)^p\right)^{1/p}\right\|_q \\ &= \lim_{k\to\infty}\Phi_{p,q}\left(p_{k,1}^{1/p}W_{k,1}AW_{k,1}^*\ ,\dots\ ,p_{k,N_k}^{1/p}W_{k,N_k}AW_{k,N_k}^*\right)\ . \end{aligned}$$

Since a limit of convex functions is convex, we see that $\Psi_{p,q}$ is convex or concave whenever $\Phi_{p,q}$ is. The reverse implication is even more elementary: To see this, suppose that the matrix A in Theorem 7.7 is the block diagonal matrix whose jth diagonal block is A_j. Then, clearly, $\Psi_{p,q}(A) = \Phi_{p,q}(A_1, A_2, \dots, A_m)$. ■

We now return to the theme with which we began this section, and explain how to deduce the strong subadditivity of quantum entropy from Theorem 7.7.

Let ρ be a density matrix on $\mathcal{H}_1\otimes\mathcal{H}_2 = \mathbb{C}^m\otimes\mathbb{C}^n$. Let $\rho_1 = \mathrm{Tr}_{\mathcal{H}_2}\rho$ and $\rho_2 = \mathrm{Tr}_{\mathcal{H}_1}\rho$ be its two partial traces. As in our previous discussion of subadditivity, we shall also use ρ_{12} to denote the full density matrix ρ.

Then a simple calculation shows that

$$\frac{\mathrm{d}}{\mathrm{d}p}\Psi_{p,1}(\rho)\bigg|_{p=1} = S(\rho_2) - S(\rho_{12})\ . \tag{7.14}$$

To see this, observe that for a positive operator A, and ε close to zero,

$$A^{1+\varepsilon} = A + \varepsilon A\ln A + \mathcal{O}(\varepsilon^2)\ .$$

At least in finite dimensions, one can take a partial trace of both sides, and the resulting identity still holds.

Applying this with $A = \rho = \rho_{12}$, we compute

$$\mathrm{Tr}_{\mathcal{H}_1}\left(\rho^{1+\varepsilon}\right) = \rho_2 + \varepsilon\mathrm{Tr}_{\mathcal{H}_1}(\rho_{12}\ln\rho_{12}) + \mathcal{O}(\varepsilon^2)\ .$$

Then, since to leading order in ε, $1/(1+\varepsilon)$ is $1-\varepsilon$,

$$\left[\mathrm{Tr}_{\mathcal{H}_1}(\rho^{1+\varepsilon})\right]^{1/(1+\varepsilon)} = \rho_2 + \varepsilon\mathrm{Tr}_{\mathcal{H}_1}((\rho_{12}\ln\rho_{12}) - \varepsilon\rho_2\ln\rho_2 + \mathcal{O}(\varepsilon^2)\ .$$

Thus,

$$\mathrm{Tr}_{\mathcal{H}_2}(\left(\left[\mathrm{Tr}_{\mathcal{H}_1}((\rho^{1+\varepsilon})\right]^{1/(1+\varepsilon)}\right) = 1 - \varepsilon S(\rho_{12}) + \varepsilon S(\rho_2) + \mathcal{O}(\varepsilon^2)\ . \tag{7.15}$$

This proves (7.14).

Now we apply Theorem 7.7. Let us take the case where $\mathcal{H}_2$ is replaced by a tensor product of two finite dimensional Hilbert spaces $\mathcal{H}_2\otimes\mathcal{H}_3$, so that we identify operators on $\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3$ with $\mathbf{M}_{mn}$ where $m = \dim(\mathcal{H}_1)$ and $n = \dim(\mathcal{H}_2)\times\dim(\mathcal{H}_3)$. Let $\mathfrak{A}$ denote the $*$-subalgebra of $\mathbf{M}_{mn}$ consisting of operators of the form $A\otimes I_{\mathcal{H}_3}$ where A is an operator on $\mathcal{H}_1\otimes\mathcal{H}_2$.

We now claim that for any density matrix ρ_{123} on $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$,

$$\Psi_{p,1}(\rho_{123}) \geq \Psi_{p,1}(\mathrm{E}_{\mathfrak{A}}(\rho_{123})) \ . \tag{7.16}$$

To see this, we apply Theorem 4.13 as in the proof of Theorem 7.7, together with the fact that the unitaries in $\mathfrak{A}'$ are of the form $I_{\mathcal{H}_1\otimes\mathcal{H}_2} \otimes U$, where U is unitary on $\mathcal{H}_3$.

Then by (7.14) and Theorem 5.7,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}p}\Psi_{p,1}(\mathrm{E}_{\mathfrak{A}}(\rho_{123}))\bigg|_{p=1} &= S(\mathrm{Tr}_{\mathcal{H}_1}[\mathrm{E}_{\mathfrak{A}}(\rho_{123})]) - S(\mathrm{E}_{\mathfrak{A}}(\rho_{123})) \\ &= S(\rho_2) - S(\rho_{12}) \ . \end{aligned}$$

Also, directly from (7.14),

$$\frac{\mathrm{d}}{\mathrm{d}p}\Psi_{p,1}(\rho_{123})\bigg|_{p=1} = S(\rho_{23}) - S(\rho_{123}) \ . \tag{7.17}$$

Then, since at $p = 1$, both sides of (7.16) equal one,

$$S(\rho_{23}) - S(\rho_{123}) \geq S(\rho_2) - S(\rho_{12}) \ ,$$

which of course is equivalent to (6.11).

