# Trace Inequalities and Quantum Entropies

Melchior Wirth

Trace inequalities for quantum entropies and the related concavity/convexity trace functionals play a fundamental role in quantum information theory. They are also very important in other areas like mathematical physics and noncommutative analysis. Since Lieb's groundbreaking result resolving a conjecture of Wigner and Yanase, great progress has been made in the past half a century. We will give an introduction of several main results towards this direction.

Background: basic linear algebras. References:

1. Eric A. Carlen. Trace inequalities and quantum entropy: an introductory course.
2. Bhatia Rajendra. Matrix Analysis
3. Bhatia Rajendra. Positive Definite Matrices
4. Mark M. Wilde. From Classical to Quantum Shannon Theory.
5. Michael M. Wolf. Quantum Channels & Operations: Guided Tour
6. Anna Vershynina, Eric A. Carlen and Elliott H. Lieb. Strong Subadditivity of Quantum Entropy
7. Anna Vershynina, Eric A. Carlen and Elliott H. Lieb. Matrix and Operator Trace Inequalities
8. Eric A. Carlen. On some convexity and monotonicity inequalities of Elliott Lieb
9. John Watrous. The Theory of Quantum Information

# Contents

# Chapter 1

# Introduction and Notations

Classically, the states of a physical system are modeled as points in a structured set like a (smooth/Riemannian/Kähler/...) manifold and the observables, that is, the measurable quantities, as (smooth/continuous/...) real-valued functions on the state space. So if the system is in state $x$, the measurement outcome of observable $f$ will be $f(x)$.

More precisely, this is the setup of theories for one particle and the states represented by a single point are called pure states. If we move to statistical physics, then (mixed) states are more generally probability measures and the measurement outcome of observable $f$ for a system in state $\mu$ is $\int f \, d\mu$. Pure states correspond to Dirac measures $\delta_x$ in this picture.

One common feature in both of these cases is that the state of a physical system completely determines the measurement outcome of an observable. Early on in the development of quantum mechanics it was recognized that this paradigm is incompatible with observations made at the atomic level. At least that is the mainstream point of view, some people still try to develop quantum physics in such a way that it is a deterministic theory. But we will not concern ourselves with these approaches in these lectures.

Instead of a deterministic theory, in which the outcomes of a measurement are determined by the state of the system, quantum theory is at its heart a probabilistic theory. The state of a physical system only determines the probabilities of measurement outcomes and not their exact value.

In quantum mechanics (of closed quantum systems), the state of system is a unit vector $\xi$ from some Hilbert space $H$, which we take finite-dimensional for simplicity's sake here, so you can think of $H = \mathbb{C}^n$ if you want. An observable is modeled by a self-adjoint operator $A$ on $H$. If we choose a basis of $H$, then $A$ can be represented as a Hermitian matrix. By the spectral theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of eigenvectors of $A$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$ (counted with multiplicity). If the system is in the state $\xi$, the possible measurement outcomes for the observable $A$ are $\lambda_1, \ldots, \lambda_n$ with probabilities $|\langle \xi, e_1 \rangle|^2, \ldots, |\langle \xi, e_n \rangle|^2$, respectively.

Again, this is the setup for one particle and the corresponding quantum states are called pure states. If we move to quantum statistical mechanics, the (mixed) states are represented by density operators. A density operator $\rho$ is a self-adjoint operator such that all its eigenvalues are positive (i.e. non-negative) and whose trace is 1. The pure state $\xi$ corresponds to the density operator $\langle \xi, \cdot \, \xi \rangle$. In fact, this is more of a half-truth. To get the full picture, we should speak about open and closed quantum systems and how pure states are not sufficient to describe open systems. But for the purpose of this introduction, this analogy is a good guiding principle to understand the mathematics.

The possible measurement outcomes for the observable $A$ in state $\rho$ are still $\lambda_1, \ldots, \lambda_n$ with probabilities $\langle e_1, \rho e_1 \rangle, \ldots, \langle e_n, \rho e_n \rangle$. In particular, the expected value of $A$ in state $\rho$ is $\mathrm{Tr}(A\rho)$.

These probabilities behave differently from the probabilities in Kolmogorov's axiomatic setup usually studied in probability theory. In fact, this difference can be quantified in terms of so-called "Bell inequalities", which have been used to experimentally rule out a classical probabilistic interpretation of quantum mechanics.

One of the striking differences between classical and quantum observables is that multiplication of functions is commutative, while the multiplication of linear operators (or matrices, if you like) is not. A physical consequence of this fact is that some observables cannot be measured at the same time with arbitrary precision – a fact known as Heisenberg's uncertainty principle.

One of the most important aspects for quantum information theory is how we describe composite systems. In classical systems, if we have two systems with (pure) state spaces $X$ and $Y$, then the pure state space of the composite system is $X \times Y$. In other words, the state of the composite system is described by the state of the two parts.

In quantum physics, if the states of two systems $A$ and $B$ are described by Hilbert spaces $H_A$ and $H_B$, then the Hilbert space for the composite system is their tensor product $H_A \otimes H_B$. This is different from classical physics in that a (pure) state of the composite system is in general not described simply by a pair of pure states for $A$ and $B$. Mathematically speaking, a unit vector $\xi \in H_A \otimes H_B$ is not necessarily of the form $\xi_A \otimes \xi_B$ for unit vectors $\xi_A \in H_A$, $\xi_B \in H_B$. This leads to the phenomenon of entanglement, which is at the heart of many interesting quantum effects, both desirable and not.

So far, this is not much of a theory. We can describe the state of a system at a given time and the possible measurement outcomes, but nothing has been said how physical systems change with time. As is customary in quantum information theory, we will not concern ourselves with the continuous times evolution of quantum systems, which is governed by the Schrödinger equation. Instead we will content ourselves with describing possible changes a system can make.

It seems natural to describe the time evolution of a systems as its state evolving in time while the observables stay unchanged. Note however that the only thing one can really measure is the outcome of an observable in a given state, not the observable itself or the state itself. Thus it is just as valid to describe the time evolution of a physical system as the observables evolving in times while the states stay unchanged. These two standpoints are called Heisenberg (observables evolve in times) and Schrödinger (states evolve in time) picture.

So, if we want a linear map $\Phi\colon B(H) \to B(H)$ to describe the change of a system in the Schrödinger picture, it should map states to states, that is, density operators to density operators. If we break this down into the two parts of the definition of density operators, this means it should

- map positive operators to positive operators (positive map),
- preserve the trace of operators (trace-preserving map).

These two requirements are certainly enough for $\Phi$ to map density operators to density operators. But it is one of the interesting quirks of quantum information theory that these two conditions are not enough to ensure that $\Phi$ describes the change of the states of a quantum system. To see this, one has to look at composite systems.

If systems $A$ and $B$ are described by $H_A$ and $H_B$ and the states of $A$ change according to $\Phi$ while the states of $B$ stay unchanged, the states of $AB$ should change according to the map $\Phi \otimes \mathrm{id}_B$. So does $\Phi \otimes \mathrm{id}_B$ always map density matrices to density matrices? Surprisingly not. It is clearly trace-preserving, but it may fail to be positive. Maps $\Phi$ with the property that $\Phi \otimes \mathrm{id}_B$ is positive for arbitrary systems $B$ are called completely positive and will occur time and again during this course.

Similarly, the maps describing the change of the observables of a system in the Heisenberg picture are unital completely positive maps. Here a linear map $\Phi\colon B(H) \to B(H)$ is called unital if $\Phi(1) = 1$.

Now how do quantum entropies and trace inequalities enter the picture? Entropy is a concept that occurs in many different shapes and forms in physics, so much that it has been called a metaphor of nature itself. One basic physical principle is that physical systems seek to maximize their entropy (the second law of thermodynamics). This play a particularly important role in understanding the dissipative behavior of open quantum systems.

A closely related quantity is the relative entropy. States that maximize the ("absolute") entropy when the total energy of the system is fixed are equilibrium states, called Gibbs states in this context. The relative entropy of a state with respect to a Gibbs state then measures the deviation of this state from equilibrium. One way to express this is the following: If a state has relative entropy $d$ with respect to the Gibbs state, then it takes $\sim \log(d/\varepsilon)$ measurements to distinguish this state from equilibrium with precision $\varepsilon$.

Mathematically, the quantum entropy $S$ of a system in state $\rho$ and the relative $D$ entropy with respect to a Gibbs state $\sigma$ are expressed as

$$\begin{aligned} S(\rho) &= -\mathrm{Tr}(\rho \log \rho), \\ D(\rho\|\sigma) &= \mathrm{Tr}(\rho(\log \rho - \log \sigma)). \end{aligned}$$

We will soon make sense of expression like $\log \rho$ for (some) matrices $\rho$. Trace inequalities enter the picture to justify that these expressions have the expected physical properties. Maybe the most prominent example is the *data processing inequality*, which states that the relative entropy decreases when a quantum channel is applied to the state of the system. Mathematically, this is reflected in certain convexity and monotonicity properties of the relative entropy.

Since this is a mathematics course after all, let us finish the first lecture by recalling some basic linear algebra (or linear analysis, really).

**Definition 1.1.** An *inner product space* (or *Hilbert space*) is a vector space $H$ equipped with a map $\langle \cdot, \cdot \rangle \colon H \times H \to \mathbb{C}$ satisfying

- $\langle \xi, \alpha\eta + \beta\zeta \rangle = \alpha\langle \xi, \eta \rangle + \beta\langle \xi, \zeta \rangle$ for all $\xi, \eta, \zeta \in H$, $\alpha, \beta \in \mathbb{C}$,
- $\langle \eta, \xi \rangle = \overline{\langle \xi, \eta \rangle}$ for all $\xi, \eta \in H$,
- $\langle \xi, \xi \rangle \geq 0$ for all $\xi \in H$ with equality if and only if $\xi = 0$.

**Remark.** We take all vector spaces to be *finite-dimensional* and over the *complex numbers*, unless otherwise stated.

As we know, every (finite-dimensional complex) vector space is isomorphic to $\mathbb{C}^n$ for some $n \in \mathbb{N}$. To determine all inner products on $\mathbb{C}^n$, we recall the notion of positive (semi-) definite matrices.

**Definition 1.2.** A matrix $A \in M_n(\mathbb{C})$ is called *positive semidefinite* (or simply *positive*) if $\xi^H A \xi \geq 0$ for all $\xi \in \mathbb{C}^n$. It is called *positive definite* if $\xi^H A \xi > 0$ for all $\xi \in \mathbb{C}^n \setminus \{0\}$. The set of all positive semidefinite matrices in $M_n(\mathbb{C})$ is denoted by $M_n(\mathbb{C})_+$ and the subset of all positive definite matrices by $M_n(\mathbb{C})_{++}$.

For $A, B \in M_n(\mathbb{C})$ we write $A \leq B$ if $B - A$ is positive semidefinite. In particular, $A \geq 0$ means that $A$ is positive semidefinite.

**Lemma 1.3.** *For every inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{C}^n$ there exists a positive definite matrix $A \in M_n(\mathbb{C})$ such that $\langle \xi, \eta \rangle = \xi^H A \eta$ for all $\xi, \eta \in \mathbb{C}^n$.*

*Proof.* Let $e_1, \dots, e_n$ be the standard basis of $\mathbb{C}^n$ and let $A_{jk} = \langle e_j, e_k \rangle$. For $\xi, \eta \in H$ we have by sesquilinearity

$$\langle \xi, \eta \rangle = \left\langle \sum_{j=1}^n \xi_j e_j, \sum_{k=1}^n \eta_k e_k \right\rangle = \sum_{j,k=1}^n \overline{\xi_j} \eta_k \langle e_j, e_k \rangle = \sum_{j,k=1}^n \overline{\xi_j} A_{jk} \eta_k = \xi^H A \eta.$$

In particular,

$$\xi^H A \xi = \langle \xi, \xi \rangle,$$

which implies that $A$ is positive definite. □

**Lemma 1.4.** *If $H$ is an inner product space, there exists $n \in \mathbb{N}$ and a linear isomorphism $U \colon H \to \mathbb{C}^n$ such that*

$$(U\xi)^H (U\eta) = \langle \xi, \eta \rangle$$

*for all $\xi, \eta \in H$.*

*Proof.* By linear algebra, there exists a linear isomorphism $V \colon H \to \mathbb{C}^n$ for some $n \in \mathbb{N}$. By the previous lemma, there exists a positive definite matrix $A \in M_n(\mathbb{C})$ such that $\langle \xi, \eta \rangle = (V\xi)^H A V \eta$ for all $\xi, \eta \in H$. Since $A$ is positive definite, there exists an invertible matrix $B \in M_n(\mathbb{C})$ such that $A = B^H B$ (see Exercise 1.1). The map $U = BV$ does the job. □

In other words, there is essentially one inner product space of dimension $n$, namely $\mathbb{C}^n$ with the inner product $\langle \xi, \eta \rangle = \xi^H \eta$. From now one we will always consider $\mathbb{C}^n$ with this specific inner product, also called the standard inner product. Notice that it has the nice property that the standard basis satisfies $\langle e_j, e_k \rangle = \delta_{jk}$, in other words, it is an orthonormal basis.

**Lemma 1.5.** *Let $H, K$ be inner product spaces. For every linear map $A \colon H \to K$ there exists a unique linear map $A^* \colon K \to H$ such that*

$$\langle A\xi, \eta \rangle = \langle \xi, A^* \eta \rangle$$

*for all $\xi \in H$, $\eta \in K$.*

*Proof.* By the previous lemma we can assume without loss of generality $H = \mathbb{C}^m$, $K = \mathbb{C}^n$ (with the standard inner products). Then $A^* = A^H$ does the job. Uniqueness is easy to see. □

From now on we will write $A^*$ instead of $A^H$ because that's what the cool kids do.

**Definition 1.6.** Let $H$ be a Hilbert space. A linear map $A \colon H \to H$ is called *self-adjoint* if $A^* = A$.

If $H = \mathbb{C}^n$, then a self-adjoint linear map $A \colon H \to H$ can be identified with a hermitian matrix in $M_n(\mathbb{C})$, and vice versa. We will use these two viewpoints interchangeably. The hermitian matrices in $M_n(\mathbb{C})$ are denoted by $M_n(\mathbb{C})_{\mathrm{sa}}$.

**Theorem 1.7** (Spectral theorem)**.** *Let $H$ be a Hilbert space. Every self-adjoint $A \colon H \to H$ can be written in the form*

$$A = \sum_{j=1}^n \lambda_j \langle \xi_j, \cdot \rangle \xi_j$$

*with real numbers $\lambda_j$ and an orthonormal basis $\xi_1, \dots, \xi_n$ of $H$.*

*Proof.* We prove this by induction over the dimension of $H$. For $n = 1$ it is clear. Suppose we have proven it for $\dim(H) = n$ and let $\dim(H) = n + 1$.

By Lemma 1.4, we can assume that $H = \mathbb{C}^{n+1} \cong \mathbb{R}^{2n+2}$ with the standard inner product.

Consider the map

$$f\colon \mathbb{C}^{n+1} \to \mathbb{C},\ \xi \mapsto \langle \xi, A\xi\rangle.$$

Since $A$ is self-adjoint, we have $\overline{f(\xi)} = \langle A\xi, \xi\rangle = \langle \xi, A\xi\rangle = f(\xi)$. In other words, $f$ is real-valued. Thus $f$ attains its maximum on the sphere $S = \{\xi \in \mathbb{C}^{n+1} : \|\xi\|^2 = 1\}$. By the Lagrange multiplier theorem, a maximizer $\xi_1$ of $f$ on $S$ must satisfy

$$2A\xi_1 = \nabla_\xi \langle \xi, A\xi\rangle\Big|_{\xi=\xi_1} = \lambda_1 \nabla_\xi (\|\xi\|^2 - 1)\Big|_{\xi=\xi_1} = 2\lambda_1\xi_1.$$

for some $\lambda_1 \in \mathbb{R}$.

If $\xi \perp \xi_1$, then

$$\langle A\xi, \xi_1, \rangle = \langle \xi, A\xi_1\rangle = \lambda_1\langle \xi, \xi_1\rangle = 0.$$

Thus $A(V^\perp) \subset V^\perp$. The subspace $V^\perp$ has dimension $n$. By induction hypothesis, there exists an orthonormal basis $\xi_2, \dots, \xi_{n+1}$ of $V^\perp$ and real numbers $\lambda_2, \dots, \lambda_{n+1}$ such that

$$A\eta = \sum_{j=2}^{n+1} \lambda_j \langle v_j, \eta\rangle v_j$$

for $\eta \in V^\perp$.

Hence if $\xi = \langle \xi_1, \xi\rangle\xi_1 + \eta$ with $\eta \in V^\perp$, then

$$A\xi = \langle \xi_1, \xi\rangle A\xi_1 + \sum_{j=2}^{n+2} \lambda_j \langle \xi_j, \eta\rangle \xi_j = \sum_{j=1}^{n+1} \lambda_j \xi_j, \xi\rangle \xi_j. \qquad \square$$

**Remark.** The operators $P_j = \langle \xi_j, \cdot\rangle\xi_j$ are projections, that is, $P_j^2 = P_j^* = P_j$. Moreover, the orthogonality relation $\langle \xi_i, \xi_j\rangle = 0$ for $i \neq j$ implies that the projections $P_i$ are orthogonal in the sense that $P_iP_j = 0$ for $i \neq j$, and the completeness of an orthonormal basis implies that the projections $P_i$ sum up to 1.

This means that every self-adjoint operator $A$ can be written as

$$A = \sum_{j=1}^{m} \lambda_j P_j$$

with real numbers $\lambda_j$ and orthogonal projections $P_j$ such that $\sum_{j=1}^m P_j = 1$. Such a representation is called *spectral decomposition* of $A$. Usually one sums up the projections belonging to the same eigenvalue.

If $A$ has spectral decomposition $A = \sum_{j=1}^m \lambda_j P_j$, then the eigenvalues of $A$ are exactly the numbers $\lambda_1, \dots, \lambda_m$.

**Definition 1.8.** If $A \in M_n(\mathbb{C})$ is self-adjoint with spectral decomposition $A = \sum_{j=1}^m \lambda_j P_j$ and $f\colon \{\lambda_1, \dots, \lambda_m\} \to \mathbb{C}$, we define

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

This notation is consistent with the usual notations $A^k$ and $(A - \mu)^{-1}$.

**Lemma 1.9.** *Let $A \in M_n(\mathbb{C})$ be self-adjoint with spectral decomposition $A = \sum_{j=1}^m \lambda_j P_j$.*

*(a) If $f(\lambda) = \lambda^k$, then $f(A) = A^k$.*

*(b) If $\mu \notin \{\lambda_1, \ldots, \lambda_m\}$ and $f(\lambda) = (\lambda - \mu)^{-1}$, then $f(A) = (A - \mu)^{-1}$.*

*Proof.* (a) We proceed by induction over $k$. For $k = 0$, the claim is true. Now assume it is true for $k$ and let us prove it for $k + 1$. We have

$$\begin{aligned} A^{k+1} = A^k A &= \left( \sum_{i=1}^m \lambda_i^k P_i \right) \left( \sum_{j=1}^m \lambda_j P_j \right) \\ &= \sum_{i,j=1}^m \lambda_i^k \lambda_j P_i P_j. \end{aligned}$$

Since the maps $P_i$ are orthogonal projections, we have $P_i P_j = 0$ if $i \neq j$ and $P_i^2 = P_i$. Thus $A^{k+1} = \sum_{j=1}^m \lambda_j^{k+1} P_j = f(A)$.

(b) Since $\mu \notin \{\lambda_1, \ldots, \lambda_m\}$, the operator $A - \lambda$ is injective, hence also surjective by dimension considerations. This means that the inverse $(A - \mu)^{-1}$ exists. Moreover,

$$(A - \mu) f(A) = \left( \sum_{i=1}^m (\lambda_i - \mu) P_i \right) \left( \sum_{j=1}^m (\lambda_j - \mu)^{-1} P_j \right) = \sum_{i,j=1}^m (\lambda_i - \mu)(\lambda_j - \mu)^{-1} P_i P_j$$

Again, we can use the orthogonality relation of the projections $P_i$ to reduce the last term to $\sum_{j=1}^m (\lambda_j - \mu)^{-1}(\lambda_j - \mu) P_j = \sum_{j=1}^m P_j = 1$. It follows from the uniqueness of (right) inverses that $f(A) = (A - \mu)^{-1}$.

□

## Exercises

**Exercise 1.1.** For a matrix $A \in M_n(\mathbb{C})$, show that the following are equivalent:

(i) $A$ is positive semidefinite (positive definite).

(ii) $A = A^*$ and all eigenvalues of $A$ are nonnegative (strictly positive).

(iii) $A = B^* B$ for some (invertible) $B \in M_n(\mathbb{C})$.

*Proof.* (i) $\Longrightarrow$ (ii): This implication uses the very practical *polarization identity*:

$$\langle \xi, A\eta \rangle = \frac{1}{4} \sum_{k=0}^3 (-i)^k \langle \xi + i^k \eta, A(\xi + i^k \eta) \rangle.$$

To prove it, you just have to sit down and expand the inner product. Not very pleasant, but it works.

Since $A$ is positive, all summands on the right side are real. Thus

$$\begin{aligned}\frac{1}{4}\sum_{k=0}^{3}(-i)^k\langle\xi+i^k\eta, A(\xi+i^k\eta)\rangle &= \frac{1}{4}\sum_{k=0}^{3}(-i)^k\overline{\langle\xi+i^k\eta, A(\xi+i^k\eta)\rangle}\\ &= \frac{1}{4}\sum_{k=0}^{3}(-i)^k\langle A(\xi+i^k\eta), \xi+i^k\eta\rangle\end{aligned}$$

Then we can apply the polarization identity again to get

$$\frac{1}{4}\sum_{k=0}^{3}(-i)^k\langle A(\xi+i^k\eta), \xi+i^k\eta\rangle = \langle A\xi, \eta\rangle = \langle\xi, A^*\eta\rangle.$$

Hence, $A = A^*$.

If $\xi$ is an eigenvector of $A$ to the eigenvalue $\lambda$, then $\langle\xi, A\xi\rangle = \lambda\|\xi\|^2$. Hence $\lambda \geq 0$ (resp. $\lambda > 0$) if $A$ is positive semi-definite (resp. positive definite).

(ii) $\Longrightarrow$ (iii): Let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}_+$ denote the eigenvalues of $A$. By the spectral theorem, there exist orthogonal projections $P_1, \ldots, P_n$ such that

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

Let $B = \sum_{j=1}^{n}\sqrt{\lambda_j}P_j = \sqrt{B}$. Then $B = B^*$ and $B^2 = A$. If $A$ the eigenvalues of $A$ are strictly positive, then $B$ is invertible with inverse $B^{-1} = \sum_j \lambda_j^{-1}P_j$.

(iii) $\Longrightarrow$ (i): If $A = B^*B$, then

$$\langle\xi, A\xi\rangle = \langle\xi, B^*B\xi\rangle = \langle B\xi, B\xi\rangle \geq 0$$

for all $\xi \in \mathbb{C}^n$. Thus $A$ is positive semi-definite.

If $B$ is invertible and $\xi \neq 0$, then $B\xi \neq 0$, hence $\langle B\xi, B\xi\rangle > 0$. □

**Exercise 1.2.** Show that for every $A \in M_n(\mathbb{C})$ there exist positive semidefinite matrices $A_1, \ldots, A_4 \in M_n(\mathbb{C})$ such that $A = A_1 - A_2 + i(A_3 - A_4)$.

**Exercise 1.3.** For matrices $A, B \in M_n(\mathbb{C})$ define their Hadamard product $A \circ B$ as the matrix with entries $A_{j,k}B_{j,k}$. Show that the Hadamard product of two positive semi-definite matrices is again positive semi-definite.

**Exercise 1.4.** The absolute value $|A|$ of a matrix $A \in M_n(\mathbb{C})$ is defined as $|A| = (A^*A)^{1/2}$.

(a) Show that for every $A \in M_n(\mathbb{C})$ there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that $A = U|A|$.

(b) Let $A, B \in M_n(\mathbb{C})$ be self-adjoint with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and $\mu_1 \leq \cdots \leq \mu_n$, respectively. Show that if $\lambda_k \leq \mu_k$ for all $k \in \{1, \ldots, n\}$, then there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that $A \leq U^*BU$.

(c) Show that for every $A \in M_n(\mathbb{C})$ there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that $\frac{1}{2}(A + A^*)_+ \leq U^*|A|U$ (Hint: Use the minmax principle.)

(d) Show that for all $A, B \in M_n(\mathbb{C})$ there exist unitary matrices $U, V \in M_n(\mathbb{C})$ such that $|A + B| \leq U^*|A|U + V^*|B|V$.

(e) Show that there exist $A, B \in M_n(\mathbb{C})$ such that $|A+B| \not\leq |A| + |B|$.

**Exercise 1.5.** (a) Show that if $\varphi\colon M_n(\mathbb{C}) \to \mathbb{C}$ is a linear map such that $\varphi(AB) = \varphi(BA)$ for all $A, B \in M_n(\mathbb{C})$, then $\varphi = \frac{\varphi(1)}{n}\mathrm{Tr}$.

(b) Let $H$ be an infinite-dimensional Hilbert space and $B(H)$ the set of all bounded linear operators on $H$. Show that if $\varphi\colon B(H) \to \mathbb{C}$ is a linear map such that $\varphi(AB) = \varphi(BA)$ for all $A, B \in B(H)$, then $\varphi = 0$ (harder).

**Exercise 1.6** (Schur complement theorem)**.** Let $A \in M_n(\mathbb{C})$ be invertible, $B \in M_{n,m}(\mathbb{C})$ and $C \in M_m(\mathbb{C})$. Show that the Block matrix

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

is positive if and only if $A \geq 0$, $C \geq 0$ and $C - B^*A^{-1}B \geq 0$.

**Exercise 1.7.** (a) Let $V$ be a subspace of $M_n(\mathbb{C})$ such that $ABC \in V$ for all $A, C \in M_n(\mathbb{C})$ and $B \in V$. Show that $V = \{0\}$ or $V = M_n(\mathbb{C})$.

