Mathematical Introduction to Quantum Information Processing

(growing lecture notes, 2019/2022)

Michael M. Wolf

Mar 2023

Contents

These are (incomplete but hopefully growing) lecture notes of courses taught in 2019 and 2022 at the Department of Mathematics at the Technical University of Munich.

Chapter 1

Mathematical framework

1.1 Hilbert spaces

This section will briefly summarize relevant concepts and properties of Hilbert spaces.

A complex Hilbert space is a vector space over the complex numbers, equipped with an inner product $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ and an induced norm $\|\psi\| := \langle \psi, \psi \rangle^{1/2}$ w.r.t. which it is complete.[1] Hence, every Hilbert space is in particular a Banach space. We will use the physicists' convention that the inner product is linear in the second and conjugate-linear in its first argument so that $\langle \psi, c\varphi \rangle = c\langle \psi, \varphi \rangle = \langle \overline{c}\psi, \varphi \rangle$, $\forall c \in \mathbb{C}$.

The most import inequality for the inner product is the *Cauchy-Schwarz inequality*, which immediately follows[2] from the identity

$$\|\psi\|^2 \|\varphi\|^2 - |\langle \psi, \varphi \rangle|^2 = \frac{1}{\|\varphi\|^2} \left\| \|\varphi\|^2 \psi - \langle \varphi, \psi \rangle \varphi \right\|^2 \geq 0.$$

This also shows that equality holds iff φ and ψ are linearly dependent.

A characteristic property of any norm that is induced by an inner product is that it satisfies the *parallelogram law*

$$\|\psi + \varphi\|^2 + \|\psi - \varphi\|^2 = 2\left(\|\psi\|^2 + \|\varphi\|^2\right). \tag{1.1}$$

In fact, whenever a norm satisfies Eq.(1.1) for all ψ, φ, then we can reconstruct a corresponding inner product via the *polarization identity*, which in the case of a complex space reads

$$\langle \psi, \varphi \rangle = \frac{1}{4} \sum_{k=0}^{3} i^k \left\| \varphi + i^k \psi \right\|^2. \tag{1.2}$$

[1] That is, every Cauchy sequence converges.

[2] Note that the derivation of Cauchy-Schwarz does not use that $\langle \psi, \psi \rangle = 0 \Rightarrow \psi = 0$. It only requires that $\langle \psi, \psi \rangle \geq 0$.

A central concept that is enabled by an inner product is *orthogonality*: ψ, φ are called orthogonal if $\langle \psi, \varphi \rangle = 0$. In that case $\|\psi + \varphi\| = \|\psi - \varphi\|$ so that the parallelogram law becomes the *Pythagoras identity* $\|\psi + \varphi\|^2 = \|\psi\|^2 + \|\varphi\|^2$. For any subset $S \subseteq \mathcal{H}$ the *orthogonal complement* $S^\perp$ is defined as the subset of $\mathcal{H}$ whose elements are orthogonal to every element in S. $S^\perp$ is then necessarily a closed linear subspace. Every closed linear subspace $\mathcal{H}_1 \subseteq \mathcal{H}$, in turn, gives rise to a unique decomposition of any element $\psi \in \mathcal{H}$ as $\psi = \psi_1 + \psi_2$, where $\psi_1 \in \mathcal{H}_1$ and $\psi_2 \in \mathcal{H}_2 = \mathcal{H}_1^\perp$. In this way, the Hilbert space decomposes into an orthogonal direct sum $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$. The ψ_i's can equivalently be characterized as those elements in $\mathcal{H}_i$ closest to ψ. Uniqueness of the ψ_i's enables the definition of two *orthogonal projections* $P_i : \mathcal{H} \to \mathcal{H}_i, \psi \mapsto \psi_i$, which are linear idempotent maps related via $P_2 = \mathbb{1} - P_1$, where $\mathbb{1}$ denotes the identity map on $\mathcal{H}$.

If we think the idea of orthogonal decompositions of a Hilbert space further, we are led to the concept of an *orthonormal basis*. An orthonormal basis is a set $\{e_i\} \subseteq \mathcal{H}$ whose linear span is dense in $\mathcal{H}$ and whose elements satisfy $\langle e_i, e_j \rangle = \delta_{ij}$. Its cardinality defines the *dimension* of the Hilbert space. *Separability* of $\mathcal{H}$ means that there is a countable orthonormal basis. In that case, for every $\psi \in \mathcal{H}$ we have $\psi = \sum_i \langle e_i, \psi \rangle e_i$ (converging in norm) and the *Parseval identity* $\|\psi\|^2 = \sum_i |\langle e_i, \psi \rangle|^2$ holds. An orthonormal set of vectors can always be extended to an orthonormal basis.

Another property that Hilbert spaces share with their Euclidean ancestors is expressed by the *Riesz representation theorem*: it states that every continuous linear map from $\mathcal{H}$ into $\mathbb{C}$ is of the form $\psi \mapsto \langle \varphi, \psi \rangle$ for some $\varphi \in \mathcal{H}$, and vice versa. In other words, there is a conjugate linear bijection between $\mathcal{H}$ and its *topological dual space* $\mathcal{H}'$ (i.e. the space of all continuous linear functionals).

The possible identification of $\mathcal{H}$ and $\mathcal{H}'$ motivates the so-called *Dirac-notation* that writes $|\psi\rangle$ for elements of $\mathcal{H}$ and $\langle \varphi|$ for elements of $\mathcal{H}'$. These symbols are then called *ket* and *bra*, respectively and the inner product in this notation reads $\langle \varphi | \psi \rangle$ (forming a "bra(c)ket"). When we would restrict ourselves to Euclidean spaces, kets and bras would be nothing but column vectors and row vectors, respectively. Dirac notation also enables the introduction of a *ket-bra* $|\psi\rangle\langle\varphi| : \mathcal{H} \to \mathcal{H}$ that defines a map $|\phi\rangle \mapsto |\psi\rangle\langle\varphi|\phi\rangle$. Using ket-bras, a necessary and sufficient condition for a set of orthonormal vectors to form a basis of a separable Hilbert space is given by

$$\sum_k |e_k\rangle\langle e_k| = \mathbb{1}. \tag{1.3}$$

To write expressions of this form even more compactly, the elements of a fixed orthonormal basis are often simply specified by their label so that one writes $|k\rangle$ instead of $|e_k\rangle$.

So far, this has been abstract Hilbert space theory. Before we proceed, some concrete examples of Hilbert spaces:

Example 1.1. $\mathbb{C}^n$ becomes a Hilbert space when equipped with the standard inner product $\langle \psi, \varphi \rangle = \sum_{i=1}^n \overline{\psi_i} \varphi_i$.

Example 1.2. The sequence space $l_2(\mathbb{N}) := \left\{\psi \in \mathbb{C}^{\mathbb{N}} \middle| \sum_k |\psi_k|^2 < \infty\right\}$ becomes a Hilbert space when equipped with the standard inner product $\langle \psi, \varphi \rangle = \sum_k \overline{\psi_k} \varphi_k$. The standard orthonormal basis in this case is given by sequences $e_k, k \in \mathbb{N}$ such that the l'th element in e_k equals δ_{lk}.

Example 1.3. The function space $L_2(\mathbb{R}) := \left\{f : \mathbb{R} \to \mathbb{C} \middle| \int_{\mathbb{R}} |\psi(x)|^2 dx < \infty\right\} / \sim$ where $\psi \sim \varphi \Leftrightarrow \int_{\mathbb{R}} |\psi(x) - \varphi(x)|^2 dx = 0$ becomes a separable Hilbert space with $\langle \psi, \varphi \rangle = \int_{\mathbb{R}} \overline{\psi}(x) \varphi(x) dx$.