This shows that the strong subadditivity of quantum entropy can be viewed as a consequence, via differentiation in p, of the inequality of Theorem 7.7. One may therefore view the inequality of Theorem 7.7 as a generlaization of the strong subadditivity inequality. For another L^p inequality that can be differentiated to yield strong subadditivity, namely a Minkowski type inequality for traces of operators on a tensor product of three Hilbert spaces, see [**3, 4**]. For other applications of Theorem 7.2, see [**4**].

7.3. Proof of the convexity of $A \mapsto \mathrm{Tr}\left[(B^*A^pB)^{q/p}\right]$. We now close this section by proving Theorem 7.2, and thus completing the proofs of all of the theorems in this subsection.

We prepare the way for the proof of Theorem 7.2 with some lemmas. The proof of convexity of $\Upsilon_{p,q}$ divides into two cases, namely $1 \leq q \leq p \leq 2$ and $1 \leq p \leq 2$ with $q > p$.

The latter case, $q > p$, is the easier one, and the next lemma takes care of it:

7.8. Lemma. *For $1 \leq p \leq 2$ and $q \geq p$, $\Upsilon_{p,q}$ is convex on $\mathbf{H}_n^+$.*

Proof: Since $r := q/p \geq 1$ and since $B^*A^pB \geq 0$, we can write

$$\|B^*A^pB\|_r = \sup_{\substack{\|Y\|_{r'} \leq 1, \\ Y \geq 0}} \mathrm{Tr}(B^*A^pBY) \tag{7.18}$$

where $1/r + 1/r' = 1$. Since A^p is well known to be operator convex in A for $1 \leq p \leq 2$, so is B^*A^pB. Since the right side of (7.18) is the supremum of a family of convex functions (note that $Y \geq 0$ is needed here) we conclude that $\|B^*A^pB\|_r$ is convex. ($\Upsilon_{p,q}(A)$ is the rth power of this quantity and is therefore convex.) ■

The case $q < p$ requires more tools. The first of these is a variational formula for pth roots.

For $r > 1$, and $c, x > 0$, the arithmetic–geometric mean inequality says

$$\frac{1}{r}c^r + \frac{r-1}{r}x^r \geq cx^{r-1} ,$$

and hence

$$c = \frac{1}{r}\inf\left\{\frac{c^r}{x^{r-1}} + (r-1)x \ : \ x > 0\right\} . \tag{7.19}$$

With the infimum replaced by a supremum, the resulting formula is valid for $0 < r < 1$, as one easily checks.

We shall build this into a variational formula for $\Upsilon_{p,q}$. It is first useful to note that since B^*A^pB and $A^{p/2}BB^*A^{p/2}$ have the same spectrum,

$$\Upsilon_{p,q}(A) = \mathrm{Tr}\left[(A^{p/2}BB^*A^{p/2})^{q/p}\right] . \tag{7.20}$$

7.9. Lemma. *For any positive $n \times n$ matrix A, and with $r = p/q > 1$,*

$$\Upsilon_{p,q}(A) = \frac{1}{r}\inf\left\{\mathrm{Tr}\left[A^{p/2}B\frac{1}{X^{r-1}}B^*A^{p/2} + (r-1)X\right] \ : \ X > 0\right\} \tag{7.21}$$

where the infimum is taken over all positive $n \times n$ matrices X. Likewise, if the infimum replaced by a supremum, the resulting formula is valid for $r = p/q < 1$.

Proof: Let $C = B^*A^{p/2}$. By continuity we may assume that C^*C is strictly positive. Then, for $r > 1$, there is a minimizing X. Let

$$Y = X^{1-r}$$

and note that minimizing

$$\mathrm{Tr}\left[A^{p/2}B\frac{1}{X^{r-1}}B^*A^{p/2} + (r-1)X\right]$$

with respect to $X > 0$ is the same as minimizing

$$\mathrm{Tr}\left(CC^*Y + (r-1)Y^{-1/(r-1)}\right)$$

with respect to $Y > 0$. Since the minimizing Y is strictly positive, we may replace the minimizing Y by $Y + tD$, with D self adjoint, and set the derivative with respect to t equal to 0 at $t = 0$. This leads to $\mathrm{Tr}D[CC^* - Y^{-r/(r-1)}] = 0$. Therefore $Y^{-r/(r-1)} = CC^*$ and we are done. The variational formula for $p/q < 1$ is proved in the same manner. ■

7.10. Lemma. *If $f(x, y)$ is jointly convex, then $g(x)$ defined by $g(x) = \inf_y f(x, y)$ is convex. The analogous statement with convex replaced by concave and infimum replaced by supremum is also true.*

Proof: For $\varepsilon > 0$, choose (x_0, y_0) and (x_1, y_1) so that

$$f(x_0, y_0) \leq g(x_0) + \varepsilon \qquad \text{and} \qquad f(x_1, y_1) \leq g(x_1) + \varepsilon .$$

Then:

$$\begin{aligned} g((1-\lambda)x_0+\lambda x_1) &\leq f((1-\lambda)x_0+\lambda x_1,(1-\lambda)y_0+\lambda y_1)\\ &\leq (1-\lambda)f(x_0,y_0)+\lambda f(x_1,y_1)\\ &\leq (1-\lambda)g(x_0)+\lambda g(x_1)+\varepsilon\ . \end{aligned}$$ ■

On account of Lemmas 7.9 and 7.10 we shall be easily able to prove the stated convexity and concavity properties of $\Upsilon_{p,q}$ once we have proved:

7.11. LEMMA. *The map*

$$(A,X)\mapsto \mathrm{Tr}\left(A^{p/2}B^*\frac{1}{X^{r-1}}BA^{p/2}\right) \tag{7.22}$$

is jointly convex on $\mathbf{H}_n^+\times\mathbf{H}_n^+$ *for all* $1\leq r\leq p\leq 2$ *and is jointly concave for all* $0<p<r<1$.