(b) Show that if $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is a linear map such that $\Phi(AB) = \Phi(A)\Phi(B)$ for all $A, B \in \mathbb{R}$, then either $\Phi = 0$ or $\Phi$ is injective.

# Chapter 2

# Complete positivity

**Definition 2.1** (Tensor product of vector spaces/Hilbert spaces)**.** For Hilbert spaces $H$ and $K$ let $\mathrm{Bil}(H \times K; \mathbb{C})$ be the vector space of all sesquilinear maps from $H \times K$ to $\mathbb{C}$. For $\xi \in H$ and $\eta \in K$ define

$$\xi \otimes \eta \colon \mathrm{Bil}(H \times K; \mathbb{C}) \to \mathbb{C},\ \varphi \mapsto \varphi(\xi, \eta).$$

The tensor product $H \otimes K$ is the linear span of all elements $\xi \otimes \eta$ with $\xi \in H$ and $\eta \in K$. It is a Hilbert space when endowed with the inner product

$$\langle \xi_1 \otimes \eta_1, \xi_2 \otimes \eta_2 \rangle = \langle \xi_1, \xi_2 \rangle \langle \eta_1, \eta_2 \rangle.$$

**Remark.** We already know that $\mathbb{C}^m \otimes \mathbb{C}^n$ must be isomorphic (as inner product space) to $\mathbb{C}^k$ for some $k \in \mathbb{N}$. It is not hard to see that the elementary tensors $e_i \otimes e_j$ with $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$ form an orthonormal basis of $\mathbb{C}^m \otimes \mathbb{C}^n$. Thus $\mathbb{C}^m \otimes \mathbb{C}^n \cong \mathbb{C}^{mn}$ as inner product spaces.

**Definition 2.2** (Tensor product of matrices/maps)**.** If $\Phi \colon H_1 \to K_1$ and $\Psi \colon H_2 \to K_2$ are linear maps, then their tensor product $\Phi \otimes \Psi$ is the linear map from $H_1 \otimes H_2$ to $K_1 \otimes K_2$, defined on elementary tensors by

$$(\Phi \otimes \Psi)(\xi \otimes \eta) = \Phi(\xi) \otimes \Psi(\eta).$$

The linear span of all elements $\Phi \otimes \Psi$ with $\Phi \in M_{m,k}(\mathbb{C})$ and $\Psi \in M_{n,l}(\mathbb{C})$ is denoted by $M_{m,k}(\mathbb{C}) \otimes M_{n,l}(\mathbb{C})$.

**Remark.** We will always identify elements of $M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ with $mn \times mn$ matrices in the following way: Let $(E_{ij})$ be the matrix units in $M_n(\mathbb{C})$, that is, $E_{ij}$ is the matrix whose $(i,j)$-entry is 1 and all other entries are 0. The matrix $A \otimes E_{ij}$ is identified with the block matrix in $M_{mn}(\mathbb{C})$ with blocks of size $m \times m$, where the block at position $(i,j)$ is $A$ and all other blocks are zero. Here is an example:

$$A \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 & A \\ 0 & 0 \end{pmatrix}.$$

Since the matrix units form a basis of $M_n(\mathbb{C})$, this identification can be linearly extended to all of $M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$. For example,

$$A \otimes \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} aA & bA \\ cA & dA \end{pmatrix}.$$

In other words, the elementary tensor $A \otimes B$ is identified with the Kronecker product of $A$ and $B$ (which is also denoted by $A \otimes B$ for this reason).

**Definition 2.3** (Completely positive maps and quantum channels)**.** A linear map $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is called *positive* if it maps positive semi-definite matrices to positive semi-definite matrices. For $k \geq 1$, a linear map $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is said to be *k-positive* if $\Phi \otimes \mathrm{id}_k\colon M_m(\mathbb{C}) \otimes M_k(\mathbb{C}) \to M_n(\mathbb{C}) \otimes M_k(\mathbb{C})$ is positive. It is said to be *completely positive* if it is $k$-positive for any $k \geq 1$.

In general, characterizing $k$-positive maps from $M_m(\mathbb{C})$ to $M_n(\mathbb{C})$ is a hard task. The situation is much better for completely positive maps. Let us start with a few (non-) examples.

**Example 2.4.** The transpose map $T(A) = A^T$ is positive but not 2-positive (exercise).

**Example 2.5.** The depolarizing channel $\Phi(A) = \lambda A + (1-\lambda)\mathrm{Tr}(A)1$ for $\lambda \in [0,1]$ is completely positive.

**Example 2.6.** The following maps are completely positive:

1. If $V \in M_{m,n}(\mathbb{C})$, then the map $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C}), A \mapsto V^*AV$ is completely positive.
2. $*$-homomorphism $\pi$, that is, a linear map $\pi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ such that $\pi(AB) = \pi(A)\pi(B)$ and $\pi(A^*) = \pi(A)^*$ for all $A, B \in M_m(\mathbb{C})$. For a more concrete example, take
$$\pi\colon M_n(\mathbb{C}) \to M_{2n}(\mathbb{C}),\ A \mapsto \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}.$$

We will soon see that all completely positive maps can be constructed from these two examples.

**Lemma 2.7.** *A linear map $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is completely positive if and only if for every $N \in \mathbb{N}$, all $A_1, \ldots, A_N \in M_m(\mathbb{C})$ and $\xi_1, \ldots, \xi_N \in \mathbb{C}^n$ we have*
$$\sum_{j,k=1}^{l} \langle \xi_j, \Phi(A_j^* A_k)\xi_k \rangle \geq 0.$$

*Proof.* First assume that $\Phi$ is completely positive. We have
$$\begin{aligned}
\sum_{j,k=N}^{l} \langle \xi_j, \Phi(A_j^* A_k)\xi_k \rangle &= \sum_{i,j,k,l=1}^{N} \langle \xi_i \delta_{i,j}, \Phi(A_j^* A_k)\xi_l \delta_{k,l} \rangle \\
&= \sum_{i,j,k,l=1}^{N} \langle \xi_i, \Phi(A_j^* A_k)\xi_l \rangle \langle e_i, E_{jk} e_l \rangle \\
&= \sum_{i,j,k,l=1}^{N} \langle \xi_i \otimes e_i, (\Phi \otimes \mathrm{id})(A_j^* A_k \otimes E_{jk})(\xi_l \otimes e_l) \rangle \\
&= \left\langle \sum_{i=1}^{N} \xi_i \otimes e_i, (\Phi \otimes \mathrm{id}) \left( \left( \sum_{j=1}^{N} A_j \otimes E_{1j} \right)^* \left( \sum_{k=1}^{N} A_k \otimes E_{1k} \right) \right) \sum_{l=1}^{N} \xi_l \otimes e_l \right\rangle.
\end{aligned}$$
Since $\left(\sum_{j=1}^N A_j \otimes E_{1j}\right)^* \left(\sum_{k=1}^N A_k \otimes E_{1k}\right)$ is positive and $\Phi$ is completely positive, the last expression is nonnegative.

Conversely, any positive element of $M_m(\mathbb{C}) \otimes M_N(\mathbb{C})$ is of the form
$$\left( \sum_{i,j}^{N} B_{ij} \otimes E_{ij} \right)^* \left( \sum_{i,j}^{N} B_{ij} \otimes E_{ij} \right) = \sum_{i,j=1}^{N} A_i^* A_j \otimes E_{ij}$$

with $A_i = \sum_{k=1}^N B_{ki}$. Moreover, any $\xi \in M_n(\mathbb{C}) \otimes M_N(\mathbb{C})$ is of the form

$$\xi = \sum_{j=1}^N \xi_j \otimes e_j$$

with $\xi_j \in \mathbb{C}^n$. Since

$$\left\langle \xi, (\Phi \otimes \mathrm{id}_{M_N(\mathbb{C})}) \left( \sum_{i,j=1}^N A_i^* A_j \otimes E_{ij} \right) \xi \right\rangle = \sum_{i,j=1}^N \langle \xi_i, \Phi(A_i^* A_j) \xi_j \rangle,$$

the map $\Phi \otimes \mathrm{id}_{M_N(\mathbb{C})}$ is positive if and only if the inequality from the lemma holds. □

**Theorem 2.8** (Stinespring's dilation theorem)**.** *Any completely positive map $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ can be represented as*

$$\Phi(A) = V^* \pi(A) V$$

*with $V \in M_{k,n}(\mathbb{C})$ and a unital $*$-homorphism $\pi\colon M_m(\mathbb{C}) \to M_k(\mathbb{C})$ for some $k \in \mathbb{N}$.*

*Proof.* The proof is reminiscent of the GNS construction, and in fact, it can be understood as a generalization of it. On $M_m(\mathbb{C}) \otimes \mathbb{C}^n$ define

$$\left\langle \sum_j A_j \otimes \xi_j, \sum_k B_k \otimes \eta_k \right\rangle_K = \sum_{j,k} \langle \xi_j, \Phi(A_j^* B_k) \eta_k \rangle.$$

This map is clearly sesquilinear, and by the previous lemma it is also positive semi-definite. It may fail to be non-degenerate, so we define $K$ as the quotient of $M_m(\mathbb{C}) \otimes \mathbb{C}^n$ by the kernel of $\langle \cdot, \cdot \rangle_K$. We write $X \otimes_K \xi$ for the image of $A \otimes \xi$ in $K$ under the quotient map.

Then

$$\left\langle \sum_j A_j \otimes_K \xi_j, \sum_k B_k \otimes_K \eta_k \right\rangle := \left\langle \sum_j A_j \otimes \xi_j, \sum_k B_k \otimes \eta_k \right\rangle_K$$

defines an inner product on $K$, making it a Hilbert space. Clearly, $K$ is finite-dimensional, so that $K \cong \mathbb{C}^k$ for some $k \in \mathbb{N}$.

Define $V\colon \mathbb{C}^n \to K$, $\xi \mapsto \mathbf{1} \otimes_K \xi$ and $\pi\colon M_m(\mathbb{C}) \to B(K)$, $\pi(A)(B \otimes_K \xi) = AB \otimes_K \xi$. To show that $\pi(A)$ is well-defined, first note that $A^*A \leq \lambda 1$, where $\lambda$ is the largest eigenvalue of $A^*A$. Thus there exists $C$ such that $\lambda 1 - A^*A = C^*$. Hence

$$\begin{aligned} \lambda \sum_{j,k} \langle \xi_j, \Phi(B_j^* B_k) \xi_k \rangle - \sum_{j,k} \langle \xi_j, \Phi(B_j^* A^* A B_k) \xi_k \rangle &= \sum_{j,k} \langle \xi_j, \Phi(B_j^* (\lambda 1 - A^* A) B_k) \xi_k \rangle \\ &= \sum_{j,k} \langle \xi_j, \Phi(B_j^* C^* C B_k) \xi_k \rangle \\ &\geq 0 \end{aligned}$$

by the previous lemma. In particular, if $\sum_j B_j \otimes_K \xi_j = 0$, then $\sum_j AB_j \otimes_K \xi_j = 0$.

Let us compute the adjoint of $V$. For $\xi, \eta \in \mathbb{C}^n$ and $A \in M_m(\mathbb{C})$ we have

$$\langle \xi, V^*(A \otimes_K \eta) \rangle = \langle V\xi, A \otimes_K \eta \rangle = \langle 1 \otimes_K \xi, A \otimes_K \eta \rangle = \langle \xi, \Phi(A) \eta \rangle.$$

Hence $V^*(A \otimes_K \eta) = \Phi(A)\eta$ and we conclude that

$$V^* \pi(A) V \xi = V^* \pi(A)(\mathbf{1} \otimes_K \xi) = V^*(A \otimes_K \xi) = \Phi(A)\xi.$$

□

**Remark.** The construction in the proof is essentially forced upon us by the statement of the Stinespring dilation theorem: Let us assume there is a (finite-dimensional) Hilbert space $K$, a linear map $V\colon \mathbb{C}^n \to K$ and a unital $*$-homomorphism $\pi\colon M_m(\mathbb{C}) \to B(K)$ such that

$$\Phi(A) = V^*\pi(A)V$$

for all $A \in M_m(\mathbb{C})$.

Without loss of generality we may assume that elements of the form $\pi(A)V\xi$ with $A \in M_m(\mathbb{C})$ and $\xi \in \mathbb{C}^n$ linearly span $K$. Since the map $(A, \xi) \mapsto \pi(A)V\xi$ is bilinear, there exists a surjection $q\colon M_m(\mathbb{C}) \otimes \mathbb{C}^n \to K$ such that $q(A \otimes \xi) = \pi(A)V\xi$. In particular, $K$ is a quotient of $M_m(\mathbb{C}) \otimes \mathbb{C}^n$. Moreover, the inner product on $K$ must satisfy

$$\langle \pi(A)V\xi, \pi(B)V\eta\rangle = \langle \xi, V^*\pi(A^*B)V\eta\rangle = \langle \xi, \Phi(A^*B)\eta\rangle.$$

If we pull this back to $M_m(\mathbb{C}) \otimes \mathbb{C}^n$ via

$$\langle A \otimes \xi, B \otimes \eta\rangle := \langle q(A \otimes \xi), q(B \otimes \eta)\rangle,$$

one gets exactly the sesquilinear form from the proof.

For the following result recall the definition of the operator norm: If $A$ is a linear operator between the Hilbert spaces $H$ and $K$, then

$$\|A\| = \sup_{\substack{\xi \in H \\ \langle \xi, \xi\rangle \leq 1}} \langle A\xi, A\xi\rangle^{1/2}.$$

**Theorem 2.9** (Kadison–Schwarz inequality)**.** *If $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is a completely positive map, then we have*

$$\Phi(A)^*\Phi(A) \leq \|\Phi(\mathbf{1})\|\Phi(A^*A).$$

*Proof.* By Stinespring's dilation theorem there exists $V \in M_{k,n}(\mathbb{C})$ and a unital $*$-homomorphism $\pi\colon M_m(\mathbb{C}) \to M_k(\mathbb{C})$ such that

$$\Phi(A) = V^*\pi(A)V$$

for all $A \in M_m(\mathbb{C})$. Thus

$$\begin{aligned}\Phi(A)^*\Phi(A) &= V^*\pi(A)^*VV^*\pi(A)V \\ &\leq \|VV^*\|V^*\pi(A)^*\pi(A)V \\ &= \|VV^*\|\Phi(A^*A).\end{aligned}$$

Now it suffices to notice that

$$\|VV^*\| = \|V^*V\| = \|\Phi(1)\|. \qquad \square$$

**Remark.** In fact, the Kadison–Schwarz inequality holds more generally for 2-positive maps (exercise).

**Lemma 2.10.** *Let $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ be a linear map. If the Choi matrix*

$$C_\Phi := (\Phi \otimes \mathrm{id}_{M_m(\mathbb{C})})\left(\sum_{i,j=1}^m E_{ij} \otimes E_{ij}\right) \in M_{mn}(\mathbb{C})$$

*is positive, then there exist $V_1, \ldots, V_k \in M_{m,n}(\mathbb{C})$ such that*

$$\Phi(A) = \sum_{l=1}^k V_l^* A V_l.$$

*Proof.* Since $C_\Phi$ is positive, there exists $B \in M_{mn}(\mathbb{C})$ such that $B^*B = C_\Phi$. Let $b_1, \dots, b_{mn} \in \mathbb{C}^{mn}$ be the row vectors of $B$, so that

$$C_\Phi = B^*B = \sum_{l=1}^{mn} b_l^* b_l.$$

Further let

$$J_l \colon \mathbb{C}^n \to \mathbb{C}^n \otimes \mathbb{C}^m,\ \xi \mapsto \xi \otimes e_l$$

Note that

$$J_k^* C_\Phi J_l \xi = \sum_{i,j=1}^{m} J_k^*(\Phi(E_{ij})\xi \otimes E_{ij} e_l) = \sum_{i,j=1}^{m} \delta_{jl}\delta_{ik}\Phi(E_{ij})\xi = \Phi(E_{kl})\xi.$$

Let $V_l$ be the matrix with rows $b_l J_1, \dots, b_l J_m$. A direct calculation shows

$$\sum_{l=1}^{mn} V_l^* E_{ij} V_l = \Phi(E_{ij}),$$

from which the claim follows by linearity. □

**Remark.** As we have seen in the proof, one can always take $k = mn$ in the previous lemma. However, the minimal number of $V_l$ may be smaller. For example, some rows of $B$ in the proof may be zero.

**Theorem 2.11** (Kraus, Choi)**.** *Any $m$-positive map $\Phi \colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ can be represented as*

$$\Phi(A) = \sum_{j=1}^{k} V_j^* A V_j.$$

*Furthermore, $\Phi$ is unital if and only if*

$$\sum_{j=1}^{k} V_j^* V_j = \mathbf{1}$$

*and trace-preserving if and only if*

$$\sum_{j=1}^{k} V_j V_j^* = \mathbf{1}.$$

*Proof.* Since

$$\left(\sum_{i,j=1}^{m} E_{ij} \otimes E_{ij}\right)^* \left(\sum_{i,j=1}^{m} E_{ij} \otimes E_{ij}\right) = m \sum_{i,j=1}^{m} E_{ij} \otimes E_{ij},$$

the matrix $\sum_{i,j} E_{ij} \otimes E_{ij}$ is positive. As $\Phi$ is assumed to be $m$-positive, this implies that the Choi matrix $C_\Phi$ is positive. Now the first claim follows from the previous lemma.

For the second claim observe that

$$\operatorname{Tr}(\Phi(A)) = \sum_{j=1}^{k} \operatorname{Tr}(V_j^* A V_j) = \sum_{j=1}^{k} \operatorname{Tr}(A V_j V_j^*),$$

so that $\Phi$ is trace-preserving if and only if $\sum_j V_j V_j^* = \mathbf{1}$. The claim for unital maps is immediate. □

**Theorem 2.12** (Choi's criterion of completely positive maps)**.** *Let $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ be a linear map. The following are equivalent:*

*(i)* $\Phi$ *is $m$-positive.*

*(ii) The Choi matrix*

$$C_\Phi := (\Phi \otimes \mathrm{id}_{M_m(\mathbb{C})}) \left( \sum_{i,j=1}^{m} E_{ij} \otimes E_{ij} \right) \in M_{mn}(\mathbb{C})$$

*is positive, where $E_{ij}, 1 \leq i, j \leq m$ are the matrix units.*

*(iii)* $\Phi$ *is completely positive.*

*Proof.* (i) $\implies$ (ii) was shown in the proof of the previous theorem. (ii) $\implies$ (iii): If the Choi matrix is positive, then $\Phi(A) = \sum_l V_l^* A V_l$ by Lemma 2.10. It is easy to see that maps of this form are completely positive. (iii) $\implies$ (i) is obvious. □

**Theorem 2.13** (Uhlmann, Lindblad)**.** *For any quantum channel $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$, there exist $N > 0$ and a pure state $\delta \in \mathcal{M}_N$ such that*

$$\Phi(A) \otimes \frac{\mathbf{1}_N}{N} = \int U^*(A \otimes \delta) U \, dU,$$

*where $dU$ is the Haar measure on the unitary group.*

## Exercises

**Exercise 2.1.** Show that $\mathbb{C}^k \otimes \mathbb{C}^m$ has the following universal property: For every bilinear map $\varphi\colon \mathbb{C}^k \times \mathbb{C}^m \to \mathbb{C}^n$ there exists a unique linear map $\Phi\colon \mathbb{C}^k \otimes \mathbb{C}^m \to \mathbb{C}^n$ such that $\varphi(x, y) = \Phi(x \otimes y)$.

**Exercise 2.2.** For each $n \geq 1$, find maps that are $n$-positive but not $(n+1)$-positive.

**Exercise 2.3.** Show that the maps from Example 2.6 are completely positive.

**Exercise 2.4** (Completely positive maps are completely bounded)**.** Recall that the operator norm of $A \in M_n(\mathbb{C})$ is the square root of the largest eigenvalue of $A^*A$. Accordingly, the norm of a linear map $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is

$$\|\Phi\| = \sup_{\|A\|=1} \|\Phi(A)\|.$$

Show that if $\Phi$ is completely positive, then $\|\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})}\| \leq \|\Phi\|$ for all $k \in \mathbb{N}$.

**Exercise 2.5** (GNS construction for states)**.** (a) Show that any positive linear functional $\varphi\colon M_n(\mathbb{C}) \to \mathbb{C}$ is completely positive.

(b) A unital positive linear map $\varphi\colon M_n(\mathbb{C}) \to \mathbb{C}$ is called a state. Show that for every state $\varphi$ there exists a unital $*$-homomorphism $\pi\colon M_n(\mathbb{C}) \to M_k(\mathbb{C})$ for some $k \in \mathbb{N}$ and a unit vector $\xi \in \mathbb{C}^k$ such that

$$\varphi(A) = \langle \xi, \pi(A)\xi \rangle$$

for all $A \in M_n(\mathbb{C})$.

**Exercise 2.6.** Show that every 2-positive map satisfies the Kadison–Schwarz inequality.

**Exercise 2.7** (Hilbert modules I)**.** Recall that a right $M_n(\mathbb{C})$-module is a vector space $E$ together with an associative and distributive product $E \times M_n(\mathbb{C}) \to E$. A right Hilbert $M_n(\mathbb{C})$-module is a right $M_n(\mathbb{C})$-module together with a sesquilinear map

$$(\cdot|\cdot)\colon E \times E \to M_n(\mathbb{C})$$

such that

- $(\xi|\eta A) = (\xi|\eta)A$ for all $\xi, \eta \in E$, $A \in M_n(\mathbb{C})$,
- $(\eta|\xi) = (\xi|\eta)^*$ for all $\xi, \eta \in M_n(\mathbb{C})$,
- $(\xi|\xi)$ is positive semidefinite for all $\xi \in E$,
- $(\xi|\xi) = 0$ if and only if $\xi = 0$.

(a) Show that $M_{m,n}(\mathbb{C})$ with the usual right multiplication and $(A|B) = A^*B$ is a right Hilbert $M_n(\mathbb{C})$-module.

(b) (maybe not so easy) Show that for every (finite-dimensional) right Hilbert $M_n(\mathbb{C})$-module $E$ there exists $m \in \mathbb{N}$ and a linear isomorphism $\alpha\colon E \to M_{m,n}(\mathbb{C})$ such that

- $\alpha(\xi A) = \alpha(\xi)A$ for all $\xi \in E$, $A \in M_n(\mathbb{C})$,
- $\alpha(\xi)^*\alpha(\eta) = (\xi|\eta)$ for all $\xi, \eta \in M_n(\mathbb{C})$.

**Exercise 2.8** (Hilbert modules II)**.** Let $E, F$ be (finite-dimensional) right Hilbert $M_n(\mathbb{C})$-modules. A linear map $T\colon E \to F$ is called adjointable if there exists a linear map $T^*\colon F \to E$ such that

$$(T\xi|\eta) = (\xi|T^*\eta)$$

for all $\xi \in E$, $\eta \in F$. The set of all adjointable operators from $E$ to $F$ is denoted by $\mathcal{L}(E, F)$.

(a) Show that a linear operator $T\colon E \to F$ is adjointable if and only if

$$T(\xi A) = (T\xi)A$$

for all $\xi \in E$, $A \in M_n(\mathbb{C})$.

Hint: One direction is easy. For the other one use the inner product $\langle \xi, \eta \rangle = \operatorname{tr}((\xi|\eta))$ on $E$ and $F$.

(b) Show that every adjointable map $T\colon M_{m,n}(\mathbb{C}) \to M_{k,n}(\mathbb{C})$ is of the form

$$T(B) = AB$$

for some $A \in M_{k,m}(\mathbb{C})$.

**Exercise 2.9** (Hilbert modules III)**.** A Hilbert $M_m(\mathbb{C})$-$M_n(\mathbb{C})$-module is a right Hilbert $M_n(\mathbb{C})$-module $E$ together with a unital $*$-homomorphism $\pi\colon M_m(\mathbb{C}) \to \mathcal{L}(E, E)$.

(a) Show that $M_{km,n}(\mathbb{C})$ with

$$\pi_L \colon M_m(\mathbb{C}) \to \mathcal{L}(M_{km,n}(\mathbb{C}), M_{km,n}(\mathbb{C})),\ \pi_L(A)B = \begin{pmatrix} A & & \\ & \ddots & \\ & & A \end{pmatrix} B$$

is a Hilbert $M_m(\mathbb{C})$-$M_n(\mathbb{C})$-module.

(b) Show that for every Hilbert $M_m(\mathbb{C})$-$M_n(\mathbb{C})$-module $(E, \pi)$ there exists $k \in \mathbb{N}$ and a linear isomorphism $\alpha \colon E \to M_{kn,n}(\mathbb{C})$ such that

- $\alpha(\xi B) = \alpha(\xi)B$ for all $\xi \in E$, $B \in M_n(\mathbb{C})$,
- $(\alpha(\xi)^*\alpha(\eta) = (\xi|\eta)$ for all $\xi, \eta \in E$,
- $\alpha \circ \pi(A) = \pi_L(A) \circ \alpha$ for all $A \in M_m(\mathbb{C})$.

(c) Show that for every completely positive map $\Phi \colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ there exists a Hilbert $M_m(\mathbb{C})$-$M_n(\mathbb{C})$-module $(E, \pi)$ and an adjointable operator $V \colon M_n(\mathbb{C}) \to E$ such that

$$\Phi(A) = V^*\pi(A)V$$

for all $A \in M_m(\mathbb{C})$.

(d) Deduce Theorem 2.11 from (b) and (c).

# Chapter 3

# Conditional expectations and partial trace

**Definition 3.1.** The *Hilbert–Schmidt inner product* on $M_n(\mathbb{C})$ is defined as

$$\langle \cdot, \cdot \rangle_{\mathrm{HS}} \colon M_n(\mathbb{C}) \times M_n(\mathbb{C}) \to \mathbb{C},\ (A, B) \mapsto \mathrm{Tr}(A^* B).$$

If $\Phi \colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is a linear map, we denote its adjoint with respect to the Hilbert–Schmidt inner product by $\Phi^\dagger$.