Example 1.4. The space $\mathbb{C}^{n \times m}$ of complex $n \times m$ matrices becomes a Hilbert space with $\langle A, B \rangle = \operatorname{tr}[A^* B]$.

Two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are called *isomorphic* if there is a bijection $U : \mathcal{H}_1 \to \mathcal{H}_2$ that preserves all inner products. U, which is called a *Hilbert space isomorphism*, is then necessarily linear and it turns out that Hilbert spaces are isomorphic iff they have the same dimension. Hence, all separable Hilbert spaces are isomorphic to either $\mathbb{C}^n$ or $l_2(\mathbb{N})$, in particular, $L_2(\mathbb{R}) \simeq l_2(\mathbb{N})$.

Sometimes one has to deal with inner product spaces that are not complete. In these cases the following theorem comes in handy and allows to 'upgrade' every such space to a Hilbert space:

Theorem 1.1 (Completion theorem)**.** *For every inner product space $\mathcal{X}$ there is a Hilbert space $\mathcal{H}$ and a linear map $V : \mathcal{X} \to \mathcal{H}$ that preserves all inner products*[3] *so that $V(\mathcal{X})$ is dense in $\mathcal{H}$ and equal to $\mathcal{H}$ if $\mathcal{X}$ is complete. The space $\mathcal{H}$ is then called the* completion *of $\mathcal{X}$. It is unique in the sense that if $(V', \mathcal{H}')$ give rise to another completion, then there is a Hilbert space isomorphism $U : \mathcal{H} \to \mathcal{H}'$ s.t. $V' = U \circ V$.*

As in the more general case of metric spaces, the completion is constructed by considering equivalence classes of Cauchy-sequences in $\mathcal{X}$. Usually, this construction is, however, hardly used beyond the proof of this theorem, and it is sound to regard $\mathcal{H}$ as a superspace of $\mathcal{X}$ that has been constructed from $\mathcal{X}$ by adding all the elements that were missing for completeness.

We finally state two properties that distinguish Hilbert spaces from other normed spaces. The first one is a distant relative of the completion theorem and allows to extend Lipschitz-maps while preserving the Lipschitz constant:

Theorem 1.2 (Kirszbraun)**.** *Let $\mathcal{H}, \mathcal{H}'$ be two Hilbert spaces and $A \subset \mathcal{H}$. Then every L-Lipschitz map $f : A \to \mathcal{H}'$ has an extension $F : \mathcal{H} \to \mathcal{H}'$ (i.e., $F|_A = f$) that is also L-Lipschitz.*

The second property that is a specialty of inner product spaces is often used in various applications in the form of a *dimension reduction* argument:

Lemma 1.3 (Johnson-Lindenstrauss)**.** *There is a universal constant $c \in \mathbb{R}$ such that for any $\epsilon \in (0, 1]$, Hilbert space $\mathcal{H}$, $n \in \mathbb{N}$, $\psi_1, \ldots, \psi_n \in \mathcal{H}$ there is a linear map $L : \mathcal{H} \to \mathcal{H}_d$ that is a multiple of an orthogonal projection onto a d-dimensional subspace $\mathcal{H}_d$ with*

$$d \leq \frac{c}{\epsilon^2} \log n,$$

[3] In other words, V is an *isometry*; see next section for the definition.

so that for all $i, j \in \{1, \dots, n\}$:

$$(1-\epsilon)\left\|\psi_i - \psi_j\right\|^2 \leq \left\|L\psi_i - L\psi_j\right\|^2 \leq (1+\epsilon)\left\|\psi_i - \psi_j\right\|^2 . \tag{1.4}$$

Clearly, the map L that occurs in the Johnson-Lindenstrauss Lemma can not also approximately preserve all distances outside the set $\{\psi_i\}_{i=1}^n$. However, if one allows L to be non-linear, then a suitable choice for L also approximately preserves the distances between all pairs of points where one is taken from the given n-point set and the other one is arbitrary (cf. [1]).

The Johnson-Lindenstrauss Lemma is often stated and used for real Hilbert spaces, but is equally valid for complex ones.

From now on, we will tacitly assume that all Hilbert spaces $\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2$, *etc. are complex and separable.*

Exercise 1.1. Show that the closed unit ball of any Hilbert space is strictly convex.

Exercise 1.2. Show that any linear map $U : \mathcal{H}_1 \to \mathcal{H}_2$ that preserves norms also preserves inner products.

Exercise 1.3. a) Prove that $\psi = \varphi$ iff $\forall \phi \in \mathcal{H} : \langle \phi, \varphi \rangle = \langle \phi, \psi \rangle$.

b) Let $A : \mathcal{H} \to \mathcal{H}$ be linear, $\psi, \varphi \in \mathcal{H}$. Verify the identity

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

c) Let $A, B : \mathcal{H} \to \mathcal{H}$ be linear. Show that $A = B$ iff $\forall \psi \in \mathcal{H} : \langle \psi, A\psi \rangle = \langle \psi, B\psi \rangle$. Why is this not true for real Hilbert spaces?

Exercise 1.4. Prove that every separable, infinite-dimensional Hilbert space is isomorphic to $l_2(\mathbb{N})$.

Exercise 1.5. Let $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}^d$, $r_1, \dots, r_n \in [0, \infty)$ and denote by $B_{r_i}(x_i)$ the closed Euclidean ball with radius r_i around x_i. Assume that $\bigcap_i B_{r_i}(y_i) = \emptyset$ and that $\forall i, j : \|y_i - y_j\| \leq \|x_i - x_j\|$ holds for the Euclidean norm. Show that then $\bigcap_i B_{r_i}(x_i) = \emptyset$ holds as well.

Notes and literature Frigyes Riesz, David Hilbert, and Hilbert's student Erhard Schmidt studied various aspects of concrete Hilbert spaces, (mainly in the context of integral equations or for $l_2(\mathbb{N})$) in the first years of the 20'th century. The introduction of a geometric viewpoint, which led to the concept of orthogonality, is largely due to Schmidt. The term *Hilbert space* was coined by Frigyes Riesz for concrete Hilbert spaces and it was later used by John von Neumann for the underlying abstract concept. Herman Weyl introduced the name *unitary space* in parallel. Von Neumann, who included separability in the definition of a Hilbert space, used the concept to unify Schrödinger's *wave mechanics* with the *matrix mechanics* of Werner Heisenberg, Pascual Jordan and Max Born. An impetus of von Neumann's work were lectures given by David Hilbert in the winter term 1926/27 on the development of quantum mechanics. Von Neumann attended the lectures and quickly established a rigorous mathematical basis of what he had heard. Soon after, this led to the foundational work "*Über die Grundlagen der Quantenmechanik*" [2].
A good way to learn about the mathematics of Hilbert spaces is from Paul Halmos' "*A Hilbert space problem book*" [3].

1.2 Hilbert space operators

With *operator* we mean a linear map between vector spaces. If these, say $\mathcal{X}$ and $\mathcal{Y}$, are Banach spaces, we define $\mathcal{B}(\mathcal{X},\mathcal{Y})$ to be the set of continuous operators from $\mathcal{X}$ to $\mathcal{Y}$, and $\mathcal{B}(\mathcal{X}) := \mathcal{B}(\mathcal{X},\mathcal{X})$. $\mathcal{B}(\mathcal{X},\mathcal{Y})$ itself becomes a Banach space when equipped with the *operator norm* $||A|| := \sup_{x\neq 0} \|Ax\| \,/\, \|x\|$. So by definition, the operator norm is the smallest Lipschitz-constant of the operator. The use of the letter $\mathcal{B}$ already suggests an elementary but crucial fact: an operator between Banach spaces is continuous iff it is bounded (in the sense that its operator norm is finite).