Proof: We first rewrite the right hand side of (7.22) in a more convenient form: Define

$$Z=\begin{bmatrix} A & 0\\ 0 & X\end{bmatrix}\quad\text{and}\quad K=\begin{bmatrix} 0 & 0\\ B & 0\end{bmatrix}\quad\text{so that}\quad K^*Z^{1-p}K=\begin{bmatrix} B^*X^{1-p}B & 0\\ 0 & 0\end{bmatrix}.$$

Then, by cyclicity of the trace, $\mathrm{Tr}(Z^pK^*Z^{1-r}K)=\mathrm{Tr}\left(A^{p/2}B^*\frac{1}{X^{r-1}}BA^{p/2}\right)$. Note that convexity/concavity of the left hand side in Z is the same as convexity/concavity of the right hand side in (A,X). The result now follows from Theorem 6.1, the Lieb Concavity Theorem and Theorem 6.2, the Ando Convexity Theorem. ■

Proof of Theorem 7.2: By Lemma 7.11, the mapping in (7.22) is jointly convex for $1\leq r\leq p\leq 2$. Then taking $r=p/q$, we have from Lemma 7.9 and Lemma 7.10 $\Upsilon_{p,q}(A)=\inf_X f(A,X)$ where $f(A,X)$ is jointly convex in A and X. The convexity of $\Upsilon_{p,q}$ now follows by Lemma 7.10. The concavity statement is proved in the same way. We have already observed that $\Phi_{p,q}$ inherits its convexity and concavity properties from those of $\Upsilon_{p,q}$, and thus, having shown in the proof of Theorem 7.5 that $\Phi_{p,q}$ is neither convex nor concave for $p>2$ and $q\neq p$, the same is true for $\Upsilon_{p,q}$. ■

8. Brascamp-Lieb type inequalities for traces

We recall the original Young's inequality: For non-negative measurable functions f_1, f_2 and f_3 on $\mathbb{R}$, and $1/p_1+1/p_2+1/p_3=2$

$$\int_{\mathbb{R}^2} f_1(x)f_2(x-y)f_3(y)\mathrm{d}x\mathrm{d}y\leq \left(\int_{\mathbb{R}} f_1^{p_1}(t)\mathrm{d}t\right)^{1/p_1}\left(\int_{\mathbb{R}} f_2^{p_2}(t)\mathrm{d}t\right)^{1/p_2}\left(\int_{\mathbb{R}} f_3^{p_3}(t)\mathrm{d}t\right)^{1/p_3}. \tag{8.1}$$

Define the maps $\phi_j:\mathbb{R}^2\to\mathbb{R}$, $j=1,2,3$, by

$$\phi_1(x,y)=x\qquad \phi_2(x,y)=x-y\qquad\text{and}\qquad \phi_3(x,y)=y\ .$$

Then (8.1) can be rewritten as

$$\int_{\mathbb{R}^2}\left(\prod_{j=1}^{3} f_j\circ\phi_j\right)\mathrm{d}^2x \le \prod_{j=1}^{3}\left(\int_{\mathbb{R}} f_j^{p_j}(t)\mathrm{d}t\right)^{1/p_j} \ . \tag{8.2}$$

There is now no particular reason to limit ourselves to products of only three functions, or to integrals over $\mathbb{R}^2$ and $\mathbb{R}$, or even any Euclidean space for that matter:

8.1. Definition. Given measure spaces $(\Omega,\mathcal{S},\mu)$ and $(M_j,\mathcal{M}_j,\nu_j)$, $j = 1,\dots,N$, not necessarily distinct, together with measurable functions $\phi_j:\Omega\to M_j$ and numbers $p_1,\dots,p_N$ with $1\le p_j\le\infty$, $1\le j\le N$, we say that *a generalized Young's inequality holds for* $\{\phi_1,\dots,\phi_N\}$ *and* $\{p_1,\dots,p_N\}$ in case there is a finite constant C such that

$$\int_\Omega \prod_{j=1}^{N} f_j\circ\phi_j \mathrm{d}\mu \le C\prod_{j=1}^{N}\|f_j\|_{L^{p_j}(\nu_j)} \tag{8.3}$$

holds whenever f_j is non-negative and measurable on M_j, $j=1,\dots,N$.

8.1. A generalized Young's inequality in the context of non-commutative integration.

In non-commutative integration theory, as expounded by I. Segal [**28, 30**] and J. Dixmier [**10**], the basic data is an operator algebra $\mathcal{A}$ equipped with a positive linear functional λ. (For more information, see especially Nelson's paper [**23**].)