**Lemma 3.2.** *A matrix $A \in M_n(\mathbb{C})$ is positive if and only if $\langle A, B \rangle_{\mathrm{HS}} \geq 0$ for all $B \in M_n(\mathbb{C})$.*

*Proof.* If $A$ is positive, then

$$\langle A, B \rangle_{\mathrm{HS}} = \mathrm{Tr}(AB) = \mathrm{Tr}(A^{1/2} B A^{1/2}) \geq 0$$

for all $B \in M_n(\mathbb{C})_+$.

For the converse let $\xi \in \mathbb{C}^n$ and $B_\xi \eta = \langle \xi, \eta \rangle \eta$. Since $B_\xi^* B_\xi = B_\xi^2 = \|\xi\|^2 B_\xi$, the operator $B_\xi$ is positive. Moreover,

$$\langle A, B_\xi \rangle_{\mathrm{HS}} = \mathrm{Tr}(A^* B_\xi) = \langle A\xi, \xi \rangle.$$

Thus if $\langle A, B \rangle_{\mathrm{HS}} \geq 0$ for all $B \in M_n(\mathbb{C})_+$, then $\langle A\xi, \xi \rangle \geq 0$, that is, $A \geq 0$. □

**Remark.** By the Riesz representation theorem, every linear map $\varphi \colon M_n(\mathbb{C}) \to \mathbb{C}$ is of the form $\varphi = \mathrm{Tr}(B \cdot)$ for some $B \in M_n(\mathbb{C})$. By the previous lemma, this map is positive if and only if $B \geq 0$.

The Hilbert–Schmidt adjoint connects the Heisenberg and Schrödinger picture, as the following lemma shows.

**Lemma 3.3.** *A linear map $\Phi \colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is unital completely positive if and only if $\Phi^\dagger$ is completely positive trace-preserving.*

*Proof.* Since $(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})^\dagger = \Phi^\dagger \otimes \mathrm{id}_{M_k(\mathbb{C})}$, it suffices to show that $\Phi$ is unital positive if and only if $\Phi^\dagger$ is positive trace-preserving.

As

$$\langle \Phi(\mathbf{1}), A \rangle_{\mathrm{HS}} = \langle \mathbf{1}, \Phi^\dagger(A) \rangle_{\mathrm{HS}} = \mathrm{Tr}(\Phi^\dagger(A)),$$

we have $\Phi(\mathbf{1}) = \mathbf{1}$ if and only if $\mathrm{Tr}(\Phi^\dagger(A)) = \mathrm{Tr}(A)$ for all $A \in M_n(\mathbb{C})$.

That $\Phi^\dagger$ is positive if and only if $\Phi$ is positive is an easy consequence of the previous lemma. □

**Definition 3.4.** Let $\mathcal{M}$ be a subalgebra of $M_n(\mathbb{C})$ that is closed under taking adjoints and contains $\mathbf{1}$. Let $\iota_{\mathcal{M}} \colon \mathcal{M} \to M_n(\mathbb{C})$ be the inclusion map. The conditional expectation $E_{\mathcal{M}}$ onto $\mathcal{M}$ is $\iota_{\mathcal{M}} \circ \iota_{\mathcal{M}}^\dagger$.

**Proposition 3.5.** *The conditional expectation $E_{\mathcal{M}}$ has the following properties:*

*(a) It is idempotent, that is, $E_{\mathcal{M}}^2 = E_{\mathcal{M}}$.*

*(b) It is symmetric with respect to the Hilbert–Schmidt inner product, that is, $E_{\mathcal{M}}^\dagger = E_{\mathcal{M}}$.*

*(c) It is unital completely positive trace-preserving.*

*(d) It is a bimodule map, that is, $E_{\mathcal{M}}(ABC) = AE_{\mathcal{M}}(B)C$ for all $A, C \in \mathcal{M}$ and $B \in M_n(\mathbb{C})$.*

*Proof.* (a) By definition, $\iota_{\mathcal{M}}$ is an isometry, so that $\iota_{\mathcal{M}}^\dagger \iota_{\mathcal{M}} = \mathrm{id}_{\mathcal{M}}$ and hence

$$E_{\mathcal{M}}^2 = \iota_{\mathcal{M}} \iota_{\mathcal{M}}^\dagger \iota_{\mathcal{M}} \iota_{\mathcal{M}}^\dagger = \iota_{\mathcal{M}} \iota_{\mathcal{M}}^\dagger = E_{\mathcal{M}}$$

(b) Clear from the definition.

(c) This follows from the previous lemma.

(d) For $A, C \in \mathcal{M}$ let $L_A B = AB$, $R_C B = BC$. Clearly, $L_A$ and $R_C$ commute with $\iota_{\mathcal{M}}$. Taking adjoints yields the claim. □

**Remark.** It follows from (a) and (b) that $E_{\mathcal{M}}$ is the orthogonal projection onto $\mathcal{M}$ (with respect to the Hilbert–Schmidt inner product). This can be equivalently characterized by the following properties:

(i) $E_{\mathcal{M}}(B) \in \mathcal{M}$ and $\|B - E_{\mathcal{M}}(B)\| \leq \|A - E_{\mathcal{M}}(B)\|$ for all $A \in \mathcal{M}$ with equality if and only if $A = E_{\mathcal{M}}(B)$.

(ii) $E_{\mathcal{M}}(B) \in \mathcal{M}$ and $B - E_{\mathcal{M}}(B) \perp \mathcal{M}$.

**Example 3.6** (Trace)**.** If $\mathcal{M} = \mathbb{C}\mathbf{1}$, then $E_{\mathcal{M}}(A) = \mathrm{Tr}(A)\mathbf{1}$.

**Remark.** As remarked before, every positive linear map $\varphi \colon M_n(\mathbb{C}) \to \mathbb{C}$ is of the form $\varphi = \mathrm{Tr}(B^{1/2} \cdot B^{1/2})$ for some $B \in M_n(\mathbb{C})_+$. Since $A \mapsto B^{1/2} A B^{1/2}$ is completely positive and Tr is completely positive by the previous example, this implies that every positive map from $M_n(\mathbb{C})$ to $\mathbb{C}$ is completely positive.

**Example 3.7** (Restriction to the diagonal)**.** If $\mathcal{M}$ consists of all diagonal matrices in $M_n(\mathbb{C})$, then $E_{\mathcal{M}}(A) = \mathrm{diag}(A_{11}, \ldots, A_{nn})$.

**Example 3.8** (Hadamard product)**.** Let $\mathcal{M} \subset M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ be the subalgebra formed by all elements of the form

$$\sum_{j,k=1}^{n} A_{jk} E_{jk} \otimes E_{jk}$$

for $A \in M_n(\mathbb{C})$. Then

$$E_{\mathcal{M}}(A \otimes B) = \sum_{j,k=1}^{n} A_{jk} B_{jk} E_{jk} \otimes E_{jk}.$$

**Definition 3.9.** The *partial trace* $\mathrm{Tr}_1 \colon M_m(\mathbb{C}) \otimes M_n(\mathbb{C}) \to M_n(\mathbb{C})$ is the linear map given by

$$\mathrm{Tr}_1(A \otimes B) = \mathrm{Tr}(A)B.$$

Likewise, the partial trace $\mathrm{Tr}_2 \colon M_m(\mathbb{C}) \otimes M_n(\mathbb{C}) \to M_m(\mathbb{C})$ is given by

$$\mathrm{Tr}_2(A \otimes B) = \mathrm{Tr}(B)A.$$

**Example 3.10.** Let $\mathcal{M} = M_m(\mathbb{C}) \otimes \mathbf{1}$, $\mathcal{N} = \mathbf{1} \otimes M_n(\mathbb{C})$. The conditional expectations from $M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ onto $\mathcal{M}$ and $\mathcal{N}$, respectively, are given by $E_{\mathcal{M}}(X) = \frac{1}{n}\mathrm{Tr}_2(X) \otimes \mathbf{1}$ and $E_{\mathcal{N}}(X) = \frac{1}{m}\mathbf{1} \otimes \mathrm{Tr}_1(X)$. The factors $\frac{1}{n}$ and $\frac{1}{m}$ come from the fact that $\mathrm{Tr}(A \otimes \mathbf{1}) = n\mathrm{Tr}(A)$ and $\mathrm{Tr}(\mathbf{1} \otimes B) = m\mathrm{Tr}(B)$.

## Exercises

**Exercise 3.1.** Show again that the Hadamard product of two positive matrices is positive. (Hint: Use Example 3.8.)

**Exercise 3.2.** Let $\Phi : \mathcal{M}_m(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$ be a quantum channel. There exist a unitary $U \in M_k$ with $k = mn^2$ and a unit vector $\varphi \in \mathbb{C}^n \otimes \mathbb{C}^n$ such that

$$\Phi(\rho) = \mathrm{Tr}_E[U(\rho \otimes |\varphi\rangle \langle\varphi|)U^*],$$

where $\mathrm{Tr}_E$ is the partial trace over the first two factors of $\mathbb{C}^m \otimes \mathbb{C}^n \otimes \mathbb{C}^n$.

**Exercise 3.3.** Let $\mathrm{Tr}_1$ be the partial trace over the first tensor factor of $\mathbb{C}^m \otimes \mathbb{C}^n$. Show that for any $\rho$ over $\mathbb{C}^m \otimes \mathbb{C}^n$, we have

$$\frac{\mathbf{1}}{m} \otimes \mathrm{Tr}_1\rho = \int (u \otimes \mathbf{1})\rho(u^* \otimes \mathbf{1})du,$$

where $du$ denotes the normalized Haar measure on the unitary group over $\mathbb{C}^m$. Or, we have

$$\frac{\mathbf{1}}{m} \otimes \mathrm{Tr}_1\rho = \frac{1}{m^2}\sum_{j,k=1}^{m} (u_{jk} \otimes \mathbf{1})\rho(u_{jk}^* \otimes \mathbf{1}),$$

where $\{u_{jk}\}_{1 \le j,k \le m}$ denotes the discrete Heisenberg–Weyl group over $\mathbb{C}^m$, i.e.

$$u_{jk} = \sum_{l=1}^{m} \eta^{kl} |j + l\rangle \langle l|, \quad \eta := e^{\frac{2\pi i}{m}}.$$

# Chapter 4

# Quantum States

We begin with a piece of notation physicists are very fond of – the bra-ket notation. Let $H$ be a Hilbert space. Every vector $\xi \in H$ gives rise to a linear map from $\mathbb{C}$ to $H$ that maps $\lambda \in \mathbb{C}$ to $\lambda\xi \in H$. This linear map is denoted by $|\xi\rangle$. As $\mathbb{C}$ is a Hilbert space with the standard inner product, we can take the adjoint of $|\xi\rangle$, which is denoted by $\langle\xi|$.

Not only does every vector $\xi \in H$ give rise to a linear map from $\mathbb{C}$ to $H$, the converse is also true: If $\Phi\colon \mathbb{C} \to H$ is linear, then $\Phi = |\Phi(1)\rangle$. For this reason, we will not distinguish between elements of $H$ and linear maps from $\mathbb{C}$ to $H$. In particular, we identify elements of $\mathbb{C}$ with linear maps from $\mathbb{C}$ to $\mathbb{C}$ (in other words, we treat $1 \times 1$ matrices as complex numbers).

With these identifications, the bra-ket notation works very nicely. For example,

$$\langle\xi|\, 1\, |\eta\rangle = \langle\xi, \eta\rangle$$

and $|\xi\rangle\langle\eta|$ is the linear map that sends $\zeta$ to $\langle\eta, \zeta\rangle\xi$.

With this notation, we can write the spectral decomposition of a self-adjoint matrix $A$ with orthonormal basis $(\xi_j)$ consisting of eigenvectors and associated eigenvalues $(\lambda_j)$ as

$$A = \sum_{j=1}^{m} \lambda_j\, |\xi_j\rangle\langle\xi_j|\,.$$

**Definition 4.1.** A *density matrix* or *quantum state* is a positive semi-definite matrix $\rho \in M_n(\mathbb{C})$ with $\mathrm{Tr}(\rho) = 1$.

Since a density matrix $\rho$ is positive semi-definite, it has an orthonormal basis $(\xi_j)$ consisting of eigenvectors and the associated eigenvalues $(\lambda_j)$ are non-negative. Moreover, the condition $\mathrm{Tr}(\rho) = 1$ is equivalent to $\sum_j \lambda_j = 1$.

Thus, density matrices are exactly the matrices that can be expressed as

$$\rho = \sum_{j=1}^{n} \lambda_j\, |\xi_j\rangle\langle\xi_j|$$

with $\|\xi_j\| = 1$ and $\lambda_j \geq 0$, $\sum_j \lambda_j = 1$.

**Definition 4.2.** A quantum state of the form $|\xi\rangle\langle\xi|$ is called a *pure state*. Every other quantum state is called a *mixed state*.

From the previous discussion, we have the following result.

**Lemma 4.3.** *Every quantum state is a convex combination of pure states.*

There are several other ways to characterize pure states.

**Proposition 4.4.** *For a quantum state $\rho \in M_n(\mathbb{C})$, the following properties are equivalent:*

*(i)* *$\rho$ is a pure state.*

*(ii)* *If $\sigma \in M_n(\mathbb{C})$ is positive semi-definite and $\sigma \leq \rho$, then there exists $\lambda \geq 0$ such that $\sigma = \lambda\rho$.*

*(iii)* *If $\rho_1, \rho_2 \in M_n(\mathbb{C})$ are quantum states and $\mu \in (0,1)$ such that $\rho = \mu\rho_1 + (1-\mu)\rho_2$, then $\rho_1 = \rho_2 = \rho$.*

*Proof.* (i) $\implies$ (ii): Let $\xi_1 \in \mathbb{C}^n$ such that $\rho = |\xi_1\rangle\langle\xi_1|$ and complete it to an orthonormal basis $(\xi_j)$ of $\mathbb{C}^n$. Since $\sigma \leq \rho$, if $j \geq 2$, then

$$\|\sigma^{1/2}\xi_j\|^2 = \langle\xi_j, \sigma\xi_j\rangle \leq \langle\xi_j, \rho\xi_j\rangle = |\langle\xi_1, \xi_j\rangle|^2 = 0$$

and hence also $\sigma\xi_j = \sigma^{1/2}\sigma^{1/2}\xi_j = 0$. By symmetry of $\sigma$, if $j \geq 2$, then

$$\langle\xi_j, \sigma\xi_1\rangle = \langle\sigma\xi_j, \xi_1\rangle = 0.$$

Thus $\sigma\xi_1 = \langle\xi_1, \sigma\xi_1\rangle\xi_1$. If $\xi \in \mathbb{C}^n$ is arbitrary, then

$$\sigma\xi = \sigma\sum_{j=1}^{n}\langle\xi, \xi_j\rangle\xi_j = \langle\xi, \xi_1\rangle\langle\xi_1, \sigma\xi_1\rangle\xi_1 = \langle\xi_1, \sigma\xi_1\rangle\,|\xi_1\rangle\langle\xi_1|\,\xi.$$

Now the claim follows with $\lambda = \langle\xi_1, \sigma\xi_1\rangle$.

(ii) $\implies$ (iii): Since $\rho_1$, $\rho_2$ are positive semi-definite, we have $\mu\rho_1, (1-\mu)\rho_2 \leq \rho$. By (ii), there exist $\lambda_1, \lambda_2 \geq 0$ such that $\mu\rho_1 = \lambda_1\rho$ and $(1-\mu)\rho_2 = \lambda_2\rho$. Taking the trace on both sides, we obtain

$$\mu = \mu\mathrm{Tr}(\rho_1) = \lambda_1\mathrm{Tr}(\rho) = \lambda_1$$

and likewise $1 - \mu = \lambda_2$. Therefore $\rho_1 = \rho_2 = \rho$.

(iii) $\implies$ (i): As $\rho$ is a quantum state, there is an orthonormal basis $(\xi_j)$ and $\lambda_j \geq 0$ with $\sum_j \lambda_j = 1$ such that $\rho = \sum_j \lambda_j |\xi_j\rangle\langle\xi_j|$. If $\rho$ is not a pure state, there exist two indices $j$ for which $\lambda_j \neq 0$. Without loss of generality we may assume $\lambda_1, \lambda_2 \neq 0$. Let $\rho_1 = |\xi_1\rangle\langle\xi_1|$ and $\rho_2 = (1-\lambda_1)^{-1}\sum_{j=2}^{n}\lambda_j |\xi_j\rangle\langle\xi_j|$. Clearly $\rho_2$ is positive semi-definite and

$$\mathrm{Tr}(\rho_2) = (1-\lambda_1)^{-1}\sum_{j=2}^{n}\lambda_j = (1-\lambda_1)^{-1}(1-\lambda_1) = 1.$$

Thus $\rho_1$, $\rho_2$ are density matrices and $\lambda_1\rho_1 + \lambda_2\rho_2 = \rho$. It follows from (ii) that $\rho_1 = \rho_2 = \rho$, a contradiction. Thus $\rho$ must be a pure state. □

There is also a handy quantitative measure to decide if a state is pure or not.

**Lemma 4.5.** *Every quantum state $\rho$ satisfies $\mathrm{Tr}(\rho^2) \leq 1$ with equality if and only if $\rho$ is pure.*

*Proof.* Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\rho$, counted with multiplicity, and recall that $\lambda_j \geq 0$, $\sum_j \lambda_j = 1$. In particular, $0 \leq \lambda_j \leq 1$ which implies $\lambda_j^2 \leq \lambda_j$. Thus

$$\mathrm{Tr}(\rho^2) = \sum_{j=1}^{n}\lambda_j^2 \leq \sum_{j=1}^{n}\lambda_j = 1.$$

Equality holds only if $\lambda_j^2 = \lambda_j$ for all $j \in \{1, \ldots, n\}$, which means means $\lambda_j \in \{0, 1\}$, which can only happen for mixed states. □

Mixed states can be seen as "shadows" of pure states on a larger Hilbert space. This is the first instance in this course of the paradigm known as "Church of the Larger Hilbert Space".

**Proposition 4.6.** *If $\rho \in M_n(\mathbb{C})$ is a quantum state, then there exists a pure state $\sigma \in M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ such that $\rho = \mathrm{Tr}_1(\sigma)$.*

*Proof.* Let $\rho = \sum_{j=1}^n \lambda_j |\xi_j\rangle \langle\xi_j|$ be the spectral decomposition of $\rho$ and let $\xi = \sum_{j=1}^n \sqrt{\lambda_j}\xi_j \otimes \xi_j$. Since $\sum_j \lambda_j = 1$, we have

$$\langle \xi, \xi \rangle = \sum_{i,j=1}^n \sqrt{\lambda_i}\sqrt{\lambda_j}\langle \xi_i, \xi_j \rangle^2 = \sum_{j=1}^n \lambda_j = 1.$$

Thus $\sigma = |\xi\rangle \langle\xi|$ is a pure state in $M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$. Moreover,

$$\begin{aligned}
\mathrm{Tr}_1(\sigma) &= \mathrm{Tr}_1 \left( \sum_{i,j=1}^n \sqrt{\lambda_i}\sqrt{\lambda_j} \, |\xi_i\rangle \langle\xi_j| \otimes |\xi_i\rangle \langle\xi_j| \right) \\
&= \sum_{i,j=1}^n \sqrt{\lambda_i}\sqrt{\lambda_j} \mathrm{Tr}(|\xi_i\rangle \langle\xi_j|) \, |\xi_i\rangle \langle\xi_j| \\
&= \sum_{j=1}^n \lambda_j \, |\xi_j\rangle \langle\xi_j| \\
&= \rho.
\end{aligned}$$

□

**Definition 4.7.** If $\rho \in M_n(\mathbb{C})$ is a quantum state, any pure state $\sigma \in M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ such that $\mathrm{Tr}_1(\sigma) = \rho$ is a called *purification* of $\rho$.

**Remark.** By the previous proposition, it is always possible to take $m = n$. But in general, one can do much better. In the extreme case when $\rho$ is already a pure state or example, one can of course take $m = 1$.

**Definition 4.8.** A *quantum channel* (in the Schrödinger picture) is a completely positive trace-preserving linear map from $M_m(\mathbb{C})$ to $M_n(\mathbb{C})$.

It is immediate from the definition that quantum channels map quantum states to quantum states, which make them suitable to model changes of the state of a physical system.

## Exercises

Let

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$
$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

These matrices are called *Pauli matrices*. It can be useful to write $\sigma_0$ for the $2 \times 2$ identity matrix.

(a) Show that $(\sigma_j)_{j=0}^3$ is an orthonormal basis of the self-adjoint $2 \times 2$ matrices with the inner product $\frac{1}{2}\langle \cdot, \cdot \rangle_{\mathrm{HS}}$.

(b) Show that $\rho \in M_2(\mathbb{C})$ is a quantum state if and only if there exists $a \in \mathbb{R}^3$ with $a_1^2+a_2^2+a_3^3 \leq 1$ such that $\rho = \frac{1}{2}(I + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3)$.

The equivalence in (b) establishes a bijection between quantum states in $M_2(\mathbb{C})$ and the unit ball of $\mathbb{R}^3$. This graphical representation of qubit states is called the *Bloch sphere*.

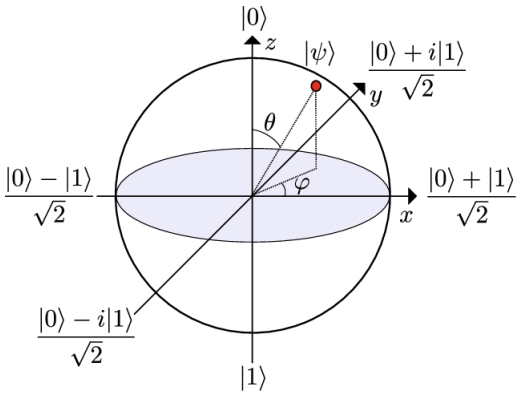


Figure 4.1: Bloch sphere. Points denoted by $|\xi\rangle$ correspond to the pure states $|\xi\rangle\langle\xi|$. The ONB of $\mathbb{C}^2$ is denoted by $|0\rangle$, $|1\rangle$.

The map $a \mapsto \frac{1}{2}(I+a_1\sigma_1+a_2\sigma_2+a_3\sigma_3)$ is affine, that is, it preserves convex combinations. Thus points on the surface of the unit ball correspond to pure states, while interior points correspond to mixed states. The center of the unit ball corresponds to the state $\frac{1}{2}I$, which is called the *maximally mixed state*.

# Chapter 5

# POVMs and quantum measurements

As discussed in the introduction, if $A$ is a self-adjoint matrix with an orthonormal basis $(\xi_j)$ consisting of eigenvectors and corresponding eigenvalues $(\lambda_j)$, then the measurement outcome for a quantum system in state $\xi$ is $\lambda_j$ with probability $|\langle \xi_j, \xi \rangle|^2$. Another way to express this is $\mathrm{Tr}(P_j |\xi\rangle \langle\xi|)$ with $P_j = |\xi_j\rangle \langle \xi_j|$.

This only describes the measurement outcomes for pure states. Mixed states can be written as convex combination of pure states, and the measurement probability are affine in the sense that a mixed state is a statistical mixture of pure states. This means that for arbitrary quantum states $\rho$, the probability to measure the value $\lambda_j$ is $\mathrm{Tr}(P_j\rho)$. When dealing with open quantum systems, this class of measurements is however to restrictive. We want to allow for a more general class of measurements, described by so-called POVMs.

**Definition 5.1.** Let $I$ be a finite set. A *projection-valued measure* (PVM) on $I$ with values in $M_n(\mathbb{C})$ is a family $(P_i)_{i\in I}$ of projections in $M_n(\mathbb{C})$ such that $\sum_{i\in I} P_i = 1$.

A positive-operator valued measure (POVM) on $I$ with values in $M_n(\mathbb{C})$ is a family $(P_i)_{i\in I}$ of positive semi-definite matrices in $M_n(\mathbb{C})$ such that $\sum_{i\in I} P_i = 1$.

Since projections are positive semi-definite, every PVM is a POVM. The measurement interpretation for POVMs is the same as for PVMs: The probability to measure outcome $i$ when the system is in state $\rho$ is given by $\mathrm{Tr}(P_i\rho)$.

**Example 5.2.** If $(\xi_i)_{i=1}^n$ is an orthonormal basis of $\mathbb{C}^n$, then $(|\xi_i\rangle \langle \xi_i|)_{i=1}^n$ is a PVM. One can also group them. For example $P_1 = |\xi_1\rangle \langle \xi_1|$, $P_2 = \sum_{i=2}^n |\xi_i\rangle \langle \xi_i|$ also forms a PVM.

**Example 5.3.** Let $\rho_1$, $\rho_2$ be quantum states and let $\rho_1 - \rho_2 = \sum_{i=1}^n \lambda_i |\xi_i\rangle \langle \xi_i|$ be the spectral decomposition of $\rho_1 - \rho_2$. A PVM on $\{0, 1, 2\}$ with values in $M_n(\mathbb{C})$ is given by

$$
\begin{aligned}
P_0 &= \sum_{i:\lambda_i=0} |\xi_i\rangle \langle \xi_i| \\
P_1 &= \sum_{i:\lambda_i>0} |\xi_i\rangle \langle \xi_i| \\
P_2 &= \sum_{i:\lambda_i<0} |\xi_i\rangle \langle \xi_i| .
\end{aligned}
$$

Thus measurement, known as Helstrom measurement, is used to optimally distinguish between the states $\rho_1$ and $\rho_2$

**Example 5.4.** For every $d \in \mathbb{N}$ the family $(1/d)_{i=1}^{d}$ is a POVM, although it is not very interesting as a measurement: $\mathrm{Tr}(1/d \cdot \rho) = \frac{1}{d}$ for every quantum state $\rho$, so it does not help to distinguish between states of a system.