A commonly used procedure is the extension of a bounded operator: if $A \in \mathcal{B}(L,\mathcal{Y})$ is defined on a dense linear subspace $L \subseteq \mathcal{X}$, then by the *BLT theorem* (for 'bounded linear transformation') there exists a unique extension $\tilde{A} \in \mathcal{B}(\mathcal{X},\mathcal{Y})$ of $A = \tilde{A}|_L$. In addition, $\|\tilde{A}\| = \|A\|$. This is often used when defining a bounded operator by first specifying its action on a set whose linear span is dense in $\mathcal{X}$ and then using that "by linearity and continuity" this extends uniquely to the whole space.

We will encounter various types of operators on Hilbert spaces:

Definition 1.4. *Let $A \in \mathcal{B}(\mathcal{H})$, $C \in \mathcal{B}(\mathcal{H}_1,\mathcal{H}_2)$.*

(i) The adjoint *$C^* \in \mathcal{B}(\mathcal{H}_2,\mathcal{H}_1)$ is defined via $\langle\psi, C\varphi\rangle =: \langle C^*\psi,\varphi\rangle$ $\forall\psi,\varphi$.*

(ii) If $A^*A = AA^*$, then A is called normal.

(iii) If $A^* = A$, then A is called Hermitian.[4]

(iv) C is called an isometry *if $C^*C = \mathbb{1}$ and a* unitary *if in addition $CC^* = \mathbb{1}$.*

(v) C is called a partial isometry *if it is an isometry on* $\ker(C)^\perp$.

(vi) If $\langle\psi, A\psi\rangle \geq 0$ $\forall\psi \in \mathcal{H}$, then A is called positive *(a.k.a.* positive semidefinite*) and we write $A \geq 0$.*

(vii) If $A^2 = A$, then A is called a projection *and an* orthogonal projection*, if in addition $A = A^*$.*

In the physics literature, A^* is often written $A^\dagger$. The adjoint operation is an involution, i.e., $(A^*)^* = A$, it preserves the operator norm $\|A^*\| = \|A\|$ and satisfies $(AB)^* = B^*A^*$. When representing the adjoint operator as a matrix in a given orthonormal basis we see that the adjoint equals the complex conjugate of the transpose since $\langle e_k, A^*e_l\rangle = \overline{\langle A^*e_l, e_k\rangle} = \overline{\langle e_l, Ae_k\rangle}$.

Example 1.5 (Pauli matrices)**.** The *Pauli matrices*

$$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.5}$$

are all Hermitian and unitary. Together with $\sigma_0 := \mathbb{1}$ they form a basis of the space of 2×2 matrices.

[4] The term *self-adjoint* is used as well.

Example 1.6 (Diagonal operators)**.** An operator $A \in \mathcal{B}(\mathcal{H})$ is called *diagonal* if there is an orthonormal basis $\{e_k\} \subseteq \mathcal{H}$ of eigenvectors with corresponding eigenvalues $\lambda_k \in \mathbb{C}$ such that $A = \sum_k \lambda_k |e_k\rangle\langle e_k|$. Obviously, every diagonal operator is normal. Less obvious is that, conversely, the diagonal operators are dense in the set of normal operators. In fact, for every normal operator $N \in \mathcal{B}(\mathcal{H})$ there is a diagonal operator A with $\mathrm{spec}(A) = \mathrm{spec}(N)$ and a sequence of unitaries U_n such that $\lim_{n\to\infty} \|U_n A U_n^* - N\|_\infty = 0$ (cf. the Berg-Weyl-von Neumann theorem discussed in [4]).

Positivity is a crucial concept for many things that follow. It induces a partial order within the set of Hermitian operators by understanding $A \geq B$ as $A - B \geq 0$. There are various ways of characterizing a positive operator. For instance, $A \geq 0$ holds iff $A = A^* \wedge \mathrm{spec}(A) \subseteq [0, \infty)$, which in turn is equivalent to the existence of a $B \in \mathcal{B}(\mathcal{H})$ so that $A = B^*B$. If such a B exists, it can always be chosen positive itself, which then uniquely defines a positive square root $B =: \sqrt{A} \geq 0$ for any $A \geq 0$. This in turn enables the definition of a positive absolute value $|A| := \sqrt{A^*A} \in \mathcal{B}(\mathcal{H})$ for any $A \in \mathcal{B}(\mathcal{H})$. The absolute value is also related to the original operator via the *polar decomposition*, which states that for any $A \in \mathcal{B}(\mathcal{H})$ there is a partial isometry U such that $A = U|A|$. Here U can be taken unitary iff $\ker(A)$ and $\ker(A^*)$ have the same dimension.

Using spectral theory, one can show that every Hermitian operator $A \in \mathcal{B}(\mathcal{H})$ admits a unique decomposition of the form

$$A = A_+ - A_- \quad \text{where} \quad A_\pm \geq 0 \quad \text{and} \quad A_+ A_- = 0. \tag{1.6}$$

In this case, the absolute value can also be expressed as $|A| = A_+ + A_-$ and $A_\pm = \frac{1}{2}(|A| \pm A)$. Another way in which linear combinations of positive operators can be used, is once again a variant of the polarization formula, which for the case of a pair of bounded operators $A, B \in \mathcal{B}(\mathcal{H})$ takes on the form

$$B^*A = \frac{1}{4} \sum_{k=0}^{3} i^k \big(A + i^k B\big)^* \big(A + i^k B\big). \tag{1.7}$$

Ideals of operators Various interesting subspaces of operators in $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ can be obtained as completions of the space of *finite-rank operators* $\mathcal{B}_0(\mathcal{H}_1, \mathcal{H}_2) := \mathrm{lin}\{|\psi\rangle\langle\varphi| \;\; |\psi \in \mathcal{H}_2, \varphi \in \mathcal{H}_1\}$. For instance, the closure of $\mathcal{B}_0(\mathcal{H}_1, \mathcal{H}_2)$ in $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ w.r.t. the operator norm yields the space of *compact operators* $\mathcal{B}_\infty(\mathcal{H}_1, \mathcal{H}_2)$. Every $A \in \mathcal{B}_\infty(\mathcal{H}_1, \mathcal{H}_2)$ admits a *Schmidt decomposition*. That is, it can be written as

$$A = \sum_k s_k |e_k\rangle\langle f_k|, \tag{1.8}$$

where $s \in \mathbb{R}_+^{\mathbb{N}}$ is a null sequence whose non-zero elements are called *singular values* of A and $\{e_k\}, \{f_k\}$ are two orthonormal sets of vectors in $\mathcal{H}_2$ and $\mathcal{H}_1$, respectively. The singular values of A are unique as a multiset. If $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$ each e_k can be chosen proportional (equal) to f_k iff A is normal (positive). In these cases, Eq.(1.8) then leads to the *spectral decomposition*, with eigenvectors e_k and eigenvalues $s_k \langle f_k, e_k\rangle$.