The algebra $\mathcal{A}$ corresponds to the algebra of bounded measurable functions, and applying the positive linear functional λ to a positive operator corresponds to taking the integral of a positive function. That is,

$$A\mapsto\lambda(A) \qquad \text{corresponds to} \qquad f\mapsto\int_M f\mathrm{d}\nu \ .$$

To frame an analog of (8.3) in an operator algebra setting, we replace the measure spaces by non-commutative integration spaces:

$$(M_j,\mathcal{M}_j,\nu_j)\longrightarrow(\mathcal{A}_j,\lambda_j) \qquad j=1,\dots,N$$

and

$$(\Omega,\mathcal{S},\mu)\longrightarrow(\mathcal{B},\lambda) \ .$$

The right hand side of (8.3) is easy to generalize to the operator algebra setting; for $A\in(\mathcal{A},\lambda)$, and $1\le q\le\infty$, we define

$$\|A\|_{q,\lambda} = (\lambda(|A|)^q)^{1/q} \ .$$

Then the natural analog of the right hand side of (8.3) is

$$\prod_{j=1}^{N}\|A_j\|_{(q_j,\lambda_j)} \ .$$

As for the left hand side of (8.3), regard $f_j\mapsto f_j\circ\phi_j$ can easily be interpreted in operator algebra terms: Think of $L^\infty(\Omega)$ and $L^\infty(M_j)$ as (commutative) operator algebras. Indeed, we can consider their elements as multiplication operators in the obvious way. Then the map $f_j \mapsto f_j \circ \phi_j$ is an operator algebra homomorphism; i.e. , a linear transformation respecting the product and the conjugation $*$.

Therefore, suppose we are given operator algebra homomorphisms

$$\phi_j : \mathcal{A}_j \to \mathcal{A} \ .$$

Then each $\phi_j(A_j)$ belongs to $\mathcal{A}$, however in the non-commutative case, the product of the $\phi_j(A_j)$ depends on the order, and need not be self adjoint even – let alone positive – even if each of the A_j are positive.

Therefore, let us return to the left side of (8.3), and suppose that each f_j is strictly positive. Then defining

$$h_j = \ln(f_j) \qquad \text{so that} \qquad f_j \circ \phi_j = e^{h_j \circ \phi_j} \ ,$$

we can rewrite (8.3) as

$$\int_\Omega \exp\left(\sum_{j=1}^N h_j \circ \phi_j\right) \mathrm{d}\mu \le C \prod_{j=1}^N \|e^{h_j}\|_{L^{p_j}(\nu_j)} \ , \tag{8.4}$$

We can now formulate our operator algebra analog of (8.3):

8.2. Definition. Given non-commutative integration spaces $(\mathcal{A}, \lambda)$ and $(\mathcal{A}_j, \lambda_j)$, $j = 1, \dots, N$, together with operator algebra homomorphisms $\phi_j : \mathcal{A}_j \to \mathcal{A}$, $j = 1, \dots, N$, and indices $1 \le p_j \le \infty$, $j = 1, \dots, N$, a *generalized Young's inequality holds for* $\{\phi_1, \dots, \phi_N\}$ *and* $\{p_1, \dots, p_N\}$ if there is a finite constant C so that

$$\lambda\left(\exp\left[\sum_{j=1}^N \phi_j(H_j)\right]\right) \le C \prod_{j=1}^N \left(\lambda_j(\exp[p_j H_j])\right)^{1/p_j} \tag{8.5}$$

whenever H_j is self adjoint in $\mathcal{A}_j$, $j = 1, \dots, N$.

We are concerned with determining the indices and the best constant C for which such an inequality holds, and shall focus on one example arising in mathematical physics.

8.2. A generalized Young's inequality for tensor products. Let $\mathcal{H}_j$, $j = 1, \dots, N$ be separable Hilbert spaces, and let $\mathcal{K}$ denote the tensor product

$$\mathcal{K} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N \ .$$

Define $\mathcal{A}$ to be $\mathcal{B}(\mathcal{K})$, the algebra of bounded linear operators on $\mathcal{K}$, and define λ to be the trace Tr on $\mathcal{K}$, so that $(\mathcal{A}, \lambda) = (\mathcal{B}(\mathcal{K}), \mathrm{Tr})$.

For any non-empty subset J of $\{1, \dots, n\}$, let $\mathcal{K}_J$ denote the tensor product

$$\mathcal{K}_J = \otimes_{j \in J} \mathcal{H}_j \ .$$

Define $\mathcal{A}_J$ to be $\mathcal{B}(\mathcal{K}_J)$, the algebra of bounded linear operators on $\mathcal{K}_J$, and define λ_J be the trace on $\mathcal{K}_J$, so that $(\mathcal{A}_J, \lambda_J) = (\mathcal{B}(\mathcal{K}_J), \mathrm{Tr}_J)$.

There are natural endomorphisms ϕ_J embedding the $2^N - 1$ algebras $\mathcal{A}_J$ into $\mathcal{A}$. For instance, if $J = \{1, 2\}$,

$$\phi_{\{1,2\}}(A_1 \otimes A_2) = A_1 \otimes A_2 \otimes I_{\mathcal{H}_3} \otimes \cdots \otimes I_{\mathcal{H}_N} \ , \tag{8.6}$$

and is extended linearly.