There is a one-to-one correspondence between POVMs on $\{1, \dots, n\}$ with values in $M_m(\mathbb{C})$ and a special class of quantum channels from $M_m(\mathbb{C})$ to $M_n(\mathbb{C})$, called quantum-to-classical channels.

**Definition 5.5.** A quantum channel $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is called *quantum-to-classical channel* if $\Phi(A)$ is diagonal for every $A \in M_m(\mathbb{C})$.

**Example 5.6.** The dephasing channel $\Phi\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$, $A \mapsto \mathrm{diag}(A_{11}, \dots, A_{nn})$ is a quantum-to-classical channel.

More generally, if $\Psi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is any quantum channel and $\Phi\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$ is the dephasing channel, then $\Psi \circ \Phi$ is a quantum-to-classical channel. Vice versa, if $\Psi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ is a quantum-to-classical channel, then $\Phi \circ \Psi = \Psi$, hence all quantum-to-classical channels are of this form.

**Proposition 5.7.** *(a) For every quantum-to-classical channel $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ there exists a unique POVM on $\{1, \dots, n\}$ with values in $M_m(\mathbb{C})$ such that*

$$\Phi(A) = \sum_{i=1}^{n} \mathrm{Tr}(P_i A) E_{ii}$$

*for all $A \in M_m(\mathbb{C})$.*

*Moreover, if the restriction of $\Phi^\dagger$ to the diagonal matrices is a $*$-homomorphism, then $(P_i)_{i=1}^{n}$ is a PVM.*

*(b) If $(P_i)_{i=1}^{n}$ is a POVM with values in $M_m(\mathbb{C})$, then*

$$\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C}),\ A \mapsto \sum_{i=1}^{n} \mathrm{Tr}(P_i A) E_{ii}$$

*is a quantum-to-classical channel.*

*Proof.* (a) Let $P_i = \Phi^\dagger(E_{ii})$. Since $\Phi$ is completely positive, so is $\Phi^\dagger$, hence $P_i$ is positive. Moreover, since $\Phi$ is trace-preserving, $\Phi^\dagger$ is unital, which implies

$$\sum_{i=1}^{n} P_i = \sum_{i=1}^{n} \Phi^\dagger(E_{ii}) = \Phi^\dagger(1) = 1.$$

Finally, since $\Phi(A)$ is diagonal,

$$\sum_{i=1}^{n} \mathrm{Tr}(P_i A) E_{ii} = \sum_{i=1}^{n} \mathrm{Tr}(\Phi^\dagger(E_{ii}) A) E_{ii} = \sum_{i=1}^{n} \mathrm{Tr}(E_i i \Phi(A)) E_{ii} = \Phi(A)$$

If the restriction of $\Phi^\dagger$ to the diagonal matrices is a $*$-homomorphism, then

$$(P_i)_{kl} = \sum_{j=1}^{m} (P_i)_{kj} (P_i)_{jl} = \sum_{j=1}^{m} \Phi^\dagger(E_{ii})_{kj} \Phi^\dagger(E_{ii})_{jl} = \Phi(E_{ii})_{kl} = (P_i)_{kl}.$$

Thus $(P_i)_{i=1}^n$ is a PVM.

(b) Clearly $\Phi(A)$ is diagonal for all $A \in M_m(\mathbb{C})$. Furthemore,

$$\operatorname{Tr}(\Phi(A)) = \sum_{i=1}^{n} \operatorname{Tr}(P_i A) = \operatorname{Tr}\left(\left(\sum_{i=1}^{n} P_i\right) A\right) = \operatorname{Tr}(A),$$

which shows that $\Phi$ is trace-preserving.

To see that $\Phi$ is completely positive, it suffices to note that $\Phi$ is the sum of the completely positive maps $A \mapsto \operatorname{Tr}(P_i^{1/2} A P_i^{1/2}) E_{ii}$. □

Just like mixed states are "shadows" of pure states, POVMs are "shadows" of PVMs. This is another instance of the church of the larger Hilbert space.

**Theorem 5.8** (Naimark dilation theorem)**.** *If $(P_i)_{i\in I}$ is a POVM with values in $M_m(\mathbb{C})$, then there exists an isometry $V\colon \mathbb{C}^m \to \mathbb{C}^k$ and a PVM $(Q_i)_{i\in I}$ with values in $M_k(\mathbb{C})$ such that*

$$P_i = V^* Q_i V$$

*for $1 \le i \le n$.*

*Proof.* Let $\Phi$ be the quantum-to-classical channel associated with $(P_i)_{i\in I}$ by the previous proposition. By the Stinespring dilation theorem, there exists $V\colon \mathbb{C}^m \to \mathbb{C}^k$ and a unital $*$-homomorphism $\pi\colon M_n(\mathbb{C}) \to M_k(\mathbb{C})$ such that $\Phi^\dagger(A) = V^*\pi(A)V$ for all $A \in M_m(\mathbb{C})$. Since $\Phi$ is trace-preserving, $\Phi^\dagger$ is unital, which implies $V^*V = 1$.

Let $E$ denote the dephasing channel on $M_k(\mathbb{C})$ and $\Psi = E \circ \pi^\dagger$, which is a quantum-to-classical channel. Moreover, since $E^\dagger = E$ is the identity on diagonal matrices, the Hilbert–Schmidt adjoint of $\Psi$ acts as $\pi$ on the diagonal matrices, which is a $*$-homomorphism. Thus the POVM given by $Q_i = \Psi^\dagger(E_{ii})$ is a PVM.

Finally,

$$V^* Q_i V = V^*\pi(E^\dagger(E_{ii}))V = V^*\pi(E_{ii})V = \Phi^\dagger(E_{ii}) = P_i.$$ □

# Chapter 6

# Basic trace inequalities and convexity/concavity results

In this chapter we will see the first glimpse of a quantum entropy, namely the von Neumann entropy. It is the trace of a matrix-valued function, which connects it with the other part of the title of this course, the trace inequalities. More specifically, we will investigate monotonicity and convexity properties of maps of the form $A \mapsto \mathrm{Tr}(f(A))$ in this chapter.

**Lemma 6.1** (Peierls Inequality)**.** *If $A \in M_n(\mathbb{C})$ is self-adjoint, $f\colon \mathbb{R} \to \mathbb{R}$ is convex and $u_1, \dots, u_n$ is an orthonormal basis of $\mathbb{C}^n$, then*

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

*and equality holds if $(u_j)$ consists of eigenvectors of $A$.*

*Proof.* Let $v_1, \dots, v_n$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $A$ and let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. Since $f$ is convex, we have

$$\begin{aligned}
\sum_{j=1}^{n} f(\langle u_j, Au_j\rangle) &= \sum_{j=1}^{n} f\left(\sum_{k=1}^{n} \lambda_k |\langle v_k, u_j\rangle|^2\right) \\
&\leq \sum_{j,k=1}^{n} f(\lambda_k)|\langle v_k, u_j\rangle|^2 \\
&= \sum_{k=1}^{n} f(\lambda_k) \\
&= \mathrm{Tr} f(A).
\end{aligned}$$

The equality case is easy to see. □

**Lemma 6.2.** *If $A, B \in M_n(\mathbb{C})$ are self-adjoint and $f\colon \mathbb{R} \to \mathbb{R}$ is continuously differentiable, then the function*

$$\varphi\colon \mathbb{R} \to \mathbb{R},\ t \mapsto \mathrm{Tr}(f(A + tB))$$

*is differentiable with*

$$\varphi'(t) = \mathrm{Tr}(f'(A + tB)B).$$

*Proof.* Let us first consider the case when $f$ is a polynomial. Otherwise replacing $A$ by $A+tB$, it suffices to prove differentiability at 0. We have

$$(A+tB)^m = A^m + t(BA^{m-1} + ABA^{m-2} + \cdots + A^{m-1}B) + o(t)$$

and thus

$$\mathrm{Tr}((A+tB)^m) = \mathrm{Tr}(A^m) + t\mathrm{Tr}(mA^{m-1}B) + o(t).$$

Hence the statement holds when $f$ is a monomial, and by linearity also when $f$ is a polynomial.

Now let $f \in C^1(\mathbb{R})$ be arbitrary. For $T > 0$ we have

$$\|A+tB\| \leq \|A\| + T\|B\|$$

if $|t| \leq T$. In particular, the spectrum of $A+tB$ is contained in the interval $I_T = [-\|A\| - T\|B\|, \|A\| + T\|B\|]$.

By the Stone–Weierstraß theorem there exists a sequence of polynomial $p_k$ such that $p_k \to f$ and $p_k' \to f'$ uniformly on $I_T$. Let

$$\varphi_k \colon \mathbb{R} \to \mathbb{R},\ t \mapsto \mathrm{Tr}(p_k(A+tB)).$$

The uniform convergence of $(p_k)$ and $(p_k')$ implies $\varphi_k \to \varphi$ and

$$\varphi_k' \to \mathrm{Tr}(f'(A+\bullet B)B)$$

uniformly on $[-T,T]$.

Since $T > 0$ was arbitrary, the function $\varphi$ is differentiable with

$$\varphi'(t) = \mathrm{Tr}(f'(A+tB)B). \qquad \square$$

**Remark.** It was crucial in the proof that the trace is invariant under cyclic permutations. In general, it is *not* true that $\frac{d}{dt}|_{t=0} f(A+tB) = f'(A)B$, as simple examples show. In the exercises you will be asked to give a correct formula for this derivative in terms of the spectral decompositions of $A$ and $B$.

**Theorem 6.3.** *Let $f \colon \mathbb{R} \to \mathbb{R}$ be a function. If $f$ is monotone increasing, then so is $A \mapsto \mathrm{Tr} f(A)$ on $M_n(\mathbb{C})_{\mathrm{sa}}$. If $f$ is convex, then so is $A \mapsto \mathrm{Tr} f(A)$ on $M_n(\mathbb{C})_{\mathrm{sa}}$.*

*Proof.* First let $f$ be monotone increasing and let $A, B \in M_n(\mathbb{C})$ with $A \leq B$. We have to show that

$$\mathrm{Tr}(f(A)) \leq \mathrm{Tr}(f(B)).$$

We can assume without loss of generality that $f$ is continuously differentiable. By the previous lemma we have

$$\begin{aligned}\mathrm{Tr}(f(B)) - \mathrm{Tr}(f(A)) &= \int_0^1 \frac{d}{dt}\mathrm{Tr}(f(A+t(B-A)))\,dt \\ &= \int_0^1 \mathrm{Tr}(f'(A+t(B-A))(B-A))\,dt.\end{aligned}$$

Since $f$ is monotone increasing, $f'(A+t(B-A)) \geq 0$. Moreover, $B - A \geq 0$ by assumption. Thus the integrand is non-negative, which implies $\mathrm{Tr}(f(A)) \leq \mathrm{Tr}(f(B))$ as desired.

Now let $f$ be convex, $A, B \in M_n(\mathbb{C})$ be self-adjoint and $\lambda \in [0,1]$. Let $(u_j)$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $\lambda A + (1-\lambda)B$. By convexity of $f$ and Peierls inequality we have

$$\begin{aligned}\mathrm{Tr}(f(\lambda A + (1-\lambda B))) &= \sum_{j=1}^{n} \langle u_j, f(\lambda A + (1-\lambda B))u_j\rangle \\ &= \sum_{j=1}^{n} f(\langle u_j, (\lambda A + (1-\lambda B))u_j\rangle) \\ &\leq \sum_{j=1}^{n} \lambda f(\langle u_j, Au_j\rangle) + (1-\lambda) f(\langle u_j, Bu_j\rangle) \\ &\leq \lambda \mathrm{Tr}(f(A)) + (1-\lambda)\mathrm{Tr}(f(B)). \qquad \square\end{aligned}$$

**Remark.** That we can assume $f$ to be continuously differentiable in the first part of the proof will be justified in the exercises.

**Corollary 6.4.** *For self-adjoint $A \in M_n(\mathbb{C})$ and $\lambda \in \mathbb{R}$ let $N(A,\lambda)$ be the number of eigenvalues of $A$ less or equal than $\lambda$, counted with multiplicity. If $A \leq B$, then $N(A,\lambda) \geq N(B,\lambda)$ for all $\lambda \in \mathbb{R}$.*

*Proof.* The function $1_{(-\infty,\lambda]}$ is decreasing and

$$N(A,\lambda) = \mathrm{Tr}(1_{(-\infty,\lambda]}(A)).$$

Now the claim follows from the previous theorem. $\square$

**Theorem 6.5** (Klein's Inequality)**.** *Let $f$ be a continuously differentiable convex function on $\mathbb{R}$. Then for any $A, B \in M_n(\mathbb{C})_{\mathrm{sa}}$, we have*

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

*Proof.* By Lemma 6.2 the function

$$\varphi\colon \mathbb{R} \to \mathbb{R},\ t \mapsto \mathrm{Tr}(f(B + t(A-B)))$$

is differentiable with

$$\varphi'(0) = \mathrm{Tr}(f'(B)(A-B)).$$

Moreover, by the previous theorem, $\varphi$ is convex. Thus

$$\mathrm{Tr}(f'(B)(A-B)) = \varphi'(0) \leq \varphi(1) - \varphi(0) = \mathrm{Tr}(f(A)) - \mathrm{Tr}(f(B)). \qquad \square$$

**Theorem 6.6** (Peierls–Bogoliubov Inequality)**.** *The function $A \mapsto \log \mathrm{Tr}\exp(A)$ is convex on $M_n(\mathbb{C})_{\mathrm{sa}}$.*

*Proof.* Let

$$\varphi\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto \log\left(\sum_{k=1}^{n} e^{x_k}\right).$$

A direct computation shows

$$\frac{\partial^2 \varphi}{\partial x_j \partial x_k} = a_j \delta_{jk} - a_j a_k,$$

where

$$a_j = \frac{e^{x_j}}{\sum_{k=1}^n e^{x_k}}.$$

For any $y \in \mathbb{R}^n$ we have

$$\sum_{j,k=1}^n \frac{\partial^2 \varphi(x)}{\partial x_j \partial x_k} y_j y_k = \sum_{j=1}^n a_j y_j^2 - \sum_{j,k=1}^n a_j a_k y_j y_k = \sum_{j=1}^n a_j y_j^2 - \left( \sum_{j=1}^n a_j y_j \right)^2 \geq 0$$

by Jensen's inequality. Thus $\varphi$ is convex.

Let $A, B \in M_n(\mathbb{C})_{\mathrm{sa}}$, $\lambda \in [0,1]$ and let $(u_j)$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $\lambda A + (1-\lambda B)$. For $x_j = \langle u_j, Au_j \rangle$, $y_j = \langle u_j, Bu_j \rangle$ we have

$$\begin{aligned} \varphi(\lambda x + (1-\lambda)y) &= \log \left( \sum_{j=1}^n \exp(\langle u_j, (\lambda A + (1-\lambda)B) u_j \rangle) \right) \\ &= \log \left( \sum_{j=1}^n \langle u_j, \exp(\lambda A + (1-\lambda)B) u_j \rangle \right) \\ &= \log \operatorname{Tr}(\exp(\lambda A + (1-\lambda)B)). \end{aligned}$$

On the other hand, since $\varphi$ is convex,

$$\begin{aligned} \varphi(\lambda x + (1-\lambda)y) &\leq \lambda \varphi(x) + (1-\lambda)\varphi(y) \\ &= \lambda \log \left( \sum_{j=1}^n e^{\langle u_j, Au_j \rangle} \right) + (1-\lambda) \log \left( \sum_{j=1}^n e^{\langle u_j, Bu_j \rangle} \right) \\ &\leq \lambda \log \operatorname{Tr}(e^A) + (1-\lambda) \log \operatorname{Tr}(e^B), \end{aligned}$$

where the last step follows from Peierls inequality. □

**Remark** (Chandler Davis convexity theorem)**.** More generally, if $\varphi \colon \mathbb{R}^n \to \mathbb{R}$ is a symmetric convex function, then the map $\Phi$ that maps a self-adjoint matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ to $\varphi(\lambda_1, \ldots, \lambda_n)$ is convex (exercise).

**Theorem 6.7** (Duality formula of the quantum entropy)**.** *If $\rho \in M_n(\mathbb{C})$ is positive and $\operatorname{Tr}(\rho) = 1$, then*

$$\operatorname{Tr}(\rho \log \rho) = \sup\{\operatorname{Tr}(H\rho) - \log \operatorname{Tr}(e^H) \mid H \in M_n(\mathbb{C})_{\mathrm{sa}}\}.$$

**Remark.** The function $\lambda \mapsto \lambda \log \lambda$ can be continuously extended to a function $f$ on $[0, \infty)$ by setting $f(0) = 0$. Here and in the following we understand $\operatorname{Tr}(\rho \log \rho)$ as $\operatorname{Tr}(f(\rho))$, so that it also makes sense if $\rho$ is not positive definite.

*Proof.* For $H \in M_n(\mathbb{C})_{\mathrm{sa}}$ let $\sigma = \frac{e^H}{\operatorname{Tr}(e^H)}$. Let $f \colon [0,\infty) \to \mathbb{R}$ be the continuous extension of $\lambda \to \lambda \log \lambda$. This function is convex. By Klein's inequality,

$$\begin{aligned} 0 &\leq \operatorname{Tr}(f(\rho) - f(\sigma) - f'(\sigma)(\rho - \sigma)) \\ &= \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(\sigma \log \sigma) - \operatorname{Tr}((\log \sigma + 1)(\rho - \sigma)) \\ &= \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(\rho \log \sigma) \\ &= \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(H\rho) + \log \operatorname{Tr}(e^H). \end{aligned}$$

Thus $\mathrm{Tr}(\rho\log\rho) \geq \mathrm{Tr}(H\rho) - \log\mathrm{Tr}(e^H)$ for all $H \in M_n(\mathbb{C})_{\mathrm{sa}}$.

For the converse inequality let $\varepsilon > 0$ and $H_\varepsilon = \log(\rho + \varepsilon\mathbf{1})$. Then

$$\mathrm{Tr}(H_\varepsilon\rho) - \log\mathrm{Tr}(e^{H_\varepsilon}) = \mathrm{Tr}(\rho\log(\rho+\varepsilon\mathbf{1})) - \log(1+\varepsilon n) \to \mathrm{Tr}(\rho\log\rho)$$

as $\varepsilon \searrow 0$. Thus

$$\mathrm{Tr}(\rho\log\rho) \leq \sup\{\mathrm{Tr}(H\rho) - \log\mathrm{Tr}(e^H) \mid H \in M_n(\mathbb{C})_{\mathrm{sa}}\}. \qquad \square$$

In the previous theorem we encountered one of the central quantities in this course, the von Neumann entropy. Moreover, another entropy quantity was hidden in the proof, namely the relative entropy, which we will re-encounter later.

**Definition 6.8.** The *von Neumann entropy* $S(\rho)$ of a quantum state $\rho$ is defined as

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

**Definition 6.9.** Given a self-adjoint matrix $H \in M_n(\mathbb{C})$ and $\beta \in [-\infty,\infty]$, the Gibbs state for the Hamiltonian $H$ at inverse temperature $\beta$ is the density matrix $\rho_{\beta,H}$ given by

$$\rho_{\beta,H} = \frac{1}{\mathrm{Tr}(e^{-\beta H})}e^{-\beta H}$$

if $\beta \in \mathbb{R}$ and $\rho_{\pm\infty,H} = \lim_{\beta\to\pm\infty}\rho_{\beta,H}$.

Gibbs states are the equilibrium states for systems with Hamiltonian $H$ for fixed energy of the system. Mathematically, this can be expressed as follows.

**Theorem 6.10.** *Let $H \in M_n(\mathbb{C})$ be a self-adjoint matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. For each $E \in [\lambda_1,\lambda_n]$ there exists $\beta \in [-\infty,\infty]$ such that $E = \mathrm{Tr}(H\rho_{\beta,H})$ and*

$$S(\rho_{\beta,H}) = \max\{S(\rho) \mid \rho \text{ quantum state}, \mathrm{Tr}(\rho H) = E\}.$$

*Moreover, the Gibbs state $\rho_{\beta,H}$ satisfying $E = \mathrm{Tr}(H\rho_{\beta,H})$ is unique.*

*Proof.* Let $\rho$ be a quantum state with $\mathrm{Tr}(\rho H) = E$. Assume there exists a Gibbs state $\rho_{\beta,H}$ such that $\mathrm{Tr}(\rho_{\beta,H}) = E$ (we will show this afterwards). By the duality formula for the quantum entropy,

$$\begin{aligned} S(\rho) &= -\sup\{\mathrm{Tr}(A\rho) - \log\mathrm{Tr}(e^A) \mid A \in M_n(\mathbb{C})_{\mathrm{sa}}\} \\ &\leq \mathrm{Tr}(\beta H\rho) + \log\mathrm{Tr}(e^{-\beta H}) \\ &= \beta E + \log\mathrm{Tr}(e^{-\beta H}). \end{aligned}$$

On the other hand,

$$\begin{aligned} S(\rho_{\beta,H}) &= -\mathrm{Tr}(\rho_{\beta,H}\log\rho_{\beta,H}) \\ &= -\frac{1}{\mathrm{Tr}(e^{-\beta H})}\mathrm{Tr}(e^{-\beta H}(-\beta H - \log\mathrm{Tr}(e^{-\beta H})) \\ &= \beta\mathrm{Tr}(\rho_{\beta,H}H) + \log\mathrm{Tr}(e^{-\beta H}) \\ &= \beta E + \log\mathrm{Tr}(e^{-\beta H}). \end{aligned}$$

Thus $S(\rho) \leq S(\rho_{\beta,H})$. The case $\beta = \pm\infty$ follows by taking limits.

Now let us turn to the existence of a Gibbs state with energy $E$. If $H = E\mathbf{1}$, we can take any $\beta \in [-\infty,\infty]$ to get $\rho_{\beta,H} = \frac{\mathbf{1}}{n}$ with energy $\mathrm{Tr}(\rho_{\beta,\mathbf{1}}) = E$. Let us assume in the following that $H$ is not a multiple of the identity.

To show that $\mathrm{Tr}(\rho_{\beta,H}H)$ takes all values between $\lambda_1$ and $\lambda_n$, we use the intermediate value theorem. We have

$$\begin{aligned}\frac{d}{d\beta}\mathrm{Tr}(\rho_{\beta,H}H) &= \frac{d}{d\beta}\frac{\mathrm{Tr}(e^{-\beta H}H)}{\mathrm{Tr}(e^{-\beta H})}\\ &= -\frac{\mathrm{Tr}(e^{-\beta H}H^2)}{\mathrm{Tr}(e^{-\beta H})} + \left(\frac{\mathrm{Tr}(e^{-\beta H}H)}{\mathrm{Tr}(e^{-\beta H})}\right)^2\\ &= \mathrm{Tr}(\rho_{\beta,H}H)^2 - \mathrm{Tr}(\rho_{\beta,H}H^2).\end{aligned}$$

By the Cauchy–Schwarz inequality,

$$\mathrm{Tr}(\rho_{\beta,H}H)^2 = \mathrm{Tr}(\rho_{\beta,H}H\mathbf{1})^2 \le \mathrm{Tr}(\rho_{\beta,H}H^2)\mathrm{Tr}(\rho_{\beta,H}\mathbf{1}^2) = \mathrm{Tr}(\rho_{\beta,H}H^2)$$

with equality if and only if $H$ is a multiple of $\mathbf{1}$, which we ruled out.

Thus $\frac{d}{d\beta}\mathrm{Tr}(\rho_{\beta,H}H) < 0$, which implies that $\beta \mapsto \mathrm{Tr}(\rho_{\beta,H}H)$ is strictly increasing. Moreover, from

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

we deduce

$$\begin{aligned}\mathrm{Tr}(\rho_{\infty,H}H) &= \lim_{\beta\to\infty}\mathrm{Tr}(\rho_{\beta,H}H) = \lambda_1,\\ \mathrm{Tr}(\rho_{-\infty,H}H) &= \lim_{\beta\to-\infty}\mathrm{Tr}(\rho_{\beta,H}H) = \lambda_n.\end{aligned}$$

Hence $\mathrm{Tr}(\rho_{\beta,H}H)$ takes all values between $\lambda_1$ and $\lambda_n$ for $\beta \in [-\infty,\infty]$.

Uniqueness of the Gibbs state with energy $E$ follows from the strict monotonicity of $\beta \mapsto \mathrm{Tr}(\rho_{\beta,H}H)$ if $H$ is not a multiple of the identity, while in the case $H = E\mathbf{1}$ we have $\rho_{\beta,H} = \frac{E}{n}\mathbf{1}$ independently of $\beta$. □

**Lemma 6.11.** *For $A \in M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ and $B \in M_m(\mathbb{C})$ we have*

$$\mathrm{Tr}(\mathrm{Tr}_2(A)B) = \mathrm{Tr}(A(B\otimes\mathbf{1})).$$

*Proof.* If $A = X \otimes Y$, then

$$\mathrm{Tr}(\mathrm{Tr}_2(A)B) = \mathrm{Tr}(Y)\mathrm{Tr}(XB) = \mathrm{Tr}(XB\otimes Y) = \mathrm{Tr}((X\otimes Y)(B\otimes\mathbf{1}).$$

The general cse follows by linearity. □

For a density matrix $\rho \in M_m(\mathbb{C})\otimes M_n(\mathbb{C})$ we write $\rho_1$ and $\rho_2$ for $\mathrm{Tr}_2(\rho)$ and $\mathrm{Tr}_1(\rho)$, respectively.