If we restrict the space of compact operators to those for which $s \in l_2(\mathbb{N})$ or $s \in l_1(\mathbb{N})$, we are led to the spaces of *Hilbert-Schmidt class* operators $\mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ and, in the case of equal spaces, the *trace-class* operators $\mathcal{B}_1(\mathcal{H})$, respectively. These become Banach spaces when equipped with the *Hilbert-Schmidt norm* $\|A\|_2 := \|s\|_2$ and the *trace-norm* $\|A\|_1 := \|s\|_1$, respectively. With respect to these norms $\mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathcal{B}_1(\mathcal{H})$ can be regarded as the completion of the space of finite-rank operators and we have the inclusion (with equalities iff $\dim(\mathcal{H}) < \infty$)

$$\mathcal{B}_0(\mathcal{H}) \subseteq \mathcal{B}_1(\mathcal{H}) \subseteq \mathcal{B}_2(\mathcal{H}) \subseteq \mathcal{B}_\infty(\mathcal{H}) \subseteq \mathcal{B}(\mathcal{H}). \tag{1.9}$$

These inclusions also reflect the norm inequalities $\|A\|_1 \geq \|A\|_2 \geq \|A\|_\infty := \|A\|$ for $A \in \mathcal{B}(\mathcal{H})$. All the spaces in Eq.(1.9) are $*$-ideals in $\mathcal{B}(\mathcal{H})$, which means that they are closed under multiplying with elements of $\mathcal{B}(\mathcal{H})$ and under taking the adjoint. Moreover, $A, B \in \mathcal{B}_2(\mathcal{H})$ implies $AB \in \mathcal{B}_1(\mathcal{H})$.

An alternative and equivalent definition of $\mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathcal{B}_1(\mathcal{H})$ is in terms of the *trace*. For a positive operator $A \in \mathcal{B}(\mathcal{H})$, the *trace* $\operatorname{tr}[A] \in [0, \infty]$ is defined as

$$\operatorname{tr}[A] := \sum_k \langle e_k, A e_k \rangle, \tag{1.10}$$

where the sum runs over all elements of an orthonormal basis. Positivity guarantees that the expression is independent of the choice of that basis. Then $\mathcal{B}_1(\mathcal{H})$ is the space of all operators for which $\operatorname{tr}[|A|] < \infty$. For all trace-class operators the trace is then unambiguously defined as well (thus the name) and $\|A\|_1 = \operatorname{tr}[|A|]$. This satisfies $|\operatorname{tr}[A]| \leq \|A\|_1$ (as can be seen from the Schmidt decomposition) and the *Hölder inequality* $\|AB\|_1 \leq \|A\|_1 \|B\|_\infty$ holds.

In a similar vein, we can express the Hilbert-Schmidt norm as $\|B\|_2 = \operatorname{tr}[B^* B]^{\frac{1}{2}}$ for any $B \in \mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$. In fact, $\mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ becomes a Hilbert space when equipped with the *Hilbert-Schmidt inner product* $\langle A, B \rangle := \operatorname{tr}[A^* B]$ (like in example 1.4).

Example 1.7 (Operator bases)**.** As a Hilbert space $\mathcal{B}_2(\mathcal{H})$ admits an orthonormal basis. A simple common choice is the set of *matrix units* $\{|k\rangle\langle l|\}$, which exploits an orthonormal basis $\{|k\rangle\}$ of $\mathcal{H}$. If $d := \dim(\mathcal{H}) < \infty$, another useful basis can be constructed from a *discrete Weyl system*: define a set $\{U_{k,l}\}_{k,l=0}^{d-1}$ of d^2 unitaries by

$$U_{k,l} := \sum_{r=0}^{d-1} \eta^{rl} |k + l\rangle\langle r|, \quad \eta := e^{\frac{2\pi i}{d}}, \tag{1.11}$$

where addition inside the ket is modulo d and $\{|k\rangle\}_{k=0}^{d-1}$ is again an orthonormal basis of $\mathcal{H}$. Then the $U_{k,l}$'s become orthonormal w.r.t. the Hilbert-Schmidt inner product when divided by $\sqrt{d}$. Note that for $d = 2$, the $U_{k,l}$'s reduce to the Pauli matrices (up to phases, i.e. scalar multiplies of modulus 1).

Since $\mathcal{B}_2(\mathcal{H})$ is a Hilbert space, the Riesz representation theorem guarantees that every continuous linear functional on $\mathcal{B}_2(\mathcal{H})$ is of the form

$$A \mapsto \operatorname{tr}[BA], \tag{1.12}$$

for some $B \in \mathcal{B}_2(\mathcal{H})$. That is, $\mathcal{B}_2(\mathcal{H})' \simeq \mathcal{B}_2(\mathcal{H})$. Via the same trace formula we also have that $\mathcal{B}_\infty(\mathcal{H})' \simeq \mathcal{B}_1(\mathcal{H})$ and $\mathcal{B}_1(\mathcal{H})' \simeq \mathcal{B}(\mathcal{H})$. $\mathcal{B}(\mathcal{H})'$, however, contains more elements than those that can be obtained from Eq.(1.12) with $B \in \mathcal{B}_1(\mathcal{H})$.

A frequently used property of the trace is that

$$\operatorname{tr}[AB] = \operatorname{tr}[BA], \tag{1.13}$$

if one of the operators is trace-class or both are Hilbert-Schmidt class. Similarly, $\operatorname{tr}[A|\psi\rangle\langle\varphi|] = \langle\varphi, A\psi\rangle$.

Convergence of operators Let us now have a look at different notions of convergence in $\mathcal{B}(\mathcal{H})$. *Norm convergence* (a.k.a. *uniform convergence*) of the form $\|A_n - A\| \to 0$ for $n \to \infty$ w.r.t. the operator norm is often too strong. The sum in Eq.(1.3), for instance, does clearly not converge in norm: if we denote the n'th partial sum by A_n, then $\|A_n - A_{n-1}\| = \||e_n\rangle\langle e_n|\| = 1$ in this case. Weaker notions of convergence are:

- *Weak convergence*, which requires $\langle\psi, (A_n - A)\varphi\rangle \to 0$ for all $\varphi, \psi \in \mathcal{H}$,
- *Weak-* convergence*[5], which requires $\operatorname{tr}[(A_n - A)B] \to 0$ $\forall B \in \mathcal{B}_1(\mathcal{H})$,
- *Strong convergence*, which requires $\|(A_n - A)\psi\| \to 0$ for all $\psi \in \mathcal{H}$.

These are generally related as follows: norm convergence implies weak-* convergence (via Hölder's inequality) and also strong convergence (via Lipschitz inequality). These two, in turn, imply weak convergence (by using $B = |\varphi\rangle\langle\psi|$ and Cauchy-Schwarz, respectively). Moreover, on norm-bounded subsets of $\mathcal{B}(\mathcal{H})$ weak and weak-* convergence are equivalent (as shown by employing Schmidt decomposition together with dominated convergence).

The expression in Eq.(1.3) is strongly convergent. More generally, any norm-bounded increasing sequence of Hermitian operators is strongly convergent in $\mathcal{B}(\mathcal{H})$. This is often useful to lift results from finite dimensions to infinite dimensions. Sometimes it is used together with the fact that if $A_n \to A$ and $B_n \to B$ each converge strongly, then $A_nB_n \to AB$ converges strongly as well, and $A_nC \to AC$ converges in norm for any $C \in \mathcal{B}_\infty(\mathcal{H})$.

Each of the mentioned notions of convergence is based on a corresponding topology on $\mathcal{B}(\mathcal{H})$. These are defined as the coarsest (i.e., smallest) topologies for which the linear functionals that appear in the respective definitions of convergence are continuous. The *weak-* topology*, for instance, can be defined as the smallest topology in which all functionals of the form $\mathcal{B}(\mathcal{H}) \ni A \to \operatorname{tr}[AB]$ are continuous for any $B \in \mathcal{B}_1(\mathcal{H})$.