It is obvious that in case $J \cap K = \emptyset$ and $J \cup K = \{1, \ldots, n\}$, then for all $H_J \in \mathcal{A}_J$ and $H_K \in \mathcal{A}_K$,

$$\text{Tr}\left(e^{H_J + H_K}\right) = \text{Tr}_J\left(e^{H_J}\right)\text{Tr}_K\left(e^{H_K}\right) \ , \tag{8.7}$$

but things are more interesting when $J \cap K \neq \emptyset$ and J and K are both proper subsets of $\{1, \ldots, N\}$. The following is proved in [**6**]:

8.3. Theorem (Generalized Young's Inequality for Tensor Products). *Let $J_1, \ldots, J_N$ be N non-empty subsets of $\{1, \ldots, n\}$ For each $i \in \{1, \ldots, n\}$, let $p(i)$ denote the number of the sets $J_1, \ldots, J_N$ that contain i, and let p denote the minimum of the $p(i)$. Then, for self adjoint operators H_j on $\mathcal{K}_{J_j}$, $j = 1, \ldots, N$,*

$$\text{Tr}\left(\exp\left(\sum_{j=1}^N \phi_{J_j}(H_j)\right)\right) \leq \prod_{j=1}^N \left(\text{Tr}_{J_j}\ e^{qH_j}\right)^{1/q} \tag{8.8}$$

for all $1 \leq q \leq p$, while for all $q > p$, it is possible for the left hand side to be infinite, while the right hand side is finite.

Note that in the generalized Young's inequality in Theorem 8.3, the constant C in Definition (8.2) is 1.

The fact that the constant $C = 1$ is best possible, and that the inequality cannot hold for $q > p$ is easy to see by considering the case that each $\mathcal{H}_j$ has finite dimension d_j, and $H_j = 0$ for each j. Then

$$\text{Tr}\left(\exp\left(\sum_{j=1}^N \phi_{J_j}(H_j)\right)\right) = \prod_{k=1}^n d_k$$

$$\prod_{j=1}^N \left(\text{Tr}_{J_j} e^{qH_j}\right)^{1/q} = \prod_{j=1}^N \prod_{k \in J_j} d_k^{1/q} = \prod_{k=1}^n d_k^{p(k)/q} \ .$$

Moreover, since for $q > p$, $\left(\text{Tr}\ e^{pH_j}\right)^{1/p} > \left(\text{Tr}\ e^{qH_j}\right)^{1/q}$, it suffices to prove the inequality (8.8) for $q = p$. We will do this later in this section.

As an example, consider the case of overlapping pairs with a periodic boundary condition:

$$J_j = \{j, j+1\} \qquad j = 1, \ldots, n-1 \qquad \text{and} \quad J_n = \{n, 1\} \ .$$

Here, $N = n$, and obviously $p = 2$. Therefore,

$$\text{Tr}\left(\exp\left(\sum_{j=1}^N \phi_j(H_j)\right)\right) \leq \prod_{j=1}^N \left(\text{Tr}\ e^{2H_j}\right)^{1/2} \ . \tag{8.9}$$

The inequality (8.9) has an interesting statistical mechanical interpretation as a bound on the *partition function* of an arbitrarily long chain of interacting spins in terms of a product of partition functions of simple constituent two–spin systems.

Again, the inequality (8.9) is non-trivial due to the "overlap" in the algebras $\mathcal{A}_j$.

8.3. Subadditivity of Entropy and Generalized Young's Inequalities. In the examples we consider, the positive linear functionals λ under consideration are either traces or normalized traces. Throughout this section, we assume that our non-commutative integration spaces $(\mathcal{A}, \lambda)$ are based on *tracial* positive linear functionals λ. That is, we require that for all $A, B \in \mathcal{A}$,

$$\lambda(AB) = \lambda(BA) \ .$$

In such a non-commutative integration space $(\mathcal{A}, \lambda)$, a *probability density* is a non-negative element ρ of $\mathcal{A}$ such that $\lambda(\rho) = 1$. Indeed, the tracial property of λ ensures that

$$\lambda(\rho A) = \lambda(A\rho) = \lambda(\rho^{1/2} A \rho^{1/2})$$

so that $A \mapsto \lambda(\rho A)$ is a positive linear functional that is 1 on the identity.

Now suppose we have N non-commutative integration spaces $(\mathcal{A}_j, \lambda_j)$ and operator algebra homomorphisms $\phi_j : \mathcal{A}_j \to \mathcal{A}$. Then these homomorphisms induce maps from the space of probability densities on $\mathcal{A}$ to the spaces of probability densities on the $\mathcal{A}_j$, as follows:

For any probability density ρ on $(\mathcal{A}, \lambda)$, let ρ_j be the probability density on $(\mathcal{A}_j, \lambda_j)$ by

$$\lambda_j(\rho_j A) = \lambda(\rho \phi_j(A))$$

for all $A \in \mathcal{A}_j$.

For example, in the setting we are discussing here, ρ_{J_j} is just the partial trace of ρ over $\otimes_{k \in J_j^c} \mathcal{H}_k$ leaving an operator on $\otimes_{k \in J_j} \mathcal{H}_k$.

In this section, we are concerned with the relations between the *entropies* of ρ and the $\rho_1, \dots, \rho_N$. The entropy of a probability density ρ, $S(\rho)$, is defined by

$$S(\rho) = -\lambda(\rho \ln \rho) \ .$$

Evidently, the entropy functional is concave on the set of probability densities.