**Proposition 6.12** (Subadditivity of quantum entropy)**.** *If $\rho \in M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ is a quantum state, then*

$$S(\rho) \le S(\rho_1) + S(\rho_2)$$

*with equality if and only if $\rho = \rho_1 \otimes \rho_2$.*

*Proof.* By the previous lemma,

$$S(\rho_1) = -\mathrm{Tr}\rho_1 \log \rho_1 = -\mathrm{Tr}\rho(\log \rho_1 \otimes \mathbf{1}),$$

and similarly

$$S(\rho_2) = -\mathrm{Tr}\rho_2 \log \rho_2 = -\mathrm{Tr}\rho(\mathbf{1} \otimes \log \rho_2).$$

Thus

$$S(\rho_1) + S(\rho_2) - S(\rho) = \mathrm{Tr}\rho(\log \rho - \log \rho_1 \otimes \mathbf{1} - \mathbf{1} \otimes \log \rho_2) = \mathrm{Tr}\rho(\log \rho - \log(\rho_1 \otimes \rho_2)) \geq 0$$

by Klein's inequality for $f(x) = x \log x$:

$$\begin{aligned} 0 &\leq \mathrm{Tr}(f(\rho) - f(\rho_1 \otimes \rho_1) - f'(\rho_1 \otimes \rho_2)(\rho - \rho_1 \otimes \rho_2)) \\ &= \mathrm{Tr}(\rho(\log \rho - \log(\rho_1 \otimes \rho_2))). \end{aligned}$$

□

**Theorem 6.13** (Golden–Thompson inequality)**.** *If $A, B \in M_n(\mathbb{C})$ are self-adjoint, then*

$$\mathrm{Tr}(e^{A+B}) \leq \mathrm{Tr}(e^A e^B).$$

*Proof.* We first show by induction that

$$|\mathrm{Tr}(C_1 \dots C_{2^k})| \leq \mathrm{Tr}(|C_1|^{2^k})^{2^{-k}} \dots \mathrm{Tr}(|C_{2^k}|^{2^k})^{2^{-k}}$$

for all $k \in \mathbb{N}$ and $C_1, \dots, C_{2^k} \in M_n(\mathbb{C})$. For $k = 1$ this follows from the Cauchy–Schwarz inequality:

$$|\mathrm{Tr}(C_1 C_2)| = |\langle C_1^*, C_2 \rangle_{\mathrm{HS}}| \leq \mathrm{Tr}(C_1 C_1^*)^{1/2} \mathrm{Tr}(C_2^* C_2)^{1/2} = \mathrm{Tr}(|C_1|^2)^{1/2} \mathrm{Tr}(|C_2|^2)^{1/2}.$$

For the induction step, we have

$$|\mathrm{Tr}(C_1 \dots C_{2^{k+1}})| \leq \mathrm{Tr}(|C_1 C_2|^{2^k})^{2^{-k}} \dots \mathrm{Tr}(|C_{2^{k+1}-1} C_{2^{k+1}}|^{2^k})^{2^{-k}}.$$

By cyclicity of the trace and a second application of the induction hypothesis,

$$\begin{aligned} \mathrm{Tr}(|C_1 C_2|^{2^k}) &= \mathrm{Tr}(C_2^* C_1^* C_1 C_2 \dots C_2^* C_1^* C_1 C_2) \\ &= \mathrm{Tr}((C_1^* C_1 C_2 C_2^*)^{2^{k-1}}) \\ &\leq \mathrm{Tr}(|C_1^* C_1|^{2^k})^{1/2} \mathrm{Tr}(|C_2 C_2^*|^{2^k})^{1/2} \\ &= \mathrm{Tr}(|C_1|^{2^{k+1}})^{1/2} \mathrm{Tr}(|C_2|^{2^{k+1}})^{1/2}. \end{aligned}$$

Applying the same argument to the other factors, we get

$$|\mathrm{Tr}(C_1 \dots C_{2^{k+1}})| \leq \mathrm{Tr}(|C_1|^{2^{k+1}})^{2^{-(k+1)}} \dots \mathrm{Tr}(|C_{2^{k+1}}|^{2^{k+1}})^{2^{-(k+1)}}$$

as desired.

Now take $C_1 = \dots = C_{2^k} = XY$ for self-adjoint $X, Y \in M_n(\mathbb{C})$. By the previous step and the cyclicity of the trace,

$$\mathrm{Tr}((XY)^{2^k}) \leq \mathrm{Tr}(|XY|^{2^k}) = \mathrm{Tr}((YX^2Y)^{2^{k-1}}) = \mathrm{Tr}((X^2Y^2)^{2^{k-1}}).$$

By induction one obtains $\mathrm{Tr}((XY)^{2^k}) \leq \mathrm{Tr}(X^{2^k} Y^{2^k})$.

If we take $X = e^{2^{-k}A}$, $Y = e^{2^{-k}B}$, then

$$\mathrm{Tr}((e^{2^{-k}A} e^{2^{-k}B})^{2^k}) \leq \mathrm{Tr}(e^A e^B).$$

By the Lie–Trotter product formula, the left side converges to $\mathrm{Tr}(e^{A+B})$ as $k \to \infty$. □

## Exercises

**Exercise 6.1.** Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuously differentiable and let $A, B \in M_n(\mathbb{C})$ be self-adjoint. Show that the map

$$\Phi\colon \mathbb{R} \to M_n(\mathbb{C}),\ t \mapsto f(A+tB)$$

is differentiable and express $\Phi'(0)$ in terms of the spectral decomposition of $A$ and $B$.

**Exercise 6.2.** Show that for every $A \in M_n(\mathbb{C})$ the map $t \mapsto e^{tA}$ is differentiable with $\frac{d}{dt}e^{tA} = Ae^{tA}$ (Hint: Use that $e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!}A^k$).

**Exercise 6.3.** Let $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n \in \mathbb{R}$ with $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ and $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$.

(a) Show that there exists a continuously differentiable increasing function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f(\lambda_k) = \mu_k$ for all $k \in \{1, \ldots, n\}$. Show that $f$ can be chosen strictly increasing if $\mu_1 < \cdots < \mu_n$.

(b) Let $g\colon \mathbb{R} \to \mathbb{R}$ be increasing and let $A_1, \ldots, A_m \in M_n(\mathbb{C})$ be self-adjoint. Show that there exists a continuously differentiable increasing function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f(A_k) = g(A_k)$ for all $k \in \{1, \ldots, n\}$.

**Exercise 6.4.** For any $A, B \in M_n(\mathbb{C})_{\mathrm{sa}}$, show that

$$\log\left(\frac{\mathrm{Tr}e^{A+B}}{\mathrm{Tr}e^{A}}\right) \geq \frac{\mathrm{Tr}(e^{A}B)}{\mathrm{Tr}e^{A}}.$$

In particular, when $\mathrm{Tr}e^{A} = 1$, we have

$$\log \mathrm{Tr}e^{A+B} \geq \mathrm{Tr}(e^{A}B).$$

**Exercise 6.5.** (a) Let $\omega\colon M_n(\mathbb{C})$ be a linear functional with $\omega(\mathbf{1}) = 1$. Show that $\omega(A) \geq 0$ for all $A \in M_n(\mathbb{C})_+$ if and only if $\|\omega\| = 1$.

(b) Show that there exists a bijection

$$f\colon \{\rho \in M_n(\mathbb{C})_+ \mid \mathrm{Tr}(\rho) = 1\} \to \{\omega\colon M_n(\mathbb{C}) \to \mathbb{C} \mid \varphi \text{ linear}, \varphi(1) = \|\varphi\| = 1\}$$

such that $f(\lambda\rho + (1-\lambda)\sigma) = \lambda f(\rho) + (1-\lambda)f(\sigma)$ for all $\rho, \sigma \in M_n(\mathbb{C})_+$ with $\mathrm{Tr}(\rho) = \mathrm{Tr}(\sigma) = 1$ and $\lambda \in [0,1]$.

# Chapter 7

# Operator monotonicity and operator concavity/convexity

Recall that for positive semi-definite square matrices $A, B$, we write $A \leq B$ if $B - A \geq 0$. If $A \leq B$, then $C^*AC \leq C^*BC$. In the following, we work with positive-definite matrices for simplicity.

**Definition 7.1.** A function $f : (0, \infty) \to \mathbb{R}$ is said to be *operator monotone* if $A \leq B$ implies $f(A) \leq f(B)$ for positive definite square matrices $A, B$ of arbitrary size.

Clearly, every operator monotone function is (scalar) monotone, as can be seen by plugging in $1 \times 1$ matrices. The converse is not true, as we shall see in the exercises.

**Example 7.2.** For $\alpha \geq 0$ and $\beta \in \mathbb{R}$ the function $x \mapsto \alpha x + \beta$ is operator monotone.

Beyond this rather trivial class of examples, it takes some work to come up with more interesting operator monotone functions. We will get two know two (well, one plus one family) in the next proposition. First, however, let us see a monotone function which is not operator monotone.

**Example 7.3.** The function $x \mapsto x^2$ is not operator monotone. Indeed, the matrices

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \; B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

clearly satisfy $A \leq B$, yet

$$B^2 - A^2 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} - \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}$$

has determinant $-1$, so that it is not positive semidefinite.

**Proposition 7.4.** *(a) The function $x \mapsto -x^{-1}$ is operator monotone.*

*(b) For $\alpha \in [0, 1]$ the function $x \mapsto x^\alpha$ is operator monotone.*

*Proof.* (a) Recall that for $C \geq 0$, $C \leq \mathbf{1}$ iff $\|C\| \leq 1$. Also, $\|X^*X\| = \|X\|^2 = \|X^*\|^2 = \|XX^*\|$. If $A, B \in M_n(\mathbb{C})$ are positive definite and $A \leq B$, then $B^{-1/2}AB^{-1/2} \leq \mathbf{1}$, hence $\|B^{-1/2}AB^{-1/2}\| \leq 1$. Thus

$$\begin{aligned} \|A^{1/2}B^{-1}A^{1/2}\| &= \|(B^{-1/2}A^{1/2})^*(B^{-1/2}A^{1/2})\| \\ &= \|(B^{-1/2}A^{1/2})(B^{-1/2}A^{1/2})^*\| \\ &= \|B^{-1/2}AB^{-1/2}\| \\ &\leq 1, \end{aligned}$$

which implies $A^{-1/2}B^{-1}A^{-1/2} \leq \mathbf{1}$. Hence $B^{-1} \leq A^{-1}$.

(b) Clearly this is true for $\alpha \in \{0, 1\}$. Let $E$ be the set of numbers $\alpha \in [0, 1]$ for which $x \mapsto x^\alpha$ is operator monotone. We will show that $E$ is convex.

Let $\alpha, \beta \in E$ and let $A, B \in M_n(\mathbb{C})_+$ with $A \leq B$. Since $x \mapsto x^\alpha$ is operator monotone, $A^\alpha \leq B^\alpha$, which implies $B^{-\alpha/2}A^\alpha B^{-\alpha/2} \leq \mathbf{1}$. Hence $\|A^{\alpha/2}B^{-\alpha/2}\| \leq 1$. Similarly $\|B^{-\beta/2}A^{\beta/2}\| \leq 1$.

Let $r(S) = \max\{|\lambda| : \lambda \in \sigma(S)\}$ for $S \in M_n(\mathbb{C})$ and note that $r(ST) = r(TS)$ for all invertible $S, T \in M_n(\mathbb{C})$. We have

$$\begin{aligned}
&r(B^{-(\alpha+\beta)/4}A^{(\alpha+\beta)/2}B^{-(\alpha+\beta)/4}) \\
&= r(B^{(\alpha-\beta)/4}B^{-(\alpha+\beta)/4}A^{(\alpha+\beta)/2}B^{-(\alpha+\beta)/4}B^{-(\alpha-\beta)/4}) \\
&= r(B^{-\beta/2}A^{(\alpha+\beta)/2}B^{-\alpha/2}) \\
&\leq \|B^{-\beta/2}A^{(\alpha+\beta)/2}B^{-\alpha/2}\| \\
&\leq \|B^{-\beta/2}A^{\beta/2}\|\|A^{\alpha/2}B^{-\alpha/2}\| \\
&\leq 1.
\end{aligned}$$

Therefore $B^{-(\alpha+\beta)/4}A^{(\alpha+\beta)/2}B^{-(\alpha+\beta)/4} \leq 1$, hence $A^{(\alpha+\beta)/2} \leq B^{(\alpha+\beta)/2}$.

Thus $(\alpha + \beta)/2 \in E$. Moreover, a continuity argument shows that $E$ is closed. Thus $E$ is convex. Together with $\{0, 1\} \subset E$ this implies $E = [0, 1]$, as desired. □

**Remark.** If $S, T \in M_n(\mathbb{C})$ are invertible, then $TS = T(ST)T^{-1}$, that is, $ST$ and $TS$ are similar. Thus they have the same eigenvalues. In particular, $r(ST) = r(TS)$ as used in the proof of the previous theorem. In general $\sigma(ST) \setminus \{0\} = \sigma(TS) \setminus \{0\}$, thus $r(ST) = r(TS)$.

**Definition 7.5.** A function $f : (0, \infty) \to \mathbb{R}$ is said to be *operator convex* if for any $n \in \mathbb{N}$, any positive definite matrices $A, B \in M_n(\mathbb{C})$ and any $\lambda \in (0, 1)$ we have

$$f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda)f(B).$$

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

As with operator monotone functions, every operator convex (resp. operator concave) function is convex (resp. concave), but the converse is not true.

**Example 7.6.** The square function $f(x) = x^2$ is operator convex. In fact, for any positive semi-definite $A, B$ and $\lambda \in (0, 1)$, we have

$$\lambda A^2 + (1 - \lambda)B^2 - (\lambda A + (1 - \lambda)B)^2 = \lambda(1 - \lambda)(A - B)^2 \geq 0.$$

**Example 7.7.** The cube function $f(x) = x^3$ is not operator convex. In fact, if $f$ is operator convex, then we must have (since $A + tB = (1 - t)A + t(A + B)$)

$$f(A + tB) \leq (1 - t)f(A) + tf(A + B),$$

for any positive semi-definite $A, B$ and $t \in (0, 1)$. The above inequality can be reformulated as

$$\frac{(A + tB)^3 - A^3}{t} \leq (A + B)^3 - A^3.$$

Letting $t \to 0^+$, we get

$$B^3 + B^2A + BAB + AB^2 \geq 0.$$

Now we choose

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then $B = B^2 = B^3 = BAB$ and

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

which is not positive semi-definite. This leads to a contradiction.

**Proposition 7.8.** *The function $x \mapsto x^{-1}$ is operator convex.*

*Proof.* Let $A, B \in M_n(\mathbb{C})$ be positive definite and let $C = A^{-1/2}BA^{-1/2}$. For $\lambda \in [0,1]$ we have

$$\begin{aligned} &\lambda A^{-1} + (1-\lambda)B^{-1} - (\lambda A + (1-\lambda)B)^{-1} \\ &= A^{-1/2}(\lambda \mathbf{1} + (1-\lambda)C^{-1} - (\lambda \mathbf{1} + (1-\lambda)C)^{-1})A^{-1/2} \end{aligned}$$

Since the real-valued function $x \mapsto x^{-1}$ is convex,

$$(\lambda + (1-\lambda)x)^{-1} \leq \lambda + (1-\lambda)x^{-1}$$

for any $x \in \mathbb{R}$. Applying this to the eigenvalues of $C$ implies

$$(\lambda \mathbf{1} + (1-\lambda)C)^{-1} \leq \lambda \mathbf{1} + (1-\lambda)C^{-1},$$

which yields the desired inequality. □

To show the operator convexity/concavity of $x \mapsto x^p$ for some (but not all, see the exercises) other values of $p$, we will use the following integral representations.

**Lemma 7.9.** *For positive definite $A \in M_n(\mathbb{C})$, the following integral formulas hold.*

$$\begin{aligned} A^p &= \frac{\sin(p+1)\pi}{\pi} \int_0^\infty t^p (t\mathbf{1} + A)^{-1}\, dt \quad \text{for } p \in (-1, 0), \\ A^p &= \frac{\sin p\pi}{\pi} \int_0^\infty t^p (t^{-1}\mathbf{1} - (t\mathbf{1} + A)^{-1})\, dt \quad \text{for } p \in (0, 1), \\ A^p &= \frac{\sin(p-1)\pi}{\pi} \int_0^\infty t^{p-1} (t^{-1}A + t(t\mathbf{1} + A)^{-1} - \mathbf{1})\, dt \quad \text{for } p \in (1, 2). \end{aligned}$$

*Proof.* Exercise. □

With this lemma, one can prove directly that

**Proposition 7.10.** *For the power functions $f_p(x) := x^p$,*

*(a) when $-1 \leq p < 0$, $-f_p$ is operator monotone and operator concave;*

*(b) when $0 \leq p \leq 1$, $f_p$ is operator monotone and operator concave;*

*(c) when $1 \leq p \leq 2$, $f_p$ is operator convex.*

*Proof.* From the integral identities in the previous lemma, it suffices to show that for any $t > 0$, $x \mapsto -(t+x)^{-1}$ is operator monotone and operator concave. □

**Proposition 7.11.** *We have the following*

*(a)* $f(x) = \log x$ *is operator concave and operator monotone;*

*(b)* $f(x) = x \log x$ *is operator convex.*

*Proof.* Exercise. □

**Theorem 7.12** (Loewner's Theorem)**.** *A function* $f : (0, \infty) \to \mathbb{R}$ *is operator monotone if and only if it is of the form*

$$f(x) = ax + b - \int_0^\infty \frac{1 - tx}{t + x} d\mu(t), \tag{7.1}$$

*where* $a \geq 0, b \in \mathbb{R}$ *and* $\mu$ *is a positive finite measure on* $(0, \infty)$*.*

*Proof.* See the book of Barry Simon. □

**Theorem 7.13.** *Let* $f$ *be a (continuous) function that maps* $(0, \infty)$ *into itself. Then the following are equivalent:*

*(a)* $f$ *is operator monotone;*

*(b)* $f$ *is operator concave.*

*Both of them imply*

*(c)* $f^{-1}$ *is operator convex.*

*Proof.* We first show (b) $\implies$ (c). Assume (b), then for any $\lambda \in (0, 1)$ and any positive definite $A, B$, we have

$$f(\lambda A + (1 - \lambda B)) \geq \lambda f(A) + (1 - \lambda) f(B).$$

By operator monotonicity and operator concavity of $-x^{-1}$, then

$$f(\lambda A + (1 - \lambda B))^{-1} \leq [\lambda f(A) + (1 - \lambda) f(B)]^{-1} \leq \lambda f(A)^{-1} + (1 - \lambda) f(B)^{-1}.$$

So we have (c).

Now we prove the equivalence of (a) and (b). Assume (b). Then for any $0 \leq A \leq B$, we will show that $f(A) \leq f(B)$. For this note that for any $\lambda \in (0, 1)$:

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

By operator concavity, we have

$$f(\lambda B) \geq \lambda f(A) + (1 - \lambda) f\left( \frac{\lambda}{1 - \lambda} (B - A) \right).$$

Since $f \geq 0$ and $B - A \geq 0$, we get $f(\lambda B) \geq \lambda f(A)$ for any $\lambda \in (0, 1)$. Letting $\lambda \to 1^-$, we get by continuity that $f(B) \geq f(A)$. So $f$ is operator monotone and we have (a).

Now assume (a). Let $A, B \in M_n(\mathbb{C})_{++}$ and $\lambda \in [0,1]$. Write $\mathbf{1}_n$ for the unit matrix in $M_n(\mathbb{C})$. Define the unitary matrix $V \in M_{2n}(\mathbb{C})$ by

$$U = \begin{pmatrix} \lambda^{1/2}\mathbf{1}_n & -(1-\lambda)^{1/2}\mathbf{1}_n \\ (1-\lambda)^{1/2}\mathbf{1}_n & \lambda^{1/2}\mathbf{1}_n \end{pmatrix}.$$

A direct computation shows

$$U^* \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} U = \begin{pmatrix} \lambda A + (1-\lambda)B & \lambda^{1/2}(1-\lambda)^{1/2}(B-A) \\ \lambda^{1/2}(1-\lambda)^{1/2}(B-A) & (1-\lambda)A + \lambda B \end{pmatrix}.$$

Let $D = -\lambda^{1/2}(1-\lambda)^{1/2}(B-A)$ and note that for $\varepsilon > 0$ we have

$$\begin{pmatrix} \lambda A + (1-\lambda)B + \varepsilon\mathbf{1}_n & 0 \\ 0 & 2\mu\mathbf{1}_n \end{pmatrix} - U^* \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} U = \begin{pmatrix} \varepsilon\mathbf{1}_n & D \\ D & 2\mu\mathbf{1}_n - (1-\lambda)A + \lambda B \end{pmatrix} \geq \begin{pmatrix} \varepsilon\mathbf{1}_n & D \\ D & \mu\mathbf{1}_n \end{pmatrix}$$

if $\mu \geq \|\lambda A + (1-\lambda)B\|$.

By the Schur complement theorem,

$$\begin{pmatrix} \varepsilon\mathbf{1}_n & D \\ D & \mu\mathbf{1}_n \end{pmatrix} \geq 0$$

if $\mu \geq \varepsilon^{-1}\|D\|^2$. Thus

$$\begin{pmatrix} \lambda A + (1-\lambda)B + \varepsilon\mathbf{1}_n & 0 \\ 0 & 2\mu\mathbf{1}_n \end{pmatrix} \geq U^* \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} U$$

for $\mu$ sufficiently large.

Since $U$ is unitary and $f$ is operator monotone, we have

$$\begin{aligned} &\begin{pmatrix} \lambda f(A) + (1-\lambda)f(B) & \lambda^{1/2}(1-\lambda)^{1/2}(f(B)-f(A)) \\ \lambda^{1/2}(1-\lambda)^{1/2}(f(B)-f(A)) & (1-\lambda)f(A) + \lambda f(B) \end{pmatrix} \\ &= U^* \begin{pmatrix} f(A) & 0 \\ 0 & f(B) \end{pmatrix} U \\ &= f\left( U^* \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} U \right) \\ &\leq f\left( \begin{pmatrix} \lambda A + (1-\lambda)B + \varepsilon\mathbf{1}_n & 0 \\ 0 & 2\mu\mathbf{1}_n \end{pmatrix} \right) \\ &= \begin{pmatrix} f(\lambda A + (1-\lambda)B + \varepsilon\mathbf{1}_n) & 0 \\ 0 & 2f(\mu)\mathbf{1}_n \end{pmatrix}. \end{aligned}$$

Hence $\lambda f(A) + (1-\lambda)f(B) \leq f(\lambda A + (1-\lambda)B + \varepsilon\mathbf{1}_n)$. Letting $\varepsilon \searrow 0$ yields (by continuity of $f$) $\lambda f(A) + (1-\lambda)f(B) \leq f(\lambda A + (1-\lambda)B)$ as desired. □

**Remark.** With a little more work one can show that every operator monotone function is automatically continuous. From the proof, we see that we only need $f \geq 0$ in deriving operator monotonicity from operator concavity. It is not true if we don't have $f \geq 0$. For example, $f(x) = -x\log x$ is operator concave, but it is not even scalar monotone.

**Lemma 7.14** (Dilation of contractions)**.** *If $A \in M_n(\mathbb{C})$ with $A^*A \leq \mathbf{1}$, then there exists $m \geq n$ and a unitary $U \in M_m(\mathbb{C})$ such that $PU|_{\mathbb{C}^n} = A$, where $P\colon \mathbb{C}^m \to \mathbb{C}^n$ is the projection onto the first $n$ coordinates.*

*Proof.* Let $m = 2n$, $B = (1 - AA^*)^{1/2}$, $C = (1 - A^*A)^{1/2}$ and

$$U = \begin{pmatrix} A & B \\ C & -A^* \end{pmatrix}.$$

We have

$$U^*U = \begin{pmatrix} A^* & C \\ B & -A \end{pmatrix} \begin{pmatrix} A & B \\ C & -A^* \end{pmatrix} = \begin{pmatrix} A^*A + C^2 & A^*B - CA^* \\ BA - AC & B^2 + AA^* \end{pmatrix}.$$

By definition, $A^*A + C^2 = B^2 + AA^* = \mathbf{1}$. Moreover, $A^*B = CA^*$ and $BA = AC$ (exercise). The property $PU|_{\mathbb{C}^n}$ is immediate from the definition of $U$. □

**Theorem 7.15** (Jensen's inequality for operators)**.** *If $f\colon [0,\infty) \to \mathbb{R}$ is operator convex, then*

$$f(\Phi(A)) \leq \Phi(f(A))$$

*for all $A \in M_m(\mathbb{C})_+$ and unital completely positive maps $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$.*

*If additionally $f(0) \leq 0$, then the same inequality holds for all contractive completely positive maps $\Phi$.*

*Proof.* Let $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ be a contractive completely positive map. By Stinespring's theorem, there exist $W \in M_{k,n}(\mathbb{C})$ and a unital $*$-homomorphism $\pi\colon M_m(\mathbb{C}) \to M_k(\mathbb{C})$ such that

$$\Phi(A) = W^*\pi(A)W.$$

Moreover, since $\Phi$ is contractive, $W^*W = \Phi(\mathbf{1}) \leq 1$.