The *Banach-Alaoglu theorem* implies that $\mathcal{B}(\mathcal{H})$ has the *Heine-Borel property* w.r.t to the weak (and weak-*) topology. That is, each closed, norm-bounded subset (such as the unit ball) is compact. Moreover, w.r.t. these two topologies, norm-bounded subsets of $\mathcal{B}(\mathcal{H})$ are metrizable, which implies that compactness and sequential compactness are equivalent.

[5] a.k.a. *ultraweak convergence* or *σ-weak convergence*.

Functional calculus If A is an operator on $\mathcal{H}$ and $f : \mathbb{C} \to \mathbb{C}$ a function, there are different ways of defining $f(A)$ depending on the properties of f and A. We will briefly survey two of them that both generalize the straightforward case of polynomial functions and both involve the spectrum of A.

Recall that the spectrum $\mathrm{spec}(A) \subseteq \mathbb{C}$ of a bounded operator is the set of complex numbers λ for which the operator $(\lambda \mathbb{1} - A)$ is not invertible (i.e. it represents a map that is not bijective). If f is holomorphic on a simply connected domain $D \supset \mathrm{spec}(A)$ and Γ a rectifiable closed curve in D that does not intersect itself and surrounds $\mathrm{spec}(A)$, then Cauchy's integral formula can be used to define

$$f(A) := \frac{1}{2\pi i} \oint_\gamma f(z)\big(z\mathbb{1} - A\big)^{-1} dz. \tag{1.14}$$

This way of defining $f(A)$ is called *holomorphic functional calculus*. The integral in Eq.(1.14) converges in operator norm and the resulting operator satisfies $\mathrm{spec}\big(f(A)\big) = f\big(\mathrm{spec}(A)\big)$. Moreover, if $g : D \to \mathbb{C}$ is another holomorphic function an gf denotes the pointwise product, then $g(A)f(A) = gf(A)$.

If f is merely continuous on a set that contains $\mathrm{spec}(A)$, then one can still define $f(A)$ if A is a normal operator. The idea is to exploit the spectral decomposition and to let f act directly on the spectrum of A. In particular, if $A \in \mathcal{B}_1(\mathcal{H})$ has spectral decomposition $A = \sum_k \lambda_k |\psi_k\rangle\langle\psi_k|$, then

$$f(A) := \sum_k f(\lambda_k) |\psi_k\rangle\langle\psi_k|, \tag{1.15}$$

where the sum converges in trace-norm. This is called *continuous functional calculus*. If f is analytic, it coincides with the holomorphic functional calculus. That is, if the assumptions of both functional calculi are satisfied, then Eq.(1.14) equals Eq.(1.15).

Unbounded operators and spectral measures In this paragraph we will have a brief look at how to generalize what we know about Hermitian bounded operators to their unbounded 'self-adjoint' relatives. An unbounded operator A can usually not be defined on the entire Hilbert space $\mathcal{H}$ so that it is necessary to introduce its *domain* $\mathcal{D}(A) \subseteq \mathcal{H}$. The adjoint A^* of an operator $A : \mathcal{D}(\mathcal{H}) \to \mathcal{H}$ also has to be defined with more care. For that it will be necessary that $\mathcal{D}(\mathcal{H})$ is a dense subspace of $\mathcal{H}$. The adjoint can then be uniquely defined on $\mathcal{D}(A^*) := \{\varphi \in \mathcal{H} | \psi \mapsto \langle\varphi, A\psi\rangle \text{ is continuous on } \mathcal{D}(\mathcal{H})\}$ so that

$$\langle\varphi, A\psi\rangle = \langle A^*\varphi, \psi\rangle \quad \forall \psi \in \mathcal{D}(A), \varphi \in \mathcal{D}(A^*). \tag{1.16}$$

This definition directly exploits the Riesz-representation theorem, which only gives rise to uniqueness of A^* if $\mathcal{D}(A)$ is dense. $\mathcal{D}(A^*)$, however, is not automatically dense – it may even happen that $\mathcal{D}(A^*) = \{0\}$.

A densely defined operator A is called *self-adjoint* if $A = A^*$ and $\mathcal{D}(A) = \mathcal{D}(A^*)$. So bounded Hermitian operators are special cases of self-adjoint operators. By the Hellinger-Toeplitz theorem a self-adjoint operator is bounded

iff it can be defined on all of $\mathcal{H}$. This underlines that considering domains is unavoidable for unbounded operators.

If A is self-adjoint, then $(A^*)^* = A$ and the ranges of $A \pm i\mathbb{1}$ are the entire Hilbert space. The latter is related to the fact that the *Calyey transform* $(A - i\mathbb{1})(A+i\mathbb{1})^{-1} =: U$ of a self-adjoint operator A defines a unitary. Exploiting this relation, von Neumann was able to use the spectral theorem for unitaries, which are necessarily bounded, to prove a spectral theorem for self-adjoint operators.

One formulation of the spectral theorem is in terms of projection-valued measures (PVMs). For any self-adjoint operator A there is a PVM $P : \mathbb{B} \to \mathcal{B}(\mathcal{H})$, where $\mathbb{B}$ is the Borel σ-algebra on $\mathbb{R}$, so that

$$A = \int_{\mathbb{R}} \lambda \, dP(\lambda). \tag{1.17}$$

The integral is understood in the following weak sense: for any $\psi \in \mathcal{D}(A), \varphi \in \mathcal{H}$ we can define a Borel-measure $\mu : \mathbb{B} \to \mathbb{C}$ via $\mu(Y) := \langle \varphi, P(Y)\psi \rangle$ that satisfies $\langle \varphi, A\psi \rangle = \int_{\mathbb{R}} \lambda \, d\mu(\lambda)$. The PVM P that is associated to A is called its *spectral measure* and one can show that there is a one-to-one correspondence between self-adjoint operators and PVMs on $(\mathbb{R}, \mathbb{B})$. Not surprisingly, $\lambda \in \mathbb{R}$ is an eigenvalue of A iff $P(\{\lambda\}) \neq 0$. In this case $P(\{\lambda\})$ is the corresponding spectral projection.

As in the compact case, the spectral representation in Eq.(1.17) leads directly to a functional calculus. For any measurable function $f : \mathbb{R} \to \mathbb{C}$ we can define

$$\begin{aligned} f(A) \quad &= \quad \int_{\mathbb{R}} f(\lambda) \, dP(\lambda) \\ &\text{on} \quad \mathcal{D}\big(f(A)\big) := \left\{ \varphi \in \mathcal{H} \middle| \int_{\mathbb{R}} |f(\lambda)|^2 \, d\langle \varphi, P(\lambda)\varphi \rangle < \infty \right\}. \end{aligned} \tag{1.18}$$

If f is bounded, then Eq.(1.18) gives rise to a bounded operator.

Exercise 1.6. Let $A, B \in \mathcal{B}(\mathcal{H})$ be Hermitian. Show that

a) $\mathrm{tr}\,[AB] \in \mathbb{R}$ if $B \in \mathcal{B}_1(\mathcal{H})$,

b) $A \geq B \wedge A \leq B$ implies $A = B$,

c) $A \geq B$ implies that $CAC^* \geq CBC^*$ for all $C \in \mathcal{B}(\mathcal{H}, \tilde{\mathcal{H}})$.

Exercise 1.7. Let $A, B \in \mathcal{B}(\mathcal{H})$ be positive and $B \in \mathcal{B}_1(\mathcal{H})$. Show that

a) $\mathrm{tr}\,[AB] \geq 0$,

b) $\mathrm{tr}\,[AB] = 0$ implies $AB = BA = 0$.