8.4. DEFINITION. Given tracial non-commutative integration spaces $(\mathfrak{A}, \lambda)$ and $(\mathfrak{A}_j, \lambda_j)$, $j = 1, \dots, N$, together with C^* algebra endomorphisms $\phi_j : \mathfrak{A}_j \to \mathfrak{A}$, $j = 1, \dots, N$, and numbers $1 \le p_j \le \infty$, $j = 1, \dots, N$, a *generalized subadditivity of entropy inequality* holds if there is a finite constant C so that

$$\sum_{j=1}^{N} \frac{1}{p_j} S(\rho_j) \ge S(\rho) - \ln C \tag{8.10}$$

for all probability densities ρ in $\mathfrak{A}$.

The following is a non-commutative version of theorem proved in [**5**]. The non-commuative version is proved in [**6**].

8.5. THEOREM (Duality for Generalized Youngs Inequalities and Entropy). *Let $(\mathfrak{A}, \lambda)$ and $(\mathfrak{A}_j, \lambda_j)$, $j = 1, \dots, N$, be tracial non-commutative integration spaces. Let $\phi_j : \mathfrak{A}_j \to \mathfrak{A}$, $j = 1, \dots, N$ be C^* algebra endomorphisms.*

Then for any numbers $1 \le p_j \le \infty$, $j = 1, \ldots, N$, and any finite constant C, the generalized subadditivity of entropy inequality

$$\sum_{j=1}^{N} \frac{1}{p_j} S(\rho_j) \ge S(\rho) - \ln C$$

is true for all probability densities ρ on $\mathfrak{A}$ if and only if the generalized Young's inequality

$$\lambda\left(\exp\left[\sum_{j=1}^{N} \phi_j(H_j)\right]\right) \le C \prod_{j=1}^{N} \left(\lambda_j \exp\left[p_j H_j\right]\right)^{1/p_j}$$

is true for all self-adjoint $H_j \in \mathfrak{A}_j$, $j = 1, \ldots, N$, with the same $p_1, \ldots, p_N$ and the same C.

Proof of Theorem 8.5: We make use of the duality formula for the entropy given in Theorem 2.13. Suppose first that the generalized Young's inequality (8.5) holds. Then, for any probability density ρ in $\mathfrak{A}$, and any self adjoint $H_j \in \mathfrak{A}_j$, $j = 1, \ldots, N$, we have

$$\begin{aligned} -S(\rho) &\ge \lambda\left(\rho\left[\sum_{j=1}^{N} \phi_j(H_j)\right]\right) - \ln\left[\lambda\left(\exp\left[\sum_{j=1}^{N} \phi_j(H_j)\right]\right)\right] \\ &= \sum_{j=1}^{N} \lambda_j(\rho_j H_j) - \ln\left[\lambda\left(\exp\left[\sum_{j=1}^{N} \phi_j(H_j)\right]\right)\right] \\ &\ge \sum_{j=1}^{N} \lambda_j(\rho_j H_j) - \ln\left[C \prod_{j=1}^{N} \lambda_j\left(e^{p_j H_j}\right)^{1/p_j}\right] \\ &= \sum_{j=1}^{N} \frac{1}{p_j}\left[\lambda_j(\rho_j[p_j H_j]) - \ln\left(\lambda_j\left(e^{[p_j H_j]}\right)\right)\right] - \ln C \ . \end{aligned}$$

Now choosing $p_j H_j$ to maximize $\lambda_j(\rho_j[p_j H_j]) - \ln\left(\lambda_j\left(e^{[p_j H_j]}\right)\right)$, we get

$$\lambda_j(\rho_j[p_j H_j]) - \ln\left(\lambda_j\left(e^{[p_j H_j]}\right)\right) = -S(\rho_j) = \lambda_j(\rho_j \ln \rho_j) \ .$$

Next, suppose that the subadditivity inequality is true. Let the self adjoint operators $H_1, \ldots, H_N$ be given, and define

$$\rho = \left[\lambda\left(\exp\left[\sum_{j=1}^{N} \phi_j(H_j)\right]\right)\right]^{-1} \exp\left[\sum_{j=1}^{N} \phi_j(H_j)\right] \ .$$

Then by Theorem 2.13,

$$\begin{aligned}\ln\left[\lambda\left(\exp\left[\sum_{j=1}^{N}\phi_j(H_j)\right]\right)\right] &= \lambda\left(\rho\left[\sum_{j=1}^{N}\phi_j(H_j)\right]\right)+S(\rho)\\ &= \sum_{j=1}^{N}\lambda_j\left[\rho_j H_j\right]+S(\rho)\\ &\le \sum_{j=1}^{N}\frac{1}{p_j}\left[\lambda_j\left[\rho_j(p_jH_j)\right]+S(\rho_j)\right]+\ln C\\ &\le \sum_{j=1}^{N}\frac{1}{p_j}\ln\left[\lambda_j\left(\exp(p_jH_j)\right)\right]+\ln C\end{aligned}$$

■

Proof of Theorem 8.3:

By Theorem 8.5, in order to prove Theorem 8.3, it suffices to prove the corresponding generalized subadditivity of entropy inequality for tensor products of Hilbert spaces, which we now formulate and prove.

The crucial tool that we use here is the strong subadditivity of the entropy; i.e., Theorem 6.6, except that we shall use a slightly different indexing of the various partial traces that is better adapted to our application.