Let $B = (\mathbf{1} - WW^*)^{1/2}$, $C = (\mathbf{1} - W^*W)^{1/2}$ and

$$X = \begin{pmatrix} \pi(A) & 0 \\ 0 & 0 \end{pmatrix},\ U = \begin{pmatrix} W & B \\ C & -W^* \end{pmatrix},\ V = \begin{pmatrix} W & -B \\ C & W^* \end{pmatrix}.$$

By the previous lemma, $U$ and $V$ are unitary. Furthermore,

$$U^*XU = \begin{pmatrix} W^*\pi(A)W & W^*\pi(A)B \\ B\pi(A)W & B\pi(A)B \end{pmatrix},\ V^*XV = \begin{pmatrix} W^*\pi(A)W & -W^*\pi(A)B \\ -B\pi(A)W & B\pi(A)B \end{pmatrix}.$$

Since $f$ is operator convex, we obtain

$$\begin{aligned} \begin{pmatrix} f(W^*\pi(A)W) & 0 \\ 0 & f(B\pi(A)B) \end{pmatrix} &= f\left(\begin{pmatrix} W^*\pi(A)W & 0 \\ 0 & B\pi(A)B \end{pmatrix}\right) \\ &= f\left(\frac{1}{2}U^*XU + \frac{1}{2}V^*XV\right) \\ &\leq \frac{1}{2}U^*f(X)U + \frac{1}{2}V^*f(X)V \\ &= \frac{1}{2}U^*\begin{pmatrix} f(\pi(A)) & 0 \\ 0 & f(0)\mathbf{1} \end{pmatrix}U + \frac{1}{2}V^*\begin{pmatrix} f(\pi(A)) & 0 \\ 0 & f(0)\mathbf{1} \end{pmatrix}V \end{aligned}$$

If $f(0) \leq 0$, then

$$\begin{aligned} &\frac{1}{2}U^*\begin{pmatrix} f(\pi(A)) & 0 \\ 0 & f(0)\mathbf{1} \end{pmatrix}U + \frac{1}{2}V^*\begin{pmatrix} f(\pi(A)) & 0 \\ 0 & f(0)\mathbf{1} \end{pmatrix}V \\ \leq\ &\frac{1}{2}U^*\begin{pmatrix} \pi(f(A)) & 0 \\ 0 & 0 \end{pmatrix}U + \frac{1}{2}V^*\begin{pmatrix} \pi(f(A)) & 0 \\ 0 & 0 \end{pmatrix}V \\ =\ &\begin{pmatrix} W^*\pi(f(A))W & 0 \\ 0 & B\pi(f(A))B \end{pmatrix}. \end{aligned}$$

Therefore

$$f(\Phi(A)) = f(W^*\pi(A)W) \leq W^*\pi(f(A))W = \Phi(f(A)).$$

On the other hand, if $\pi$ is unital, then $\mathbf{1} = \pi(\mathbf{1}) = V^*V$, hence $C = 0$. Then

$$\begin{aligned}&\frac{1}{2}U^*\begin{pmatrix} f(\pi(A)) & 0 \\ 0 & f(0)\mathbf{1}\end{pmatrix}U + \frac{1}{2}V^*\begin{pmatrix} f(\pi(A)) & 0 \\ 0 & f(0)\mathbf{1}\end{pmatrix}V \\ &= \begin{pmatrix} W^*f(\pi(A))W & 0 \\ 0 & Bf(A)B + f(0)WW^*\end{pmatrix}.\end{aligned}$$

Thus $f(W^*\pi(A)W) \leq W^*f(\pi(A))W$, and we conclude as before. □

**Corollary 7.16.** *If $f\colon [0,\infty) \to \mathbb{R}$ is operator convex, then*

$$f\left(\sum_{j=1}^m V_j^*A_jV_j\right) \leq \sum_{j=1}^m V_j^*f(A_j)V_j$$

*for all $A_1, \ldots, A_m \in M_n(\mathbb{C})$ and all $V_1, \ldots, V_m \in M_{n,k}(\mathbb{C})$ with $\sum_{j=1}^m V_j^*V_j = \mathbf{1}$.*

*If additionally $f(0) \leq 0$, the same inequality holds under the assumption $\sum_{j=1}^m V_j^*V_j \leq \mathbf{1}$.*

## Exercises

**Exercise 7.1.** For a continuously differentiable function $f\colon (0,\infty) \to \mathbb{R}$ let

$$Df\colon (0,\infty)^2 \to \mathbb{R},\ (\lambda,\mu) \mapsto \begin{cases} \frac{f(\lambda)-f(\mu)}{\lambda-\mu} & \text{if } \lambda \neq \mu, \\ f'(\lambda) & \text{if } \lambda = \mu.\end{cases}$$

Show that $f$ is operator monotone if and only if for all $n \in \mathbb{N}$ and $\lambda_1, \ldots, \lambda_n > 0$ the matrix $[Df(\lambda_j, \lambda_k)]_{j,k}$ is positive semi-definite.

**Exercise 7.2.** Show that the set

$$E = \{\log f \mid f\colon (0,\infty) \to (0,\infty) \text{ operator monotone}\}$$

is convex.

**Exercise 7.3.** Show the integral formulas from Lemma 7.9.

**Exercise 7.4.** Show that

1. $f(x) = \log x$ is operator concave and operator monotone;
2. $f(x) = x\log x$ is operator convex.
3. $f(x) = \frac{x-1}{\log x}$ is operator concave and operator monotone.

*Proof.* Hint: we have

$$\log x = \int_0^\infty \left(\frac{1}{t+1} - \frac{1}{t+x}\right)dt,$$

$$x\log x = \lim_{p\to 1^+} \frac{x^p - x}{p-1},$$

$$\frac{x-1}{\log x} = \int_0^1 x^\alpha d\alpha.$$

□

**Exercise 7.5.** Show that $x \mapsto x^p$ is neither operator convex nor operator concave if $p \notin [-1, 2]$.

**Exercise 7.6.** Give an example of an operator concave function $f\colon (0,1) \to \mathbb{R}$ that is not operator monotone.

**Exercise 7.7.** Show that every operator monotone function $f\colon (0,\infty) \to \mathbb{R}$ is continuous.

**Exercise 7.8.** Show that if $A \in M_n(\mathbb{C})$ and $f\colon [0,\infty) \to \mathbb{R}$, then $Af(A^*A) = f(AA^*)A$.

**Exercise 7.9.** Let $f\colon [0,\infty) \to \mathbb{R}$ be a continuous function such that $f(\Phi(A)) \leq \Phi(f(A))$ for all $n \in \mathbb{N}$, $A \in M_n(\mathbb{C})$ and all unital completely positive maps $\Phi\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$. Show that $f$ is operator convex.

# Chapter 8

# Lieb's concavity theorem

In 1963, Wigner, Yanase and Dyson conjectured that

$$S_p(\rho) = \frac{1}{2}\mathrm{Tr}[K, \rho^p][K, \rho^{1-p}]$$

is concave in $\rho$, where $K = K^*$ is arbitrary. The quantity $-S_p(\rho)$ is sometimes called Wigner–Yanase–Dyson skew information. It was resolved in 1973 by Lieb. Among many others, Lieb proved the following result. We write $M_n(\mathbb{C})_+$ for the positive $n \times n$ matrices and $M_n(\mathbb{C})_{++}$ for the positive definite $n \times n$ matrices.

**Theorem 8.1** (Lieb's concavity theorem and Ando's convexity theorem)**.** *For $K \in M_n(\mathbb{C})$ the function*

$$M_n(\mathbb{C})_+ \times M_n(\mathbb{C})_+ \to \mathbb{C},\ (A, B) \mapsto \mathrm{Tr}(K^* A^p K B^{1-p}),$$

*is jointly concave if $p \in [0, 1]$ and jointly convex if $p \in [-1, 0]$.*

*Remark* 8.2. The parameters can be more general, as we shall see later. The convexity result is named after Ando as he proved it in 1979, but this result is contained in another result of Lieb in the same 1973 paper.

**Corollary 8.3.** *The quantum relative entropy*

$$D(\rho||\sigma) := \mathrm{Tr}(\rho(\log \rho - \log \sigma)).$$

*is jointly convex in density matrices $\rho$ and $\sigma$.*

*Proof.* It follows from Lieb's concavity theorem and

$$D(\rho||\sigma) = \lim_{p \to 1^-} \frac{1}{p-1}\left(\mathrm{Tr}\rho^p\sigma^{1-p} - 1\right).$$

□

Let us come back to Lieb's concavity theorem. It now has a lot of proofs. We will give two here, the first one being Lieb's original proof using interpolation. For this, let us recall the three-line lemma first.

**Lemma 8.4.** *Let $S := \{z \in \mathbb{C} : 0 < \Re z < 1\}$ be the open strip and denote by $\overline{S}$ its closure. Suppose that $f : \overline{S} \to \mathbb{C}$ is bounded function such that*

1. *$f$ is analytic in $S$;*
2. *$f$ is continuous on $\overline{S}$;*
3. $\sup\{|f(k+iy)| : y \in \mathbb{R}\} := M_k < \infty, k = 0, 1.$

*Then for any $\theta \in [0,1]$, we have $|f(\theta)| \leq M_0^{1-\theta} M_1^{\theta}$.*

*Proof of Lieb's concavity theorem.* To prove the joint concavity of

$$(A, B) \mapsto \mathrm{Tr} A^p K^* B^{1-p} K, \quad 0 \leq p \leq 1,$$

it suffices to prove the concavity of

$$A \mapsto \mathrm{Tr} A^p K^* A^{1-p} K, \quad 0 \leq p \leq 1.$$

In fact, this is a doubling dimension trick:

$$\mathrm{Tr} A^p K^* B^{1-p} K = \mathrm{Tr} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}^p \begin{pmatrix} 0 & K^* \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}^{1-p} \begin{pmatrix} 0 & 0 \\ K & 0 \end{pmatrix}.$$

For any positive semi-definite $A_1, A_2$ and $\lambda \in (0,1)$, put $A := \lambda A_1 + (1-\lambda) A_2$. We need to show that

$$\lambda \mathrm{Tr} A_1^p K^* A_1^{1-p} K + (1-\lambda) \mathrm{Tr} A_2^p K^* A_2^{1-p} K \leq \mathrm{Tr} A^p K^* A^{1-p} K.$$

By approximation, we may assume that $A_1, A_2$ and $A$ are all positive definite. Set $M := A^{\frac{1-p}{2}} K A^{\frac{p}{2}}$. For $k = 1, 2$, consider the function

$$f_k(z) := \mathrm{Tr} A_k^z A^{-\frac{z}{2}} M^* A^{-\frac{1-z}{2}} A_k^{1-z} A^{-\frac{1-z}{2}} M A^{-\frac{z}{2}}, \quad z \in \overline{S}.$$

Then what we need to show can be reformulated as

$$\lambda f_1(p) + (1-\lambda) f_2(p) \leq \mathrm{Tr} M^* M.$$

We claim that the function $f_k$ is uniformly bounded on $\overline{S}$. In fact, denote $G_k(z) := A^{-\frac{z}{2}} A_k^z A^{-\frac{z}{2}}$ and we may write $f_k$ as $f_k(z) = \mathrm{Tr} M^* G_k(1-z) M G_k(z)$. By Cauchy–Schwarz,

$$|f_k(z)| \leq \left(\mathrm{Tr} M M^* G_k(1-z) G_k(1-\overline{z})\right)^{1/2} \left(\mathrm{Tr} M M^* G_k(\overline{z}) G_k(z)\right)^{1/2}.$$

For $z = x + iy$, we have

$$\mathrm{Tr} M M^* G_k(1-z) G_k(1-\overline{z}) \leq \|A^{-1}\|^{2x} \|A_k\|^{2x} \mathrm{Tr} M^* M,$$

which is uniformly bounded. Similarly $\mathrm{Tr} M M^* G_k(\overline{z}) G_k(z)$ is also uniformly bounded. Therefore, we finish the proof of the claim, so that we can use the three-line lemma to $f(z) := \lambda f_1(z) + (1-\lambda) f_2(z)$. When $\Re z = 0$:

$$f_k(0+iy) = \mathrm{Tr}\left( A_k^{\frac{iy}{2}} A^{-\frac{iy}{2}} M^* A^{\frac{-1+iy}{2}} A_k^{\frac{1}{2}} \cdot A_k^{\frac{1}{2}-iy} A^{\frac{-1+iy}{2}} M A^{-\frac{iy}{2}} A_k^{\frac{iy}{2}} \right).$$

By Cauchy–Schwarz:

$$|f_k(iy)| \leq \mathrm{Tr} M^* A^{\frac{-1+iy}{2}} A_k A^{-\frac{1+iy}{2}} M, \quad k = 1, 2.$$

So for any $y \in \mathbb{R}$

$$|f(iy)| \leq \mathrm{Tr} M^* A^{\frac{-1+iy}{2}} (\lambda A_0 + (1-\lambda) A_1) A^{-\frac{1+iy}{2}} M = \mathrm{Tr} M^* M.$$

Similarly we can prove $|f(1+iy)| \leq \mathrm{Tr} M^* M$ for all $y \in \mathbb{R}$. This concludes the proof by three-line lemma. □

Now we give another proof using perspective functions. We start with an even more general convexity/concavity result, which reduces most of the other results in this chapter to easy corollaries.

**Theorem 8.5** (ENG perspective theorem)**.** *Let $f\colon [0,\infty) \to \mathbb{R}$ be operator convex (resp. operator concave) and $g\colon (0,\infty) \to (0,\infty)$ operator concave. Assume that $f(0) \leq 0$ (resp. $f(0) \geq 0$). Then the map*

$$\begin{aligned} M_n(\mathbb{C})_{++} \times M_n(\mathbb{C})_+ &\to M_n(\mathbb{C}), \\ (A,B) &\mapsto g(A)^{1/2} f(g(A)^{-1/2} B g(A)^{-1/2}) g(A)^{1/2} \end{aligned}$$

*is jointly convex (resp. concave).*

*Proof.* We only prove the jointly convex case. The jointly concave case follows by replacing $f$ by $-f$.

Let $A_1, A_2 \in M_n(\mathbb{C})_{++}, B_2, B_2 \in M_n(\mathbb{C})_+$, $\lambda \in [0,1]$ and define $A = \lambda A_1 + (1-\lambda)A_2$, $B = \lambda B_1 + (1-\lambda)B_2$. Let $V_1 = (\lambda g(A_1))^{1/2} g(A)^{-1/2}$, $V_2 = ((1-\lambda) g(A_2))^{1/2} g(A)^{-1/2}$. Since $g$ is operator concave, we have

$$V_1^* V_1 + V_2^* V_2 = g(A)^{-1/2} (\lambda g(A_1) + (1-\lambda) g(A_2)) g(A)^{-1/2} \leq \mathbf{1}.$$

The operator Jensen inequality implies

$$\begin{aligned} & g(A)^{1/2} f(g(A)^{-1/2} B g(A)^{-1/2}) g(A)^{1/2} \\ &= g(A)^{1/2} f(V_1^* g(A_1)^{-1/2} B_1 g(A_1)^{-1/2} V_1 + V_2^* g(A_2)^{-1/2} B_2 g(A_2)^{-1/2} V_2) g(A)^{1/2} \\ &\leq g(A)^{1/2} V_1^* f(g(A_1)^{-1/2} B_1 g(A_1)^{-1/2}) V_1 g(A)^{1/2} \\ &\quad + g(A)^{1/2} V_2^* f(g(A_2)^{-1/2} B_2 g(A_2)^{-1/2}) V_2 g(A)^{1/2} \\ &= \lambda g(A_1)^{1/2} f(g(A_1)^{-1/2} B_1 g(A_1)^{-1/2}) g(A_1)^{1/2} \\ &\quad + (1-\lambda) g(A_2)^{1/2} f(g(A_2)^{-1/2} B_2 g(A_2)^{-1/2}) g(A_2). \end{aligned}$$

□

**Corollary 8.6.** *The function*

$$\Lambda_{p,q}\colon M_n(\mathbb{C})_{++} \times M_n(\mathbb{C})_+ \to M_n(\mathbb{C})_+, (A,B) \mapsto A^{q/2} (A^{-q/2} B A^{-q/2})^p A^{q/2}$$

*is jointly concave if $p, q \in [0,1]$ and jointly convex if $p \in [1,2]$ and $q \in [0,1]$.*

*Proof.* As $x \mapsto x^p$ is operator concave for $p \in [0,1]$ and operator convex for $p \in [1,2]$, the result follows immediately from the ENG perspective theorem. □

We are now in the position to prove Lieb's concavity theorem and Ando's convexity theorem.

*Proof of Theorem 8.1.* Equip $M_n(\mathbb{C})$ with the Hilbert–Schmidt inner product

$$\langle \cdot, \cdot \rangle_{\mathrm{HS}}\colon M_n(\mathbb{C}) \times M_n(\mathbb{C}) \to \mathbb{C},\ (A,B) \mapsto \mathrm{Tr}(A^* B),$$

making $M_n(\mathbb{C})$ into a Hilbert space. For $A, B \in M_n(\mathbb{C})$ define

$$L_A, R_B\colon M_n(\mathbb{C}) \to M_n(\mathbb{C}),\ L_A K = AK,\ R_B K = KB.$$

Note that $L_A$ and $R_B$ commute.

With this notation we have

$$\begin{aligned}\mathrm{Tr}(K^*A^pKB^{1-p}) &= \langle K, L_A^p R_B^{1-p} K\rangle_{\mathrm{HS}} \\ &= \langle K, L_A^{1/2}(L_A^{-1/2}R_B L_A^{-1/2})^{1-p} L_A^{1/2} K\rangle_{\mathrm{HS}} \\ &= \langle K, \Lambda_{1-p,1}(L_A, R_B)K\rangle_{\mathrm{HS}}\end{aligned}$$

and the joint convexity resp. joint concavity follows from the previous corollary. □

**Theorem 8.7.** *The operator function*

$$M_n(\mathbb{C})_{++} \times M_n(\mathbb{C})_+ \to M_n(\mathbb{C}) \otimes M_n(\mathbb{C}),\ (A,B) \mapsto A^p \otimes B^{1-p}$$

*is jointly concave when* $0 < p < 1$ *and jointly convex when* $-1 < p < 0$.

*Proof.* For $A, B \in M_n(\mathbb{C})$ let

$$S_A = A \otimes \mathbf{1},\ T_B = \mathbf{1} \otimes B.$$

Since $S_A$ and $T_B$ commute, we have

$$A^p \otimes B^{1-p} = S_A^p T_B^{1-p} = \Lambda_{1-p,1}(S_A, R_B),$$

and the joint convexity follows from the previous corollary. □

**Theorem 8.8.** *The geometric mean*

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

*is jointly concave.*

*Proof.* Since $f(x) = x^{1/2}$ and $g(x) = x$ are operator concave, and $f(0) = 0$, this result follows directly from the ENG perspective theorem. □

**Theorem 8.9.** *The harmonic mean*

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

*is jointly concave.*

*Proof.* Let $f(x) = (1 + x^{-1})^{-1} = \frac{x}{1+x} = 1 - \frac{1}{1+x}$ and $g(x) = x$. By Proposition 7.4, $f$ is operator concave. Clearly $f(0) = 0$. Since

$$\begin{aligned}M_{-1}(A,B) &= 2(A^{-1/2}(\mathbf{1} + A^{1/2}B^{-1}A^{1/2})A^{-1/2})^{-1} \\ &= 2A^{1/2}f(A^{-1/2}BA^{-1/2})A^{1/2},\end{aligned}$$

the result follows from the ENG perspective theorem. □

Recall that the arithmetic mean is given by

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

As in the scalar case, we have

**Theorem 8.10** (Arithmetic-geometric-harmonic mean inequality)**.** *For all positive definite matrices $A, B$, we have*

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

*Proof.* The first inequality is nothing but

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

which is equivalent to

$$\left(\frac{\mathbf{1}+A^{1/2}B^{-1}A^{1/2}}{2}\right)^{-1} \le (A^{-1/2}BA^{-1/2})^{1/2}.$$

This is true by the scalar inequality $(\frac{1+x}{2})^{-1} \le x^{-1/2}$ and the functional calculus.

The second inequality is

$$A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2} \le \frac{A+B}{2},$$

which is equivalent to

$$(A^{-1/2}BA^{-1/2})^{1/2} \le \frac{\mathbf{1}+A^{-1/2}BA^{-1/2}}{2}.$$

This follows from the scalar inequality $\sqrt{x} \le \frac{1+x}{2}$ and the functional calculus. □

We end with the following

**Theorem 8.11.** *The operator function*

$$(A,B) \mapsto B^*A^{-1}B$$

*is jointly convex.*

*Proof.* Let $A_1, A_2 \in M_n(\mathbb{C})_{++}$, $B_1, B_2 \in M_n(\mathbb{C})$ and $t \in [0,1]$. By the Schur complement theorem,

$$\begin{pmatrix} A_1 & B_1 \\ B_1^* & B_1^*A_1^{-1}B_1 \end{pmatrix} \ge 0,$$

and the same holds for $A_1$ and $B_1$ replaced by $A_2$ and $B_2$, respectively. Thus

$$\begin{pmatrix} tA_1+(1-t)A_2 & tB_1+(1-t)B_2 \\ tB_1^*+(1-t)B_2^* & tB_1^*A_1^{-1}B_1+(1-t)B_2^*A_2^{-1}B_2 \end{pmatrix} \ge 0.$$

Another application of the Schur complement theorem yields

$$\begin{aligned} &tB_1^*A_1^{-1}B_1+(1-t)B_2^*A_2^{-1}B_2 \\ &\quad \ge (tB_1^*+(1-t)B_2^*)(tA_1+(1-t)A_2)^{-1}(tB_1+(1-t)B_2). \end{aligned}$$

□

**Lemma 8.12.** *If $f(\cdot,\cdot)$ is jointly concave, then $\max_x f(x,y)$ is concave.*

*Proof.* For any $y_1, y_2$, suppose $x_i$ is such that $f(x_i, y_i) = \max_x f(x, y_i)$ for $i = 1, 2$. Then for any $\lambda \in (0, 1)$, we have

$$\begin{aligned}\max_x f(x, \lambda y_1 + (1-\lambda)y_2) &\geq f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\ &\geq \lambda f(x_1, y_1) + (1-\lambda) f(x_2, y_2) \\ &= \lambda \max_x f(x, y_1) + (1-\lambda) \max_x f(x, y_2).\end{aligned}$$

□

**Theorem 8.13.** *For any self-adjoint $H$, the function*

$$M_n(\mathbb{C})_{++} \to \mathbb{R},\ A \mapsto \operatorname{Tr} \exp(H + \log A)$$

*is concave.*

*Proof.* We have the following duality formula:

$$\operatorname{Tr} \exp(H + \log A) = \max_{X \geq 0} \operatorname{Tr} XH + \operatorname{Tr} X - \operatorname{Tr} X(\log X - \log A)$$

(exercise).

Then the desired concavity result follows from the joint convexity of quantum relative entropy and the above lemma. □

Using the above theorem, one can extend the Golden–Thompson inequality $\operatorname{Tr} e^{H+K} \leq \operatorname{Tr} e^H e^K$ to three matrices. Note that $\operatorname{Tr} e^H e^K e^L$ is in general not even a real number.

**Proposition 8.14.** *For all self-adjoint matrices $H, K, L$ we have*

$$\operatorname{Tr} e^{H+K+L} \leq \operatorname{Tr}[e^H T_{e^{-K}}(e^L)],$$

*where*

$$T_A(B) = \int_0^\infty \frac{1}{s+A} B \frac{1}{s+A} ds = \frac{d}{dt}|_{t=0} \log(A + tB).$$

*Proof.* Let $C \subset \mathbb{R}^N$ be a convex set such that $tx \in C$ for every $t > 0$, $x \in C$. If $f \colon C \to \mathbb{R}$ is concave and $f(tx) = tf(x)$ for all $t > 0$, $x \in C$, then

$$f(y) \leq \lim_{t \to 0^+} \frac{f(x+ty) - f(x)}{t},$$

for any $x, y \in C$.

Indeed, for any $t > 0$

$$f(x+ty) = (1+t) f\left(\frac{x}{1+t} + \frac{ty}{1+t}\right) \geq (1+t)\left[\frac{f(x)}{1+t} + \frac{tf(y)}{1+t}\right] = f(x) + tf(y).$$

Now we apply this result to $C = M_n(\mathbb{C})_{++}$ and $f(X) = \operatorname{Tr}[e^{H+K+\log X}]$. This function is concave by the previous theorem and homogeneous of degree one. For $X = e^{-K}$ and $Y = e^L$ we get

$$\operatorname{Tr} e^{H+K+L} \leq \frac{d}{dt}|_{t=0} \operatorname{Tr}[e^{H+K+\log(e^{-K}+te^L)}] = \operatorname{Tr}\left[e^{H+K+\log(e^{-K})} \frac{d}{dt}|_{t=0} \log(e^{-K} + te^L)\right],$$

where the right hand side is exactly what we need once we prove

$$\int_0^\infty \frac{1}{s+A} B \frac{1}{s+A} ds = \frac{d}{dt}|_{t=0} \log(A+tB).$$

This identity follows from the integral formula

$$\log A = \int_0^\infty \left( (s+\mathbf{1})^{-1} - (s+A)^{-1} \right) ds$$

as follows:

$$\begin{aligned} \frac{d}{dt}\Big|_{t=0} \log(X+tY) &= \frac{d}{dt}\Big|_{t=0} \int_0^\infty ((s+\mathbf{1})^{-1} - (s+X+tY)^{-1})\, ds \\ &= -\int_0^\infty \frac{d}{dt}\Big|_{t=0} (s+X+tY)^{-1}\, ds \\ &= \int_0^\infty (s+X)^{-1} Y (s+X)^{-1}. \end{aligned}$$

□

## Exercises

**Exercise 8.1.** Suppose that $p \leq q$. For $K \in M_n(\mathbb{C})$ the function

$$M_n(\mathbb{C})_+ \times M_n(\mathbb{C})_+ \to \mathbb{C},\ (A,B) \mapsto \mathrm{Tr} K^* A^p K B^q,$$

is

1. jointly concave if $0 \leq p \leq q \leq 1$ such that $p+q \leq 1$;
2. jointly convex if $-1 \leq p \leq 0, 1 \leq q \leq 2$ such that $p+q \geq 1$.

**Exercise 8.2.** Prove the duality formula:

$$\mathrm{Tr} \exp(H + \log A) = \max_{X \geq 0} \mathrm{Tr} XH + \mathrm{Tr} X - \mathrm{Tr} X(\log X - \log A).$$

**Exercise 8.3.** The parallel sum $A:B$ of $A, B \in M_n(\mathbb{C})_+$ is defined as

$$A:B = \lim_{\varepsilon \searrow 0} ((A+\varepsilon \mathbf{1})^{-1} + (B+\varepsilon \mathbf{1})^{-1})^{-1}.$$

1. Show that the limit in the definition of $A:B$ exists.
2. Show that
$$\langle \xi, (A^{-1}+B^{-1})^{-1} \xi \rangle = \inf\{ \langle \eta, A\eta \rangle + \langle \zeta, B\zeta \rangle \mid \xi = \eta + \zeta \}$$
for all $\xi \in \mathbb{C}^n$.
3. Show that $S^*(A:B)S \leq (S^*AS) : (S^*BS)$ for all $S \in M_n(\mathbb{C})$.