Exercise 1.8. Let $P \in \mathcal{B}(\mathcal{H})$ be an orthogonal projection. Show that

a) $0 \leq P \leq \mathbb{1}$,

b) if $0 \leq A \leq \mu P$ for some $\mu \in \mathbb{R}_+$ and Hermitian $A \in \mathcal{B}(\mathcal{H})$, then $A = AP = PAP$.

Exercise 1.9. Let $A_1, \ldots, A_n \in \mathcal{B}_1(\mathcal{H})$ be a set of Hermitian operators and $\mathcal{A} := \{B \in \mathcal{B}_1(\mathcal{H}) | \, \forall i : B \geq A_i\}$.

a) Show that $\mathcal{A}$ contains a unique element of minimal trace (although $\mathcal{A}$ does in general not contain a unique least element w.r.t. the operator ordering).

b) Show that for $n = 2$ the element of minimal trace in $\mathcal{A}$ is given by $\frac{1}{2}(A_1 + A_2 + |A_1 - A_2|) = A_1 + (A_1 - A_2)_-$.

Exercise 1.10. For the operator norm on $\mathcal{B}(\mathcal{H})$, show that

a) $0 \leq A \leq B$ implies that $\|A\| \leq \|B\|$,

b) $-\mathbb{1} \leq C \leq \mathbb{1}$ iff $\|C\| \leq 1$ for Hermitian C,

c) $\|AB\| \leq \|A\| \, \|B\|$,

d) $\|A^*A\| = \|A\|^2$ for all $A \in \mathcal{B}(\mathcal{H})$,

e)* $\|A\| = \sup_{\|\psi\|=1} |\langle \psi, A\psi \rangle|$ for all normal A.

Exercise 1.11. Let $Q \in \mathcal{B}(\mathcal{H})$ be positive and such that $\ker(Q) = \{0\}$. Prove that $(A, B) \mapsto \operatorname{tr}[QA^*B]$ defines an inner product on $\mathcal{B}_2(\mathcal{H})$.

Exercise 1.12. Construct a sequence of finite rank operators $A_n \in \mathcal{B}_0(\mathcal{H})$ that converges weakly to zero but not strongly.

Notes and literature A good introduction to Hilbert space operators can for instance be found in Pedersen's *Analysis Now* [5]. More comprehensive expositions are given by Conway [6], Simon [7] and Dunford and Schwartz [8].

1.3 Probabilistic structure of Quantum Theory

Quantum theory can be regarded as a general theoretical framework for physical theories. It consists out of a mathematical core that becomes a physical theory when adding a set of correspondence rules telling us which mathematical objects we have to use in different physical situations.

Quantum theory divides the description of any physical experiments into two parts: *preparation* and *measurement*. This innocent-looking step already covers one of the basic differences between the quantum and the classical world, as in classical physics there is no need to talk about measurements in the first place. Note also that the division of a physical process into preparation and measurement is sometimes ambiguous, but, fortunately, quantum theoretical predictions do not depend on the particular choice of the division.

A genuine request is that a physical theory should predict the outcome of any measurement given all the information about the preparation, i.e., the initial conditions, of the system. Quantum mechanics[6] teaches us that this is in general not possible and that all we can do is to predict the probabilities of outcomes in statistical experiments, i.e., long series of experiments where all relevant parameters in the procedure are kept unchanged. Thus, quantum mechanics does not predict individual events, unless the corresponding probability distribution happens to be tight. We will see later that there are good reasons to believe that this 'fuzziness' is not due to incompleteness of the theory and lacking knowledge about some *hidden variables* but rather part of nature's character. In fact, *entanglement* will be the leading actor in that story. The fact that the appearance of probabilities is not only due to the ignorance of the observer but at the very heart of the description, means that the measurement process can be regarded as a transition from possibilities to facts.

The *preparation* of a quantum system is the set of actions that determines all probability distributions of any possible measurement. It has to be a procedure that, when applied to a statistical ensemble, leads to converging relative frequencies and thus allows us to talk about probabilities. Since many different preparations can have the same effect in the sense that all the resulting probability distributions coincide it is reasonable to introduce the concept of a *state*, which specifies the effect of a preparation regardless of how it has actually been performed. Note that, in contrast to classical mechanics, a quantum 'state' does not refer to the attributes of an individual system but rather describes a statistical ensemble—the effect of a preparation in a statistical experiment. One should thus be careful with assigning states to individual systems. Talking about the 'state of an individual atom' is more common but not necessarily more meaningful than talking about the 'Bernoulli distribution of an individual coin'.

[6] We use *quantum mechanics* and *quantum theory* synonymously.

Preparation While the term 'state' is used for various different albeit related mathematical objects (explained further down), a mathematically unambiguous way to describe the preparation of a quantum system is the use of *density operators*:

Definition 1.5 (Density operators)**.** *$\rho \in \mathcal{B}_1(\mathcal{H})$ is called a* density operator *if it is positive and satisfies $\operatorname{tr}[\rho] = 1$. A density operator is called* pure *if there is a unit vector $\psi \in \mathcal{H}$ such that $\rho = |\psi\rangle\langle\psi|$, and it is called* mixed *otherwise.*

A pure density operator is completely specified by the corresponding unit vector ψ, which in turn is specified by the density operator up to a scalar of modulus one (a 'phase'). The term 'state' is used for both ρ and ψ. To emphasize the latter case, 'state vector' is sometimes used.[7]

On the level of state vectors, a natural mathematical operation is linear combination: for any pair of unit vectors ψ_1, ψ_2 new state vectors can be obtained as $\psi = c_1\psi_1 + c_2\psi_2$ with appropriately chosen $c_1, c_2 \in \mathbb{C}$. ψ is then said to be a *superposition* of ψ_1 and ψ_2.

On the level of density operators, a superficially similar natural mathematical operation is convex combination. As we will see below, this has, however, an entirely different physical interpretation and it will usually change the *purity* of the state.

Proposition 1.6 (Purity)**.** *Let $\rho \in \mathcal{B}(\mathcal{H})$ be a density operator. Then $0 < \operatorname{tr}\left[\rho^2\right] \leq 1$ with equality iff ρ describes a pure state. Moreover, if $d := \dim(\mathcal{H}) < \infty$, then $\operatorname{tr}\left[\rho^2\right] \geq 1/d$ with equality iff $\rho = \mathbb{1}/d$ (which is then called* maximally mixed*).*

Proof. Since $\operatorname{tr}\left[\rho^2\right] = \|\rho\|_2^2$, it is positive and non-zero. Hölder's inequality together with $\|\rho\|_1 = 1$ gives $\operatorname{tr}\left[\rho^2\right] \leq \|\rho\|$. Since the operator norm, in this case, equals the largest eigenvalue and all eigenvalues are positive and sum up to one, we get $\|\rho\| \leq 1$ with equality iff ρ has rank one.