Suppose, as in the case we are discussing, that we are given n separable Hilbert spaces $\mathcal{H}_1,\dots,\mathcal{H}_n$. As before, let $\mathcal{K}$ denote their tensor product, and for any non-empty subset J of $\{1,\dots,n\}$, let $\mathcal{K}_J$ denote $\otimes_{j\in J}\mathcal{H}_j$.

For a density matrix ρ on $\mathcal{K}$, and any non-empty subset J of $\{1,\dots,n\}$, define $\rho_J = \mathrm{Tr}_{J^c}\rho$ to be the density matrix on $\mathcal{K}_J$ induced by the natural injection of $\mathcal{B}(\mathcal{K}_J)$ into $\mathcal{B}(\mathcal{K})$. As noted above, ρ_J is nothing other than the partial trace of ρ over the complementary product of Hilbert spaces, $\otimes_{j\notin J}\mathcal{H}_j$.

The strong subadditivity of the entropy of Theorem 6.6 can be formulated as the statement that for all non-empty $J,K\subset\{1,\dots,n\}$,

$$S(\rho_J)+S(\rho_K)\ge S(\rho_{J\cup K})+S(\rho_{J\cap K})\ . \tag{8.11}$$

In case $J\cap K=\emptyset$, it reduces to the ordinary subadditivity of the entropy, which is the elementary inequality

$$S(\rho_J)+S(\rho_K)\ge S(\rho_{J\cup K})\qquad\text{for}\quad J\cap K=\emptyset\ . \tag{8.12}$$

Combining these, we have

$$\begin{aligned}S(\rho_{\{1,2\}})+S(\rho_{\{2,3\}})+S(\rho_{\{3,1\}}) &\ge S(\rho_{\{1,2,3\}})+S(\rho_{\{2\}})+S(\rho_{\{1,3\}})\\ &\ge 2S(\rho_{\{1,2,3\}})\ ,\end{aligned} \tag{8.13}$$

where the first inequality is the strong subadditivity (8.11) and the second is the ordinary subadditivity (8.12). Thus, for $n=3$ and $J_1=\{1,2\}$, $J_2=\{2,3\}$ and

$J_3 = \{3,1\}$, we obtain

$$\frac{1}{2}\sum_{j=1}^{N} S(\rho_{J_j}) \geq S(\rho)\ .$$

8.6. THEOREM. *Let $J_1,\dots,J_N$ be N non-empty subsets of $\{1,\dots,n\}$. For each $i \in \{1,\dots,n\}$, let $p(i)$ denote the number of the sets $J_1,\dots,J_N$ that contain i, and let p denote the minimum of the $p(i)$. Then*

$$\frac{1}{p}\sum_{j=1}^{N} S(\rho_{J_j}) \geq S(\rho) \tag{8.14}$$

for all density matrices ρ on $\mathcal{K} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$.

Proof: Simply use strong subadditivity to combine overlapping sets to produce as many "complete" sets as possible, as in the example above. Clearly, there can be no more than p of these. If $p(i) > p$ for some indices i, there will be "left over" partial sets. The entropy is always non-negative, and therefore, discarding the corresponding entropies gives us $\sum_{j=1}^{N} S(\rho_{J_j}) \geq pS(\rho)$, and hence the inequality. ■

Proof of Theorem 8.3: This now follows directly from Theorem 8.5 and Theorem 8.6. ■

Acknoledgements

I would like to thank the organizers of the 2009 Arizona Summer School *Entropy and the quantum*, Robert Sims and Daniel Ueltschi, for the invitation to deliver these lectures. Some of the results presented here were obtained in collaboration with Elliott Lieb, and beyond that, my view of the subject presented in these notes draws very much on the many discussions we had in the course of these collaborations. Though I take full responsibility for any shortcomings in these notes, I would like to thank Elliott Lieb for many enlightening conversations that have certainly added to whatever utility these notes may find. I would also like to thank Robert Seiringer for very carefully reading them, correcting many misprints, and making valuable suggestions.

References

[1] T. Ando: *Convexity of certain maps on positive definite matrices and applications to Hadamard products*, Lin. Alg. and Appl., **26** 203–241 (1979)

[2] H.J. Brascamp and E.H. Lieb: *The best constant in Young's inequality and its generalization to more than three functions*,. J. Math. Pures Appl. **86**, no. 2, 2006 pp. 89–99.

[3] E.A. Carlen and E.H. Lieb: *A Minkowski type trace inequality and strong subadditivity of quantum entropy*, Advances in the Mathematical Sciences, AMS Transl., **189** Series 2, (1999) 59-68. arXiv:math/0701352.

[4] E.A. Carlen and E.H. Lieb: *A Minkowski type trace inequality and strong subadditivity of the quantum entropy II* Lett. Math. Phys., Vol. 83, no. 2, (2008) pp. 107-126

[5] E.A. Carlen and D. Cordero–Erausquin: *Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities*, to appear in Geom. Funct. Anal. (2008).