**Exercise 8.4.** A quantum Markov semigroup is a family $(P_t)_{t \geq 0}$ of linear operators on $M_n(\mathbb{C})$ such that

- $P_0 = \mathrm{id}, P_s P_t = P_{s+t}, s, t \geq 0$,
- $P_t$ is unital completely positive trace preserving,

- $P_t x \to x, t \to 0$.

It has a generator $L$ that is defined via

$$L(x) := \lim_{t \to 0} \frac{x - P_t x}{t}.$$

Show that $\rho \mapsto \langle L(\rho^p), \rho^{1-p} \rangle$ is convex when $0 \leq p \leq 1$, and concave when $-1 \leq p \leq 0$.

# Chapter 9

# Entanglement

If two classical physical systems are described by the (finite) pure state spaces $X$ and $Y$, then the composite system is described the pure state space $X \times Y$. This means that the mixed states of the composite system are probability densities on $X \times Y$. For every $\rho\colon X \times Y \to [0,1]$ with $\sum_{x,y} \rho(x,y) = 1$ one has $\rho = \sum_{x,y} \rho(x,y)\mathbf{1}_{(x,y)}$. In other words, every probability density on $X \times Y$ is a convex combination of the Dirac densities $\mathbf{1}_{(x,y)}$ with $x \in X$, $y \in Y$.

The situation is markedly different for quantum systems, where the phenomenon of entanglement occurs, which is one of the key features of quantum information theory compared to classical information theory.

**Definition 9.1.** A quantum state $\rho \in M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ is called *separable* if there exist $\lambda_1, \dots, \lambda_k \in [0,1]$ with $\sum_{j=1}^k \lambda_j = 1$ and quantum states $\sigma_1^{(1)}, \dots, \sigma_k^{(1)} \in M_m(\mathbb{C})$, $\sigma_1^{(2)}, \dots, \sigma_k^{(2)} \in M_n(\mathbb{C})$ such that

$$\rho = \sum_{j=1}^{k} \lambda_j \sigma_j^{(1)} \otimes \sigma_j^{(2)}.$$

Every quantum state that is not separable is called *entangled*.

**Remark.** The notion of separable and pure states applies to states of a composite system for a given composition. For example, if we view $M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ as $M_{mn}(\mathbb{C}) \otimes M_1(\mathbb{C})$, then every quantum state is naturally separable.

Examples of separable states are easy to come by – just take your favorite quantum states in $M_m(\mathbb{C})$ and $M_n(\mathbb{C})$ and then form their tensor product and take convex combinations if you like. What is less obvious is how to find entangled states (or if they exist at all). For this purpose, the following criterion comes in handy.

**Proposition 9.2** (Horodecki criterion)**.** *A quantum state $\rho \in M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ is separable if and only if for every $k \in \mathbb{C}$ and every positive map $\Phi\colon M_m(\mathbb{C}) \to M_k(\mathbb{C})$ the matrix $(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(\rho)$ is positive.*

*Proof.* We only prove the easier implication here.

If $\rho$ is separable, then there exists $\lambda_1, \dots, \lambda_l \geq 0$ and quantum states $\sigma(1)_1, \dots, \sigma_l^{(1)} \in M_m(\mathbb{C})$, $\sigma_1^{(2)}, \dots, \sigma_l^{(2)} \in M_n(\mathbb{C})$ such that $\rho = \sum_j \lambda_j \sigma_j^{(1)} \otimes \sigma_j^{(2)}$.

If $\Phi\colon M_m(\mathbb{C}) \to M_k(\mathbb{C})$ is positive, then

$$(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(\rho) = \sum_{j=1}^{l} \lambda_j \Phi(\sigma_j^{(1)}) \otimes \sigma_j^{(2)}.$$

Since $\Phi$ is positive, the matrix $\Phi(\sigma_j^{(1)})$ is positive. Thus $(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(\rho) \geq 0$. □

**Corollary 9.3.** *Whenever $m, n \geq 2$, there exist entangled states in $M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$.*

*Proof.* Let $\Phi\colon M_m(\mathbb{C}) \to M_k(\mathbb{C})$ be a positive map that is not 2-positive. For example, we can take $k = m$ and $\Phi$ the transpose map. Then there exists a (necessarily non-zero) positive matrix $A \in M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ such that $(\Phi \otimes \mathrm{id}_{M_n(\mathbb{C})})(A)$ is not positive. By the Horodecki criterion, $A/\mathrm{Tr}(A)$ is an entangled state. □

**Example 9.4** (Werner states)**.** Let $m = n \geq 2$ and $W = \sum_{i,j} E_{ij} \otimes E_{ji}$. As $W^2 = \mathbf{1}$, the matrix $W$ has eigenvalues $\pm 1$. Let $P_{\pm 1}$ be the orthogonal projection onto the eigenspace of $W$ corresponding to the eigenvector $\pm 1$. More explicitly, $P_1 = \frac{1}{2}(1 \otimes 1 + W)$ and $P_{-1} = \frac{1}{2}(1 \otimes 1 - W)$.

A basis of the range of $P_1$ is given by $(e_i \otimes e_j + e_j \otimes e_j)_{i \leq j}$ and a basis of the range of $P_{-1}$ is given by $(e_i \otimes e_j - e_j \otimes e_i)_{i<j}$. Thus $\mathrm{Tr}(P_1) = \frac{n(n+1)}{2}$ and $\mathrm{Tr}(P_{-1}) = \frac{n(n-1)}{2}$.

A quantum state of the form $\rho_\lambda = \frac{2\lambda}{n(n-1)} P_{-1} + \frac{2(1-\lambda)}{n(n+1)} P_1$ with $\lambda \in [0,1]$ is called a *Werner state*. Werner states are entangled states. We only prove this here for $\lambda > \frac{1}{2}$.

Let $\Phi\colon M_m(\mathbb{C}) \to M_m(\mathbb{C})$ be the transpose map, which is positive, but not completely positive. We have

$$(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(W) = \sum_{i,j} E_{ij} \otimes E_{ij}$$

and therefore

$$(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(P_1) = \frac{1}{2}(1 \otimes 1) + \frac{1}{2} \sum_{i,j} E_{ij} \otimes E_{ij},$$
$$(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(P_{-1}) = \frac{1}{2}(1 \otimes 1) - \frac{1}{2} \sum_{i,j} E_{ij} \otimes E_{ij}.$$

Let $Q_0 = \frac{1}{n} \sum_{i,j} E_{ij} \otimes E_{ij}$ and $Q_1 = 1 \otimes 1 - Q_0$. From the previous identities we deduce

$$\begin{aligned}(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(\rho_\lambda) &= (\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})}) \left( \frac{2\lambda}{n(n-1)} P_{-1} + \frac{2(1-\lambda)}{n(n+1)} P_1 \right) \\ &= \frac{2\lambda - 1 + n}{n(n^2-1)} (1 \otimes 1) + \frac{(2\lambda - 1)n + 1}{n(n^2-1)} \sum_{i,j} E_{i,j} \otimes E_{ij} \\ &= \frac{1 - 2\lambda}{n} Q_0 + \left( 1 - \frac{1 - 2\lambda}{n} \right) \frac{Q_1}{n^2 - 1}.\end{aligned}$$

Observe that $Q_0$ is self-adjoint and

$$Q_0^2 = \frac{1}{n^2} \sum_{i,j,k,l} E_{ij} E_{kl} \otimes E_{ij} E_{kl} = \frac{1}{n^2} \sum_k \sum_{i,l} E_{il} \otimes E_{il} = Q_0.$$

Hence $Q_0$ and $Q_1$ are orthogonal projections with $Q_0 Q_1 = 0$. It follows that $(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(\rho_\lambda)$ has eigenvalues $(1 - 2\lambda)/n$ and $(1 - \frac{1-2\lambda}{n})/(n^2 - 1)$. In particular, $(\Phi \otimes \mathrm{id}_{M_k(\mathbb{C})})(\rho_\lambda)$ is not positive for $\lambda > \frac{1}{2}$.

# Chapter 10

# Data processing inequalities

**Definition 10.1** (Quantum relative entropy)**.** For any quantum states $\rho$ and $\sigma$, the quantum relative entropy of $\rho$ with respect to $\sigma$ is

$$D(\rho||\sigma) := \mathrm{Tr}(\rho(\log\rho - \log\sigma)).$$

Although the quantum relative entropy is not a distance, it still serves as a nice measure to distinguish quantum states.

**Lemma 10.2.** *If $\rho, \sigma \in M_n(\mathbb{C})$ are quantum states, then $D(\rho\|\sigma) \geq 0$ with equality if $\rho = \sigma$.*

*Proof.* Let $f(x) = x\log x$. Since $f$ is convex, Klein's inequality implies

$$0 \leq \mathrm{Tr}(f(\rho) - f(\sigma) - f'(\sigma)(\rho-\sigma)) = \mathrm{Tr}(\rho\log\rho - \sigma\log\sigma - \log\sigma(\rho-\sigma)) = \mathrm{Tr}(\rho(\log\rho - \log\sigma)). \ \square$$

In fact, the quantum relative entropy vanishes $D(\rho||\sigma) = 0$ if and only if $\rho = \sigma$. The converse implication follows from the equality case in Klein's inequality for strictly convex functions, which we did not discuss. However, it is also an immediate consequence of the following result. For its formulation, recall that the trace norm of a matrix $A \in M_m(\mathbb{C})$ is defined as $\|A\|_1 = \mathrm{Tr}(|A|)$.

**Theorem 10.3** (Pinsker's inequality)**.** *For any quantum states $\rho$ and $\sigma$, we have*

$$D(\rho||\sigma) \geq \frac{1}{2}\|\rho - \sigma\|_1^2.$$

To prove this, we shall need the following *monotonicity* property, sometimes called *data processing inequality* of quantum relative entropy.

**Theorem 10.4** (Data processing inequality for quantum relative entropy)**.** *For any quantum states $\rho, \sigma \in M_m(\mathbb{C})$ and any quantum channel $\Lambda\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$, we have*

$$D(\Lambda(\rho)||\Lambda(\sigma)) \leq D(\rho||\sigma).$$

*Proof.* As noticed before,

$$D(\rho\|\sigma) = \lim_{p\nearrow 1} \frac{1 - \mathrm{Tr}(\rho^p\sigma^{1-p})}{1-p}.$$

Let $f_p(x) = x^p$ for $p \in [0,1]$ and

$$L_\sigma\colon M_n(\mathbb{C}) \to M_n(\mathbb{C}),\ L_\sigma A = \sigma A$$
$$R_\rho\colon M_n(\mathbb{C}) \to M_n(\mathbb{C}),\ R_\rho A = A\rho$$

Then

$$\langle \sigma^{1/2}, f_p(L_\rho R_\sigma^{-1})\sigma^{1/2}\rangle_{\mathrm{HS}} = \mathrm{Tr}(\rho^p \sigma^{1-p}).$$

For convenience, write $\Delta_{\rho,\sigma} = L_\rho R_\sigma^{-1}$ and $\Delta_{\Lambda(\rho),\Lambda(\sigma)} = L_{\Lambda(\rho)} R_{\Lambda(\sigma)}^{-1}$ . Consider the map

$$V\colon M_n(\mathbb{C}) \to M_m(\mathbb{C}),\ A \mapsto \Lambda^\dagger(A\Lambda(\sigma)^{-1/2})\sigma^{1/2}.$$

Then $V(\Lambda(\sigma)^{1/2}) = \Lambda^\dagger(\mathbf{1})\sigma^{1/2} = \sigma^{1/2}$, where we used the trace-preserving property of $\Lambda$. So we may write

$$\mathrm{Tr}(\rho^p\sigma^{1-p}) = \langle \sigma^{1/2}, f_p(\Delta_{\rho,\sigma})\sigma^{1/2}\rangle_{\mathrm{HS}} = \langle \Lambda(\sigma)^{1/2}, V^\dagger f_p(\Delta_{\rho,\sigma})V(\Lambda(\sigma)^{1/2})\rangle_{\mathrm{HS}}.$$

Note that for any $A \in M_n(\mathbb{C})$, we have by Kadison–Schwarz inequality,

$$\begin{aligned}
\langle A, V^\dagger V(A)\rangle_{\mathrm{HS}} &= \langle V(A), V(A)\rangle_{\mathrm{HS}} \\
&= \langle \Lambda^\dagger(A\Lambda(\sigma)^{-1/2})\sigma^{1/2}, \Lambda^\dagger(A\Lambda(\sigma)^{-1/2})\sigma^{1/2}\rangle_{\mathrm{HS}} \\
&\le \mathrm{Tr}[\Lambda^\dagger(\Lambda(\sigma)^{-1/2}A^*A\Lambda(\sigma)^{-1/2})\sigma] \\
&= \langle A, A\rangle_{\mathrm{HS}},
\end{aligned}$$

and

$$\begin{aligned}
\langle A, V^\dagger \Delta_{\rho,\sigma} V(A)\rangle_{\mathrm{HS}} &= \langle \Lambda^\dagger(A\Lambda(\sigma)^{-1/2})\sigma^{1/2}, \Delta_{\rho,\sigma}\Lambda^\dagger(A\Lambda(\sigma)^{-1/2})\sigma^{1/2}\rangle_{\mathrm{HS}} \\
&= \langle \Lambda^\dagger(A\Lambda(\sigma)^{-1/2}), \rho\Lambda^\dagger(A\Lambda(\sigma)^{-1/2})\rangle_{\mathrm{HS}} \\
&\le \mathrm{Tr}[\Lambda^\dagger(A\Lambda(\sigma)^{-1}A^*)\sigma] \\
&= \mathrm{Tr}[A\Lambda(\sigma)^{-1}A^*\Lambda(\rho)] \\
&= \langle A, \Delta_{\Lambda(\rho),\Lambda(\sigma)}(A)\rangle_{\mathrm{HS}}.
\end{aligned}$$

So we have $V^\dagger V \le \mathbf{1}$ and $V^\dagger \Delta_{\rho,\sigma} V \le \Delta_{\Lambda(\rho),\Lambda(\sigma)}$. Then we get

$$V^\dagger f_p(\Delta_{\rho,\sigma})V \le f_p(V^\dagger \Delta_{\rho,\sigma} V) \le f_p(\Delta_{\Lambda(\rho),\Lambda(\sigma)}),$$

where in the first inequality we used operator Jensen's inequality ($f_p$ is operator concave since it is operator monotone), and in the second inequality we used the operator monotonicity. Therefore,

$$\begin{aligned}
D(\rho\|\sigma) &= \lim_{p\nearrow 1} \frac{1-\mathrm{Tr}(\rho^p\sigma^{1-p})}{1-p} \\
&= \lim_{p\nearrow 1} \frac{1-\langle \Lambda(\sigma)^{1/2}, V^\dagger f_p(\Delta_{\rho,\sigma})V\Lambda(\sigma)^{1/2}\rangle_{\mathrm{HS}}}{1-p} \\
&\ge \frac{1-\langle \Lambda(\sigma)^{1/2}, f_p(\Delta_{\Lambda(\rho),\Lambda(\sigma)})(\Lambda(\sigma)^{1/2})\rangle_{\mathrm{HS}}}{1-p} \\
&= \lim_{p\nearrow 1} \frac{1-\mathrm{Tr}(\Lambda(\rho)^p\Lambda(\sigma)^{1-p})}{1-p} \\
&= D(\Lambda(\rho)\|\Lambda(\sigma)).
\end{aligned}$$

□

**Definition 10.5.** A vector $p \in \mathbb{C}^n$ is called *probability vector* if $p_j \ge 0$ for $1 \le j \le n$ and $\sum_{j=1}^n p_j = 1$. A *classical channel* is a linear map $\Phi\colon \mathbb{C}^m \to \mathbb{C}^n$ that maps probability vectors to probability vectors.

If $p, q \in \mathbb{C}^n$ are probability vectors, the *relative entropy of $p$ with respect to $q$* is defined as

$$D(p\|q) = \sum_{j=1}^n p_j \log \frac{p_j}{q_j}.$$

**Theorem 10.6.** *Let $p, q \in \mathbb{C}^m$ be probability vectors.*

*(a) If $\Phi\colon \mathbb{C}^m \to \mathbb{C}^n$ is a classical channel, then*

$$D(\Phi(p)\|\Phi(q)) \le D(p\|q).$$

*(b) We have the Pinsker's inequality*

$$D(p\|q) \ge \frac{1}{8}\|p-q\|_1^2.$$

*Proof.* (a) If $\rho = \operatorname{diag}(p)$, $\sigma = \operatorname{diag}(q)$, then $D(\rho\|\sigma) = D(p\|q)$. Moreover, let $E\colon M_m(\mathbb{C}) \to \mathbb{C}^m$, $E(A) = (A_{11}, \dots, A_{mm})$. Then the map $\Lambda\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$, $\Lambda(A) = \operatorname{diag}(\Phi(E(A)))$ is a quantum channel. It follows from the data processing inequality for quantum entropies that

$$D(\Phi(p)\|\Phi(q)) = D(\operatorname{diag}(\Phi(E(\rho))), \operatorname{diag}(\Phi(E(\sigma)))) = D(\Lambda(\rho)\|\Lambda(\sigma)) \le D(\rho\|\sigma) = D(p\|q).$$

(b) As $x \mapsto x\log x$ has second derivative $\frac{1}{x}$, the Taylor expansion at 1 gives

$$x\log x = x - 1 + \sum_{k=0}^{\infty} \frac{(x-1)^{k+2}}{(k+2)!} \frac{d^k}{dx^k}\bigg|_{x=1} \frac{1}{x}.$$

Since $x\log x$ is convex, we have $x\log x \ge x-1$. Moreover, if $x \le 1$, then all terms in the Taylor expansion are positive. Thus

$$x\log x \ge (x-1) + \frac{1}{2}(1-x)_+^2.$$

It follows that

$$\begin{aligned} D(p\|q) &= \sum_{j=1}^n \frac{p_j}{q_j} \log\left(\frac{p_j}{q_j}\right) q_j \\ &\ge \sum_{j=1}^n \left(\frac{p_j}{q_j} - 1\right) q_j + \frac{1}{2}\sum_{j=1}^n \left(1 - \frac{p_j}{q_j}\right)_+^2 q_j \\ &\ge \frac{1}{2}\left(\sum_{j=1}^n \left(1 - \frac{p_j}{q_j}\right)_+ q_j\right)^2. \end{aligned}$$

Since $\sum_j (1 - p_j/q_j)p_j = 0$, we have $\sum_j (1-p_j/q_j)_+ q_j = \sum_j (1-p_j/q_j)_- q_j$. Thus

$$\|p-q\|_1 = \sum_{j=1}^n \left|1 - \frac{p_j}{q_j}\right| q_j = 2\sum_{j=1}^n \left(1 - \frac{p_j}{q_j}\right)_+ q_j \le 2\sqrt{2} D(p\|q)^{1/2}. \qquad \square$$

**Remark.** The constant $\frac{1}{8}$ in Pinsker's inequality can be improved to $\frac{1}{2}$, but that requires a bit more work or some slick probabilistic arguments.

**Theorem 10.7.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ be convex, $A \in M_m(\mathbb{C})$ self-adjoint and $\Phi\colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ a completely positive map with $\Phi(\mathbf{1}) \le \mathbf{1}$. If $f(0) \le 0$ or $\Phi(\mathbf{1}) = \mathbf{1}$, then*

$$\operatorname{Tr}(f(\Phi(A))) \le \operatorname{Tr}(\Phi(f(A))).$$

*Proof.* Let $A = \sum_{k=1}^{l} \lambda_k P_k$ be the spectral decomposition of $A$. For an eigenvector $\xi$ of $\Phi(A)$ with $\|\xi\| = 1$ define $\mu_k = \langle \xi, \Phi(P_k)\xi \rangle$ for $1 \leq k \leq l$ and $\mu_{l+1} = 1 - \sum_{k=1}^{l} \mu_k$. Clearly, $\mu_k \geq 0$ for $k \leq l$, and since $\sum_k P_k = \mathbf{1}$ and $\Phi(\mathbf{1}) \leq \mathbf{1}$, we also have $\mu_{l+1} \geq 0$.

Using the convexity of $f$, we get

$$\begin{aligned}
\langle \xi, f(\Phi(A))\xi \rangle &= f(\langle \xi, \Phi(A)\xi \rangle) \\
&= f\left( \sum_k \lambda_k \langle \xi, \Phi(P_k)\xi \rangle \right) \\
&= f\left( \sum_{k=1}^{l} \lambda_k \mu_k + 0 \cdot \mu_{k+1} \right) \\
&\leq \sum_{k=1}^{l} \mu_k f(\lambda_k) + \mu_{l+1} f(0).
\end{aligned}$$

If $\Phi(\mathbf{1}) = \mathbf{1}$, then $\mu_{l+1} = 0$, and if $f(0) \leq 0$, then $\mu_{l+1} f(0) \leq 0$. In either case, we get

$$\langle \xi, f(\Phi(A))\xi \rangle \leq \sum_{k=1}^{l} \mu_k f(\lambda_k) = \sum_k \langle \xi, \Phi(f(\lambda_k) P_k)\xi \rangle = \langle \xi, \Phi(f(A))\xi \rangle.$$

Summing over an orthonormal eigenbasis for $\Phi(A)$, the desired inequality follows.

In the case when □

For the following recall that a POVM is a family $(P_i)_{i=1}^n$ of positive matrices with $\sum_i P_i = \mathbf{1}$ and that for every POVM $(P_i)_{i=1}^n$ the map

$$\Phi \colon M_m(\mathbb{C}) \to M_n(\mathbb{C}),\ A \mapsto \sum_{i=1}^{k} \mathrm{Tr}(P_i A) E_{ii}$$

is a quantum channel.

**Lemma 10.8.** *Let $\rho, \sigma \in M_m(\mathbb{C})$ be two quantum states and $\Lambda \colon M_m(\mathbb{C}) \to M_n(\mathbb{C})$ any quantum channel.*

*(a) The trace distance is monotone under quantum channels:*

$$\|\Lambda(\rho) - \Lambda(\sigma)\|_1 \leq \|\rho - \sigma\|_1.$$

*(b) There exists a POVM $(P_i)_{i=1}^n$ with associated quantum-to-classical channel $\Phi$ such that*

$$\|\rho - \sigma\|_1 = \|\Phi(\rho) - \Phi(\sigma)\|_1 = \sum_{i=1}^{n} |\mathrm{Tr}(P_i \rho) - \mathrm{Tr}(P_i \sigma)|.$$

*Proof.* (a) Let $A = \rho - \sigma$. Since $f(x) = |x|$ is convex and $f(0) = 0$, we deduce from the previous theorem

$$\|\Lambda(A)\|_1 = \mathrm{Tr}(|\Lambda(A)|) \leq \mathrm{Tr}(\Lambda(|A|)) = \mathrm{Tr}(|A|) = \|A\|_1.$$

(b) Consider the spectral decomposition of $\rho - \sigma = \sum_j \lambda_j Q_j$. Then

$$P_1 := \sum_{\lambda_j \geq 0} Q_j, \quad P_2 := \sum_{\lambda_j < 0} Q_j$$

give a POVM $P = (P_i)_{i=1}^2$. By definition,

$$\|\Phi(\rho) - \Phi(\sigma)\|_1 = \sum_{i=1}^{2} |\mathrm{Tr}(P_i\rho) - \mathrm{Tr}(P_i\sigma)| = \left|\sum_{\lambda_j \geq 0} \lambda_j\right| + \left|\sum_{\lambda_j < 0} \lambda_j\right| = \sum_j |\lambda_j| = \|\rho - \sigma\|_1. \qquad \square$$

Now we are ready to prove the quantum Pinsker's inequality.

*Proof of Theorem 10.3.* We only prove the weaker version with constant $\frac{1}{8}$ instead of $\frac{1}{2}$. Take $\Lambda$ as the quantum-to-classical channel in the previous lemma. Then from the monotonicity of quantum relative entropy and classical Pinsker's inequality:

$$D(\rho\|\sigma) \geq D(\Lambda(\rho)\|\Lambda(\sigma)) \geq \frac{1}{8}\|\Lambda(\rho) - \Lambda(\sigma)\|_1^2 = \frac{1}{8}\|\rho - \sigma\|_1^2. \qquad \square$$

Recall that $S(\rho) = -\mathrm{Tr}(\rho \log \rho)$ is the quantum entropy. For any bipartite state $\rho$ over $H_1 \otimes H_2$, we denote $\rho_1 = \mathrm{Tr}_2(\rho)$ and $\rho_2 = \mathrm{Tr}_1(\rho)$. We have seen the following subadditivity result of entropy in the previous lectures:

$$S(\rho) \leq S(\rho_1) + S(\rho_2),$$

which follows from the non-negativity of quantum relative entropy:

$$S(\rho_1) + S(\rho_2) - S(\rho) = \mathrm{Tr}\rho(\log \rho - \log(\rho_1 \otimes \rho_2)) = D(\rho||\rho_1 \otimes \rho_2) \geq 0.$$

Moreover, the equality $S(\rho) = S(\rho_1) + S(\rho_2)$ holds iff $\rho = \rho_1 \otimes \rho_2$.

Actually, the quantum entropy satisfies the following *strong subadditivity* (SSA). For a multipartite state $\rho \in M_{l_1}(\mathbb{C}) \otimes \cdots \otimes M_{l_n}(\mathbb{C}) \otimes M_n(\mathbb{C})$ we use the notation $\rho_{j_1 \ldots j_N}$ to denote the state $\mathrm{Tr}_{k_1} \ldots \mathrm{Tr}_{k_M}(\rho)$, where $k_1, \ldots, k_M$ are chosen such that $\{j_1, \ldots, j_N\} \sqcup \{k_1, \ldots, k_M\} = \{1, \ldots, n\}$. In particular, if $\rho$ is a tripartite state, then $\rho_{123} = \rho$ etc. Note that this notation depends on the splitting of our quantum system into subsystems.