For the lower bound in finite dimensions, we can invoke the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product in order to get:

$$1 = \operatorname{tr}[\mathbb{1}\rho]^2 \leq \operatorname{tr}[\mathbb{1}] \operatorname{tr}\left[\rho^2\right] = d \operatorname{tr}\left[\rho^2\right].$$

Equality in the Cauchy-Schwarz inequality holds iff ρ is a multiple of $\mathbb{1}$, and $\operatorname{tr}[\rho] = 1$ determines the prefactor. □

Example 1.8 (Bloch ball)**.** There is a bijection between the set of density operator on $\mathbb{C}^2$ and the set of vectors $r \in \mathbb{R}^3$ with Euclidean norm $\|r\| \leq 1$, given

[7] There is yet another, more general, mathematical meaning of the term 'state', namely as a positive normalized linear functional. Clearly, every density operator induces such a functional via $A \mapsto \operatorname{tr}[\rho A]$. In fact, every weak-* continuous positive normalized linear functional on $\mathcal{B}(\mathcal{H})$ is of this form. If one drops or relaxes the continuity requirement, there are, however, other 'states' as well. Those arising from density operators are then called *normal states* and the other ones *singular states*.

by

$$\rho = \frac{1}{2}\Big(\mathbb{1} + \sum_{i=1}^{3} r_i \sigma_i\Big). \tag{1.19}$$

The purity is then expressible as $\operatorname{tr}\left[\rho^2\right] = \frac{1}{2}\big(1 + \|r\|^2\big)$. Consequently, the boundary coincides with the set of pure states and the origin corresponds to the maximally mixed state. Physically, a two-level density operator (a '*qubit*') might for instance model:

- An atom in a double-well potential. $\rho = |0\rangle\langle 0|$ and $\rho = |1\rangle\langle 1|$ would then correspond to the atom being left or right, respectively.
- A two-level atom with $\rho = |0\rangle\langle 0|$, $\rho = |1\rangle\langle 1|$ referring to the ground and exited state, respectively.
- The spin of an electron with $\rho = |0\rangle\langle 0| \hat{=}$ spin up, $\rho = |1\rangle\langle 1| \hat{=}$ spin down.
- Polarization degrees of freedom of light. North-/south pole correspond to left-/right circular polarization while the east-/west pole correspond to horizontal/vertical polarization. The center $\rho = \frac{\mathbb{1}}{2}$ then describes unpolarized light.

The case $\dim(\mathcal{H}) = 2$ is very special in many ways. For instance, a nice geometric representation of the set of all density operators as in Eq.(1.19) is not possible in higher dimensions.

In infinite dimensions, as seen in Exercise 1.12, weak convergence can be a rather weak property, indeed, even when restricted to finite-rank operators. On the set of density operators, however, normalization and positivity assure that weak convergence implies every other from of convergence:

Theorem 1.7 (Convergence to a density operator)**.** *Let $\rho_n \in \mathcal{B}_1(\mathcal{H})$ be a sequence of positive operators that converges weakly to a density operator ρ and satisfies $\operatorname{tr}[\rho_n] \to 1$. Then $\|\rho_n - \rho\|_1 \to 0$.*

Proof. Exploiting the spectral decomposition of ρ, we can find a finite-dimensional orthogonal projection P for which $1 - \operatorname{tr}[\rho P] =: \epsilon$ is arbitrarily small. That is, for any $\varepsilon > 0$, we can achieve $\epsilon < \varepsilon$ in this way. With $P^\perp := \mathbb{1} - P$ we can bound

$$\|\rho - \rho_n\|_1 \leq \|P(\rho - \rho_n)P\|_1 + 2\left\|P(\rho - \rho_n)P^\perp\right\|_1 + \left\|P^\perp(\rho - \rho_n)P^\perp\right\|_1. \tag{1.20}$$

The first term on the r.h.s. converges to zero, since it involves only finite-dimensional operators on which weak convergence implies norm convergence (in any norm). For the second term on the r.h.s. of Eq.(1.20) we first use that $P\rho P^\perp = 0$ and then bound the remaining part via

$$\begin{aligned}\left\|P\rho_n P^\perp\right\|_1 \leq \left\|\rho_n P^\perp\right\|_1 &= \operatorname{tr}\left[U^*\sqrt{\rho_n}\sqrt{\rho_n}P^\perp\right] \leq \sqrt{\operatorname{tr}[\rho_n]\operatorname{tr}[\rho_n P^\perp]} \\ &= \sqrt{\operatorname{tr}[\rho_n]\left(\operatorname{tr}[\rho_n] - \operatorname{tr}[P\rho_n P]\right)} \to \sqrt{\epsilon}.\end{aligned}$$

Here, we have first used Hölder's inequality, then the polar decomposition $\rho_n P^\perp = U|\rho_n P^\perp|$, and in the third step the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product.

Finally, an upper bound for the third term on the r.h.s. of Eq.(1.20) is

$$\begin{aligned} \left\|P^\perp(\rho-\rho_n)P^\perp\right\|_1 &\leq \operatorname{tr}\left[P^\perp\rho P^\perp\right] + \operatorname{tr}\left[P^\perp\rho_n P^\perp\right] \\ &= \epsilon + \operatorname{tr}\left[\rho_n\right] - \operatorname{tr}\left[P\rho_n P\right] \to 2\epsilon. \end{aligned}$$

□

In fact, the property just proven extends to the entire space of trace-class operators: if $T_n \in \mathcal{B}_1(\mathcal{H})$ converges weakly to $T \in \mathcal{B}_1(\mathcal{H})$ and $\|T_n\|_1 \to \|T\|_1$, then $T_n \to T$ in trace-norm [9].

Measurements Let X be the set of all possible measurement outcomes in a given description of an experiment. We will denote by $\mathbb{B}$ a σ-algebra over X. If X is discrete, then $\mathbb{B}$ is usually just the power set and if X is a manifold (in particular, if $X = \mathbb{R}$), then the canonical choice for $\mathbb{B}$ is the corresponding Borel σ-algebra. For the moment, we will treat the elements of X just as labels without further physical meaning. The mathematical object assigned to each measurement apparatus is then a *positive operator valued measure* (POVM):

Definition 1.8 (POVMs)**.** *A* positive operator valued measure *(POVM) on a measurable space $(X, \mathbb{B})$ is a map $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ that satisfies $M(Y) \geq 0$ for all $Y \in \mathbb{B}$ and*

$$\sum_k M(X_k) = \mathbb{1} \tag{1.21}$$

for every countable, disjoint partition $X = \cup_k X_k$ with $X_k \in \mathbb{B}$. A POVM is called sharp *if $M(Y)$ is an* orthogonal projection *for any $Y \in \mathbb{B}$. In this case, M is also called a* projection valued measure *(PVM).*

Due to Eq.(1.21), M is also called *resolution of identity* in the literature. If X is discrete, M is determined by the tuple of operators $M_x := M(\{x\})$ that correspond to the singletons $x \in X$. Then $M(Y) = \sum_{x\in Y} M_x$ for any $Y \subseteq X$ and with a slight abuse of terminology, one often calls the tuple $(M_x)_{x\in X}$ of positive operators that sum up $\mathbb{1}$ the POVM.

Positivity of the $M(Y)$ together with the normalization requirement in Eq.(1.21) implies $0 \leq M(Y) \leq \mathbb{1}$.[8] Moreover:

Lemma 1.9. *Let $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ be a POVM and $J, Y \in \mathbb{B}$.*

(1) If $J \subseteq Y$, then $M(J) + M(Y\backslash J) = M(Y)$ and $M(J) \leq M(Y)$,

(2) $M(J \cup Y) \leq M(J) + M(Y)$ with equality if $Y \cap J = \emptyset$.

[8] An element $E \in \mathcal{B}(\mathcal{H})$ that satisfies $0 \leq E \leq \mathbb{1}$ is in this context often called *effect operator*.

Proof. Using Eq.(1.21) twice, we get

$$\mathbb{1} = \begin{cases} M(Y) + M(X\backslash Y) \\ M(J) + M(Y\backslash J) + M(X\backslash Y). \end{cases}$$

By subtraction of the two lines we obtain $M(Y) - M(J) = M(Y\backslash J) \geq 0$, which proves (1). In order to arrive at (2), we exploit (1) for the sets J and $J \cup Y$. Then $M(J \cup Y) = M(J) + M((J \cup Y)\backslash J) \leq M(J) + M(Y)$. □

If a POVM M is projection valued, then (cf. Exercise 1.17) $M(Y)M(J) = 0$ whenever $Y \cap J = \emptyset$. This in turn implies that on a d-dimensional Hilbert space any PVM can have at most d distinct possible measurement outcomes.