[6] E.A. Carlen and E.H. Lieb: *Bascamp-Lieb inequalities for non-commutative integration* Documenta Mathematica, **13** (2008) 553-584

[7] C. Davis *A Schwarz inequality ofr operator convex functions.* Proc. Amer. Math. Soc., **8**, 42–44, (1957)

[8] C. Davis *Various averaging operations onto subalgebras.* Illinois J. Math., **3**, 528–553, (1959)

[9] C. Davis *Notions generalizing convexity for functions defined on spaces of matrices*, in *Convexity*, Proceedings of Symposia in Pure Mathematics, **VII**, ed. V. Klee, Amer. Math. Soc., Providence, 1963, pp 187–202

[10] J. Dixmier: *Formes linéaires sur un anneau d'opérateurs*, Bull. Soc. Math. France **81**, 1953, pp. 222–245

[11] H. Epstein: *Remarks on two theorems of E. Lieb*, Commun. Math. Phys. **31** 317-325 (1973)

[12] S. Golden: *Lower bounds for Helmholtz functions.* Phys. Rev. **137B**, 1965, pp. 1127-1128

[13] E. Heinz: *Beiträge zur Störungstheorie der Spektralzerlegung*, Math. Ann. **123**, 415–438 (1951)

[14] F. Hiai: *Concavity of certain matrix trace functions.* Taiwanese J. Math., **5**, no. 3, 535–554 (2001)

[15] F. Hansen: *The Wigner-Yanase entropy is not subadditive*, J. of Stat. Phys., **126**, No. 3, 643–648 (2007)

[16] G.H. Hardy, J.E. Littlewood and G. Pólya: *Inequalities.* Cambridge University Press, Cambridge, 1934

[17] T Kato: *Notes on some inequalities for linear operators*, Math. Ann., **125**, 208–212 (1952)

[18] E.H. Lieb: *Convex trace functions and the Wigner-Yanase-Dyson Conjecture*, Adv. Math. **11** 267-288 (1973)

[19] E.H. Lieb: *Some convexity and subadditivity properties of entropy*, Bull. Amer. Math. Soc. **81**, 1975, pp. 1-13.

[20] E.H. Lieb and M.B. Ruskai: *Proof of the strong subadditivity of quantum-mechanical entropy*, J. Math. Phys. **14**, 1973, pp. 1938-1941.

[21] E.H. Lieb and M.B. Ruskai: *Some operator inequalities of the Schwarz type*, Adv. in Math. **12**, 269-273 (1974).

[22] E.H. Lieb and W. Thirring, *Inequalities for the moments of the eigenvalues of the Schrödinger hamiltonian and their relation to Sobolev inequalities*, in *Studies in Mathematical Physics*, E.H. Lieb, B. Simon, A. Wightman eds., Princeton University Press, 269-303 (1976).

[23] E. Nelson: *Notes on non-commutative integration*, Jour. Funct. Analysis, **15**, 1974, pp. 103-116

[24] D. Petz: *A variational expression for the relative entropy*, Comm. Math. Phys., **114**, 1988, pp. 345–349

[25] D.W.. Robinson and D. Ruelle: *Mean entropy of states in classical statistical mechanics* Commun. Math. Phys **5**, 1967, pp.288-300

[26] T. Rockafellar: *Conjugate duality and optimization*, Vol. 16, *Regional conference series in applied mathematics*, SIAM, Philadelphia, 1974

[27] M.B. Ruskai, *Inequalities for quantum entropy: A review with conditions for equality*, J. Math. Phys. **43**, 2005, pp. 4358-4375 (2002). Erratum *ibid* **46**, pp. 019901

[28] I.E. Segal: *A non-commutative extension of abstract integration*, Annals of Math., **57**, 1953, pp. 401–457

[29] I.E. Segal.: *Tensor algebras over Hilbert spaces II*, Annals of Math., **63**, 1956, pp. 160–175

[30] I.E. Segal.: *Algebraic integration theory*, Bull. Am. Math. Soc.., **71**, 1965, no. 3, pp. 419-489

[31] B. Segal.: *Trace ideals and their applications*, London Math. Soc. Lecture Note Series, **35**, Cambridge University Press, Cambridge, 1979

[32] C. Thompson: *Inequality with application in statistical mechanics* J. Math. Phys. **6**, 1812–1813 (1965)

[33] W. Thirring: *A course in mathematical physics, Volume IV: quantum mechanics of large systems* Springer Verlag, New York, 1983

[34] A. Uhlmann: *Sätze über Dichtematrizen*, Wiss. Z. Karl-Marx Univ. Leipzig **20**, 633-53 (1971)

[35] A. Uhlmann: *Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory* Commun. Math. Phys., **54**, 1977, pp. 21-32

[36] H. Umegaki: *Conditional expectation in operator algebras I*, Tohoku Math. J., **6**, 1954, pp. 177-181

[37] J. von Neumann, *Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren*, Math. Ann. **102**, 370427 (1929)

[38] E.P. Wigner and M.M. Yanase *Information contents of distributions* Proc. Nat. Acad. Sci **49** 910–918 (1963)

[39] E.P. Wigner and M.M. Yanase *On the positive semidefinite nature of certain matrix expressions* Canad. J. Math., **16** 397–406 (1964)

[40] W.H.. Young: *On the multiplication of successions of Fourier constants*, Proc. Royal soc. A., **97**, 1912, pp. 331-339

[41] W.H.. Young: *Sur la généralization du théoreme du Parseval*, *Comptes rendus.*, **155**, 1912, pp. 30-33

Department of Mathematics, Hill Center, Rutgers University, 110 Frelinghuysen Road, Piscataway NJ 08854-8019 USA

E-mail address: carlen@math.rutgers.edu