**Theorem 10.9** (Strong subadditivity of the quantum entropy)**.** *If $\rho \in M_l(\mathbb{C}) \otimes M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ is a quantum state, then*

$$S(\rho_{12}) + S(\rho_{23}) \geq S(\rho_{123}) + S(\rho_2).$$

*This inequality reduces to the subadditivity of the quantum entropy when $m = 1$.*

*Proof.* Similar to the computations in the proof of the subadditivity of the quantum entropy, one obtains

$$\begin{aligned} D(\rho_{123}\|\rho_{12} \otimes \rho_3) &= \mathrm{Tr}(\rho_{123}(\log \rho_{123} - \log \rho_{12} \otimes 1 - 1 \otimes \log \rho_3)) \\ &= -S(\rho_{123}) - \mathrm{Tr}(\mathrm{Tr}_3(\rho_{123} \log \rho_{12} \otimes 1)) - \mathrm{Tr}(\mathrm{Tr}_{12}(\rho_{123} 1 \otimes \log \rho_3)) \\ &= -S(\rho_{123}) - \mathrm{Tr}(\rho_{12} \log \rho_{12}) - \mathrm{Tr}(\rho_3 \otimes \rho_3) \\ &= S(\rho_{12}) + S(\rho_3) - S(\rho_{123}) \end{aligned}$$

Now consider the quantum channel $\Lambda = \mathrm{Tr}_1$. We have

$$\Lambda(\rho_{123}) = \rho_{23}, \quad \Lambda(\rho_{12} \otimes \rho_3) = \rho_2 \otimes \rho_3.$$

By a similar computation as above,

$$D(\Lambda(\rho_{123})\|\Lambda(\rho_{12} \otimes \rho_3)) = S(\rho_2) + S(\rho_3) - S(\rho_{23}).$$

Thus it follows from the data processing inequality that

$$\begin{aligned} S(\rho_{12}) + S(\rho_3) - S(\rho_{123}) &= D(\rho_{123} \| \rho_{12} \otimes \rho_3) \\ &\geq D(\Lambda(\rho_{123}) \| \Lambda(\rho_{12} \otimes \rho_3)) \\ &= S(\rho_2) + S(\rho_3) - S(\rho_{23}). \end{aligned}$$ □

**Definition 10.10.** For any tripartite state $\rho_{123}$ the *conditional mutual information of 1 and 2 given 3,* $I(1,2|3)$, is defined as

$$I(1,2|3) := S(\rho_{13}) + S(\rho_{23}) - S(\rho_{123}) - S(\rho_3).$$

Then SSA says $I(1,2|3) \geq 0$. For any bipartite state $\rho_{12}$ the *squashed entanglement* of $\rho_{12}$ is defined as

$$E_{sq}(\rho_{12}) := \frac{1}{2} \inf\{I(1,2|3) : \rho_{123} \text{ is any tripartite extension of } \rho_{12}\}.$$

The functional $E_{sq}$ provides a faithful measure of entanglement.

**Theorem 10.11.** *A bipartite state $\rho_{12}$ is separable if and only if $E_{sq}(\rho_{11}) = 0$.*

So if $E_{sq}(\rho) > 0$, then $\rho$ is entangled. The following extended SSA provides a lower bound of $E_{sq}$:

**Theorem 10.12** (Extended SSA)**.** *For any tripartite state $\rho_{123} \in M_l(\mathbb{C}) \otimes M_m(\mathbb{C}) \otimes M_n(\mathbb{C})$ we have*

$$S(\rho_{13}) + S(\rho_{23}) - S(\rho_{123}) - S(\rho_3) \geq 2 \max\{S(\rho_1) - S(\rho_{12}), S(\rho_2) - S(\rho_{12}), 0\}.$$

*As a corollary,*

$$E_{sq}(\rho_{12}) \geq \max\{S(\rho_1) - S(\rho_{12}), S(\rho_2) - S(\rho_{12}), 0\}.$$

*So if either of the conditional entropies $S(\rho_{12}) - S(\rho_1)$ or $S(\rho_{12}) - S(\rho_2)$ is strictly negative, then $E_{sq}(\rho_{12}) > 0$ and thus $\rho_{12}$ is entangled.*

To prove this theorem, we need a purification trick that is very useful. Let us come back to the quantum entropy $S(\rho)$. It is clear that $S(\rho) = 0$ iff $\rho$ is pure. If $\rho = \rho_{12}$ is a bipartite state, then (exercise) $S(\rho_{12}) = 0$ implies $S(\rho_1) = S(\rho_2)$. The following theorem says that if $S(\rho_{12})$ is small, then $S(\rho_1)$ is close to $S(\rho_2)$:

**Theorem 10.13.** *For any bipartite state $\rho_{12}$ we have*

$$|S(\rho_1) - S(\rho_2)| \leq S(\rho_{12}).$$

For the proof, we recall the following purification result from Chapter 4: If $\rho \in M_n(\mathbb{C})$ is a quantum state, then there exists a unit vector $\xi \in \mathbb{C}^n \otimes \mathbb{C}^n$ such that $\rho = \mathrm{Tr}_1(|\xi\rangle\langle\xi|) = \mathrm{Tr}_2(|\xi\rangle\langle\xi|)$. This is not quite the statement of Proposition 4.6, but can be deduced from its proof.

Now we can prove Theorem 10.13 as follows:

*Proof of Theorem 10.13.* Consider a purification $\rho_{123}$ of $\rho_{12}$ as described above. Then $S(\rho_{12}) = S(\rho_3)$ and $S(\rho_1) = S(\rho_{23})$. By the additivity of quantum entropy:

$$S(\rho_1) = S(\rho_{23}) \leq S(\rho_2) + S(\rho_3) = S(\rho_2) + S(\rho_{12}).$$

So $S(\rho_{12}) \geq S(\rho_1) - S(\rho_2)$, and the proof is finished by symmetry. □

*Proof of Theorem 10.12.* Consider any purification $\rho_{1234}$ of $\rho_{123}$. Since $\rho_{1234}$ is pure, $S(\rho_{14}) = S(\rho_{23})$ and $S(\rho_{124}) = S(\rho_3)$. Then

$$S(\rho_{12}) + S(\rho_{23}) - S(\rho_1) - S(\rho_3) = S(\rho_{12}) + S(\rho_{14}) - S(\rho_1) - S(\rho_{124}) \geq 0.$$

So

$$S(\rho_{12}) + S(\rho_{23}) \geq S(\rho_1) + S(\rho_3).$$

Similarly we have

$$S(\rho_{13}) + S(\rho_{23}) \geq S(\rho_1) + S(\rho_2).$$

Adding the above two inequalities, we get

$$S(\rho_{12}) + S(\rho_{13}) + 2S(\rho_{23}) \geq 2S(\rho_1) + S(\rho_2) + S(\rho_3).$$

This is independent of $H_4$. Consider again the purification $\rho_{1234}$ of $\rho_{123}$, then the above argument shows:

$$S(\rho_{14}) + S(\rho_{13}) + 2S(\rho_{34}) \geq 2S(\rho_1) + S(\rho_4) + S(\rho_3).$$

Since $\rho_{1234}$ is pure, $S(\rho_{14}) = S(\rho_{23}), S(\rho_{34}) = S(\rho_{12})$ and $S(\rho_4) = S(\rho_{123})$. Hence,

$$S(\rho_{23}) + S(\rho_{13}) + 2S(\rho_{12}) \geq 2S(\rho_1) + S(\rho_{123}) + S(\rho_3),$$

which is nothing but

$$S(\rho_{13}) + S(\rho_{23}) - S(\rho_{123}) - S(\rho_3) \geq 2S(\rho_1) - 2S(\rho_{12}).$$

Similarly, we can derive a lower bound in terms of $2S(\rho_2) - 2S(\rho_{12})$. □

## Exercises

**Exercise 10.1.** Prove the classical Pinsker inequality (you can use the monotonicity of classical relative entropy). Let $p, q$ be two probability densities over some finite set $\mathcal{X}$. We have

$$D(p||q) \geq \frac{1}{2}\|p - q\|_1^2.$$

**Exercise 10.2.** Suppose that $\{\Lambda_x\}_{x\in\mathcal{X}}$ is a POVM, i.e. a finite set of operators such that

$$\Lambda_x \geq 0, \forall x \in \mathcal{X}, \text{ and } \sum_x \Lambda_x = \mathbf{1}.$$

Show that this gives a quantum channel $\Lambda$ such that for any quantum state $\rho$, $\Lambda(\rho)$ is a classical probability density over $\mathcal{X}$ satisfying $\Lambda(\rho)(x) = \mathrm{Tr}(\rho\Lambda_x)$.

**Exercise 10.3.** Show that for any quantum channel $\Lambda$ and any matrix $X$, we have $\|\Lambda(X)\|_p \leq \|X\|_p, 1 \leq p \leq \infty$.

**Exercise 10.4.** Find an entangled state.

**Exercise 10.5.** Show that for any pure bipartite state $\rho$ over $H_1 \otimes H_2$, we have $S(\rho_1) = S(\rho_2)$. Show also that any quantum state can be purified. That is, for any quantum state $\rho$ over $H$, there exists a pure state $|\psi\rangle\langle\psi|$ on $H \otimes H$ such that

$$\rho = \mathrm{Tr}_1[|\psi\rangle\langle\psi|] = \mathrm{Tr}_2[|\psi\rangle\langle\psi|].$$

**Exercise 10.6.** Show that the monotonicity of quantum relative entropy implies the joint convexity.

# Chapter 11

# Supplement: Quantum Markov Semigroups and Logarithmic Sobolev Inequalities

**Definition 11.1.** A quantum Markov semigroup (QMS) on $M_n(\mathbb{C})$ is a family $(P_t)_{t\geq 0}$ of unital completely positive maps on $M_n(\mathbb{C})$ such that

- $P_0 = \mathrm{id}_{M_n(\mathbb{C})}$, $P_s P_t = P_{st}$ for all $s, t \geq 0$,
- $P_t \to P_0$ as $t \to 0$.

**Remark.** By the Heisenberg–Schrödinger duality, if $(P_t)$ is a QMS, then $P_t^\dagger$ is completely positive trace-preserving for all $t \geq 0$. In particular, $P_t^\dagger$ maps quantum states to quantum states for all $t \geq 0$.

**Theorem 11.2** (Lindblad)**.** *If $(P_t)$ is a quantum Markov semigroup on $M_n(\mathbb{C})$, then for each $A \in M_n(\mathbb{C})$ the limit*

$$\mathcal{L}(A) = \lim_{t \searrow 0} \frac{1}{t}(A - P_t(A))$$

*exists, $\mathcal{L}$ is a linear map from $M_n(\mathbb{C})$ to itself and $P_t = e^{-t\mathcal{L}}$.*

*Moreover, if $\mathcal{L}\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$ is a linear map, then $(e^{-t\mathcal{L}})$ is a quantum Markov semigroup if and only if there exist $G \in M_n(\mathbb{C})$ and a completely positive map $\Phi\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$ with $\Phi(\mathbf{1}) = G + G^*$ such that*

$$\mathcal{L}(A) = GA + AG^* - \Phi(A)$$

*for all $A \in M_n(\mathbb{C})$.*

**Remark.** For a linear map $\mathcal{L}\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$, the exponential $e^{-t\mathcal{L}}$ is defined as

$$e^{-t\mathcal{L}} = \sum_{k=0}^{\infty} \frac{(-1)^k t^k}{k!} \mathcal{L}^k.$$

*Proof of Lindblad's theorem.* Clearly, the set $V = A \in M_n(\mathbb{C}) \mid \lim_{t\searrow 0} \frac{1}{t}(A - P_t(A))$ exists$\}$ is a subspace of $M_n(\mathbb{C})$. For $A \in M_n(\mathbb{C})$ and $\delta > 0$ let

$$A_\delta = \int_0^\delta P_t(A)\, dt.$$

Since $t \mapsto P_t(A)$ is continuous, we have $\delta^{-1}A_\delta \to A$ as $\delta \to 0$. Moreover,

$$\begin{aligned}\frac{1}{t}(A_\delta - P_t(A_\delta)) &= \frac{1}{t}\left(\int_0^\delta P_s(A)\,ds - \int_0^\delta P_{s+t}(A)\,ds\right)\\ &= \frac{1}{t}\left(\int_0^\delta P_s(A)\,ds - \int_t^{t+\delta} P_s(A)\,ds\right)\\ &= \frac{1}{t}\int_0^t P_s(A)\,ds - \frac{1}{t}\int_\delta^{t+\delta} P_s(A)\,ds\\ &\to A - P_\delta(A)\end{aligned}$$

as $t \to 0$. Thus $A_\delta \in V$, and it follows that $V$ is dense in $M_n(\mathbb{C})$. Since $M_n(\mathbb{C})$ is finite-dimensional, we must have $V = M_n(\mathbb{C})$.

Now let $A \in M_n(\mathbb{C})$. We want to show that $P_t(A) = e^{-t\mathcal{L}}(A)$. Note that

$$\frac{1}{h}(P_{t+h}(A) - P_t(A)) = P_t\left(\frac{1}{h}(P_h(A) - A)\right) \to -P_t(\mathcal{L}(A))$$

as $h \to 0$. In other words, the map $t \mapsto P_t(A)$ solves the initial-value problem

$$\begin{aligned}\frac{d}{dt}A(t) &= -\mathcal{L}A(t)\\ A(0) &= A.\end{aligned}$$

A direct computation shows that $t \mapsto e^{-t\mathcal{L}}(A)$ solves the same IVP. It follows from the uniqueness theorem for ordinary differential equations that $P_t(A) = e^{-t\mathcal{L}}(A)$ for all $t \geq 0$.

For the second part we first assume that $(e^{-t\mathcal{L}})$ is a quantum Markov semigroup. Let $U(n) = \{U \in M_n(\mathbb{C}) \mid U^*U = UU^* = \mathbf{1}\}$. There exists a unique probability measure $\mu$ on $U(n)$ such that

$$\int_{U(n)} f(UVW)\,d\mu(V) = \int_{U(n)} f(V)\,d\mu(V)$$

for all $U, W \in U(n)$ and all continuous $f\colon U(n) \to \mathbb{C}$.

Let $G = \int_{U(n)} \mathcal{L}(U^*)U\,d\mu(U)$ and $\Phi(A) = GA + AG^* - \mathcal{L}(A)$. If $V \in M_n(\mathbb{C})$ is unitary, then

$$\int_{U(n)} \mathcal{L}(VU^*)U\,d\mu(U) = \int_{U(n)} \mathcal{L}((UV^*)^*)UV^*\,d\mu(U)V = \int_{U(n)} \mathcal{L}(U^*)U\,\mu(U)V = GV,$$

where we used the invariance property of $\mu$.

Since every element of $M_n(\mathbb{C})$ is a linear combination of four unitary matrices, we conclude

$$\int_{U(n)} \mathcal{L}(AU^*)U\,d\mu(U) = GA$$

for all $A \in M_n(\mathbb{C})$.

Thus

$$\begin{aligned}\Phi(A^*A) &= (GA^*)A + A^*(GA^*)^* - \mathcal{L}((UA)^*UA)\\ &= \int_{U(n)} (\mathcal{L}((UA)^*)UA + (UA)^*\mathcal{L}(UA) - \mathcal{L}((UA)^*(UA)))\,d\mu(U).\end{aligned}$$

Note that

$$\mathcal{L}((UA)^*)UA + (UA)^*\mathcal{L}(UA) - \mathcal{L}((UA)^*(UA)) = \lim_{t\to 0}\frac{1}{t}(P_t((UA)^*(UA)) - P_t(UA)^*P_t(UA)) \geq 0$$

by the Kadison–Schwarz inequality. Thus $\Phi$ is positive.

If we replace $\mathcal{L}$ by $\mathcal{L}\otimes \mathrm{id}_{M_k(\mathbb{C})}$, then $G$ is replaced by $G\otimes \mathbf{1}_k$ and $\Phi$ by $\Phi\otimes \mathrm{id}_{M_k(\mathbb{C})}$. Since $e^{-t(\mathcal{L}\otimes \mathrm{id})} = P_t\otimes \mathrm{id}$ again satisfies the Kadison–Schwarz inequality, the argument from above implies that $\Phi\otimes \mathrm{id}_{M_k(\mathbb{C})}$ is positive for all $k\in\mathbb{N}$.

Finally,

$$\mathcal{L}(\mathbf{1}) = \lim_{t\to 0}\frac{1}{t}(\mathbf{1} - P_t(\mathbf{1})) = 0.$$

Hence

$$\Phi(\mathbf{1}) = G\mathbf{1} + \mathbf{1}G^* - \mathcal{L}(\mathbf{1}) = G + G^*. \qquad \square$$

**Remark.** The unique probability measure $\mu$ on $U(n)$ such that

$$\int_{U(n)} f(UVW)\,d\mu(V) = \int_{U(n)} f(V)\,d\mu(V)$$

for all $U, W\in U(n)$ and all continuous $f\colon U(n)\to\mathbb{C}$ is called the (normalized) *Haar measure* on $U(n)$. More generally, a Haar measure exists for any compact group (and any locally compact group if one drops the assumption that the measure be finite). In general, there is no explicit formula for it.

**Remark.** By duality one obtains that $t\mapsto P_t^\dagger(A)$ is also differentiable for all $A\in M_n(\mathbb{C})$ and $\frac{d}{dt}P_t^\dagger(A) = -\mathcal{L}^\dagger(P_t^\dagger(A))$.

**Definition 11.3.** If $(P_t)_{t\geq 0}$ is a quantum Markov semigroup on $M_n(\mathbb{C})$, then the unique linear map $\mathcal{L}\colon M_n(\mathbb{C})\to M_n(\mathbb{C})$ such that $e^{-t\mathcal{L}} = P_t$ for all $t\geq 0$ is called the *generator* of $(P_t)$.

**Corollary 11.4** (Gorini–Kossakowski–Lindblad–Sudarshan)**.** *If $(P_t)$ is a quantum Markov semigroup on $M_n(\mathbb{C})$ with generator $\mathcal{L}$, then there exist $H\in M_n(\mathbb{C})_{\mathrm{sa}}$ and $V_1,\dots,V_m\in M_n(\mathbb{C})$ such that*

$$\mathcal{L}(A) = i[H,A] + \sum_{j=1}^m\left(\frac{1}{2}V_j^*V_jA + \frac{1}{2}AV_j^*V_j - V_j^*AV_j\right)$$

*for all $A\in M_n(\mathbb{C})$.*

*Proof.* By Lindblad's theorem, there exist $G\in M_n(\mathbb{C})$ and $\Phi\colon M_n(\mathbb{C})\to M_n(\mathbb{C})$ completely positive such that $\Phi(\mathbf{1}) = G + G^*$ and $\mathcal{L}(A) = GA + AG^* - \Phi(A)$ for all $A\in M_n(\mathbb{C})$. By Kraus' theorem, there exist $V_1,\dots,V_m\in M_n(\mathbb{C})$ such that $\Phi(A) = \sum_{j=1}^m V_j^*AV_j$ for all $A\in M_n(\mathbb{C})$. Let $H = \frac{1}{2i}(G - G^*)$.

Then we have $G = \frac{1}{2}\Phi(\mathbf{1}) + iH$ and thus

$$\begin{aligned}\mathcal{L}(A) &= GA + AG^* - \Phi(A)\\ &= iHA - iAH + \frac{1}{2}\Phi(\mathbf{1})A + \frac{1}{2}A\Phi(\mathbf{1}) - \Phi(A)\\ &= i[H,A] + \frac{1}{2}\sum_{j=1}^m(V_j^*V_jA + AV_j^*V_j) - \sum_{j=1}^m V_j^*AV_j. \qquad \square\end{aligned}$$

**Remark.** One calls an operator $\mathcal{L}$ of this form a Lindbladian and $A \mapsto i[H, A]$ the conservative (or Hamiltonian part) and $A \mapsto \sum_{j=1}^m \left(\frac{1}{2}V_j^* V_j A + \frac{1}{2}AV_j^* V_j - V_j^* A V_j\right)$ the dissipative part of $\mathcal{L}$. The matrices $V_j$ are called jump operators.

**Theorem 11.5.** *Let $(P_t)$ be a quantum Markov semigroup on $M_n(\mathbb{C})$, $\sigma \in M_n(\mathbb{C})$ a full-rank quantum state such that $P_t^\dagger(\sigma) = \sigma$ for all $t \geq 0$ and $\alpha \geq 0$. The following conditions are equivalent:*

*(i) $D(P_t^\dagger(\rho)\|\sigma) \leq e^{-\alpha t} D(\rho\|\sigma)$ for all quantum states $\rho \in M_n(\mathbb{C})$ and $t \geq 0$,*

*(ii) $\alpha D(\rho\|\sigma) \leq \mathrm{Tr}(\mathcal{L}^\dagger(\rho)(\log\rho - \log\sigma))$ for all full-rank quantum states $\rho \in M_n(\mathbb{C})$.*

*Proof.* (i) $\Longrightarrow$ (ii): Let $f(t) = e^{\alpha t} D(P_t^\dagger(\rho)\|\sigma)$. By (i), $f(t) \leq f(0)$ for all $t \geq 0$. We have

$$f'(t) = \alpha f(t) + e^{\alpha t}\frac{d}{dt}\mathrm{Tr}(P_t^\dagger(\rho)(\log P_t^\dagger(\rho) - \log\sigma)).$$

To compute $\frac{d}{dt}\mathrm{Tr}(P_t^\dagger(\rho)\log P_t^\dagger(\rho))$, we can use a similar argument as in the section on monotonicity of trace functionals to see that

$$\frac{d}{dt}\mathrm{Tr}(P_t^\dagger(\rho)\log P_t^\dagger(\rho)) = -\mathrm{Tr}((\log P_t^\dagger(\rho) + \mathbf{1})\mathcal{L}^\dagger(P_t^\dagger(\rho))) = -\mathrm{Tr}(\mathcal{L}^\dagger(P_t^\dagger(\rho))\log P_t^\dagger(\rho)),$$

where we used that

$$\mathrm{Tr}(\mathbf{1}\mathcal{L}^\dagger(P_t^\dagger(\rho))) = \mathrm{Tr}(\mathcal{L}(\mathbf{1})P_t^\dagger(\rho)) = 0.$$

Clearly, $\frac{d}{dt}\mathrm{Tr}(P_t^\dagger(\rho)\log\sigma) = -\mathrm{Tr}(\mathcal{L}^\dagger(P_t^\dagger(\rho))$. Thus

$$f'(t) = \alpha f(t) - e^{\alpha t}\mathrm{Tr}(\mathcal{L}^\dagger(P_t^\dagger(\rho))(\log P_t^\dagger(\rho) - \log\sigma)).$$

In particular,

$$0 \leq f'(0) = \alpha D(\rho\|\sigma) - \mathrm{Tr}(\mathcal{L}^\dagger(\rho)(\log\rho - \log\sigma)).$$

(ii) $\Longrightarrow$ (i): Again let $f(t) = e^{\alpha t} D(P_t^\dagger(\rho)\|\sigma)$. We have seen above that

$$f'(t) = e^{\alpha t}(\alpha D(P_t^\dagger(\rho)\|\sigma) - \mathrm{Tr}(\mathcal{L}^\dagger(P_t^\dagger(\rho))(\log P_t^\dagger(\rho) - \log\sigma))).$$

By (ii), $f'(t) \leq 0$ for all $t \geq 0$. Hence

$$D(\rho\|\sigma) = f(0) \geq f(t) = e^{\alpha} D(P_t^\dagger(\rho)\|\sigma).$$ □

**Example 11.6** (Depolarizing semigroup)**.** Let $\sigma \in M_n(\mathbb{C})$ be a full-rank quantum state and $E(A) = \mathrm{Tr}(A\sigma)\mathbf{1}$. Then the operators $P_t = e^{-t}\mathrm{id}_{M_n(\mathbb{C})} + (1-e^{-t})E$, $t \geq 0$, form a quantum Markov semigroup with generator $\mathcal{L} = \mathrm{id} - E$. This semigroup is called the (generalized) *depolarizing semigroup*.

Moreover, $P_t^\dagger(\sigma) = \sigma$ and

$$D(P_t^\dagger(\rho)\|\sigma) \leq e^{-t} D(\rho\|\sigma).$$

Indeed, $P_t$ is unital completely positive as convex combination of two unital completely positive maps. Moreover,

$$\begin{aligned}
P_s(P_t(A)) &= P_s(e^{-t}A + (1-e^{-t})E(A)) \\
&= e^{-t}(e^{-s}A + (1-e^{-s})E(A)) + (1-e^{-t})(e^{-s}E(A) + (1-e^{-s})(E^2(A))) \\
&= e^{-(s+t)}A + (e^{-t}(1-e^{-s}) + (1-e^{-t}))E(A) \\
&= e^{-(s+t)}A + (1-e^{-(s+t)})E(A) \\
&= P_{s+t}(A).
\end{aligned}$$

The property $P_0 = \mathrm{id}$ and the continuity of $t \mapsto P_t$ are clear. Note moreover that $P_t^\dagger(A) = e^{-t}A + (1-e^{-t})\mathrm{Tr}(A)\sigma$.

To see that exponential decay of the relative entropy, recall that $D$ is convex. Thus

$$\begin{aligned} D(P_t^\dagger(\rho)\|\sigma) &= D(e^{-t}\rho + (1-e^{-t})\sigma\|\sigma) \\ &\leq e^{-t}D(\rho\|\sigma) + (1-e^{-t})D(\sigma\|\sigma) \\ &= e^{-t}D(\rho\|\sigma). \end{aligned}$$

## Exercises

**Exercise 11.1.** Find a Lindblad form for the QMS from Example 11.6 in the case $\sigma = \mathbf{1}/n$.

**Exercise 11.2.** Let $(P_t)$ be a quantum Markov semigroup with generator $\mathcal{L}$. Show that $\rho \mapsto \mathrm{Tr}(\mathcal{L}(\rho^p)\rho^{1-p})$ is convex when $0 \leq p \leq 1$, and concave when $-1 \leq p \leq 0$.