Example 1.9. For $m \in \mathbb{N}$, every m-outcome PVM $\big(M_x \in \mathcal{B}(\mathcal{H})\big)_{x=1}^{m} =: M$ can be obtained from a unitary $U \in \mathcal{B}(\mathcal{H})$ of order m (i.e., U satisfies $U^m = \mathbb{1}$) and vice versa: given M, we can define $U := \sum_{x=1}^{m} \exp(\frac{2\pi i x}{m}) M_x$ and conversely, given U we can obtain (or recover) a PVM via $M_x := \frac{1}{m} \sum_{k=1}^{m} \exp(\frac{2\pi i k}{m}) U^k$.

We will later see (in *Naimark's theorem* Thm.1.65) that every POVM can be obtained from a PVM that acts on a larger Hilbert space.

Probabilities Having introduced the basic mathematical objects that are assigned to preparation and measurement, it remains to see how these are combined in a way that eventually leads to probabilities. This is what the following postulate is doing:

Postulate 1.10 (Born's rule)**.** *The probability $p(Y|\rho, M)$ of measuring an outcome in $Y \in \mathbb{B}$ if preparation and measurement are described by a density operator $\rho \in \mathcal{B}_1(\mathcal{H})$ and a POVM $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$, respectively, is given by*

$$p(Y|\rho, M) = \operatorname{tr}\left[\rho M(Y)\right]. \tag{1.22}$$

If ρ and M are clear from the context, we will simply write $p(Y) := p(Y|\rho, M)$ and if X is discrete and $\mathbb{B}$ the corresponding power set, we will write $p(x)$ or p_x for $p(\{x\})$.

The defining properties of density operators and POVMs now nicely play together so that $p(Y|\rho, M)$ has all the necessary properties for an interpretation in terms of probabilities:

Corollary 1.11. *For any density operator $\rho \in \mathcal{B}_1(\mathcal{H})$ and POVM $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$, the map $p : Y \mapsto p(Y)$ that appears in Born's rule defines a probability measure on $(X, \mathbb{B})$.*

Proof. First observe that $\forall Y \in \mathbb{B} : 0 \leq p(Y) \leq 1$. The lower bound follows from positivity of ρ and $M(Y)$ (cf. Exercise 1.7a) and the upper bound from Eq.(1.21) applied to the trivial partition of X together with $\operatorname{tr}[\rho] = 1$. When applying Eq.(1.21) to $X = X \cup \emptyset$ together with positivity of M, we obtain further that $p(X) = 1$ and $p(\emptyset) = 0$.

Finally, we have to show that $\sum_k p(X_k) = 1$ for any countable disjoint partition $X = \cup_k X_k$ with $X_k \in \mathbb{B}$. This again follows from Eq.(1.21) since

$$\sum_k p(X_k) = \sum_k \mathrm{tr}\left[\rho M(X_k)\right] = \mathrm{tr}\left[\rho \sum_k M(X_k)\right] = \mathrm{tr}\left[\rho \mathbb{1}\right] = 1. \tag{1.23}$$

Here interchanging the sum with the one in the trace is justified by positivity of all expressions and Fubini-Tonelli. □

If M and ρ are given, Born's rule tells us how to compute quantum theory's prediction of the measurement probabilities. In practice, we typically know M and ρ only for some simple cases together with some mathematical rules (yet to be formalized in this lecture) telling us how to reduce more general cases to these simple ones. The largest part of quantum theory (Schrödinger equation, composite systems, etc.) is about those rules and their consequences.

Traditional textbook quantum theory often assumes ρ to be pure and M to be sharp. We will soon see in which sense this is justified.

As a first application of the formalism, let us consider the problem of information transmission via a d-level quantum system, i.e., one for which $\mathcal{H} = \mathbb{C}^d$. Given an alphabet X of size $|X| = m$, is it possible to encode all its elements into a d-level quantum system so that the information can finally be retrieved exactly or at least with a small probability of error?

Following the rules of the formalism, we assign a density operator $\rho_x \in \mathcal{B}(\mathcal{H})$ to each $x \in X$. Similarly, we assume that there is a measurement apparatus that has X as the set of possible measurement outcomes so that a positive operator $M_x \in \mathcal{B}(\mathcal{H})$ is assigned to each outcome and that $\sum_{x\in X} M_x = \mathbb{1}$. If ρ_x has been prepared, the probability for measuring the correct outcome is then, according to Born's rule: $p_x := \mathrm{tr}\left[\rho_x M_x\right]$. Now consider the average probability of success, averaged uniformly over all $x \in X$:

Proposition 1.12. *The average probability of success, when transmitting an alphabet of size m over a d-level quantum system satisfies $\frac{1}{m}\sum_x p_x \leq \frac{d}{m}$.*

Proof. The claim follows from the defining properties of POVMs and density operators for instance via the use of Hölder's inequality and the fact that $\|\rho_x\|_\infty \leq 1$:

$$\frac{1}{m}\sum_x p_x = \frac{1}{m}\sum_x \mathrm{tr}\left[\rho_x M_x\right] \leq \frac{1}{m}\sum_x \|\rho_x\|_\infty \|M_x\|_1 \leq \sum_x \mathrm{tr}\left[M_x\right] = \frac{d}{m}.$$

□

This should be compared with the performance of the following naive classical (= non-quantum) protocol that aims at transmitting a random element from the alphabet X using only d of its elements: fix any subset $D \subseteq X$ of $d = |D|$ elements; send x if $x \in D$ and send an arbitrary element from D if $x \notin D$. The

probability of success of this protocol is d/m. Prop.1.12 tells us that this can not be outperformed by any quantum protocol.

As a second simple application of the formalism, let us analyze to what extent a change in ρ or M can alter the probability of a measurement outcome:

Corollary 1.13 (Lipschitz-bounds for probabilities)**.** *Let $\rho, \rho' \in \mathcal{B}_1(\mathcal{H})$ be density operators, $M, M' : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ POVMs on a common measurable space $(X, \mathbb{B})$ and $Y \in \mathbb{B}$. Then*

$$\big|p(Y|\rho, M) - p(Y|\rho', M)\big| \leq \frac{1}{2} \left\|\rho - \rho'\right\|_1 , \tag{1.24}$$

where equality can be attained for every pair ρ, ρ' by a suitable choice of the POVM M. Similarly,

$$\sup_{\rho} \big|p(Y|\rho, M) - p(Y|\rho, M')\big| = \left\|M(Y) - M'(Y)\right\|_\infty . \tag{1.25}$$

Proof. Consider the decomposition $(\rho - \rho') = \Delta_+ - \Delta_-$ into orthogonal positive and negative parts (as introduced in Eq.(1.6)) and denote by P_+ the orthogonal projection onto the closure of the range of Δ_+. Then $\Delta_\pm \geq 0$, $P_+\Delta_+ = \Delta_+$ and $P_+\Delta_- = 0$. Moreover, $\mathrm{tr}\,[\rho - \rho'] = 0$ implies $\mathrm{tr}\,[\Delta_+] = \mathrm{tr}\,[\Delta_-]$ and $|\rho - \rho'| = \Delta_+ + \Delta_-$ implies further that $\|\rho - \rho'\|_1 = 2\mathrm{tr}\,[\Delta_+]$. W.l.o.g. we assume that $\mathrm{tr}\,[\Delta_+ M(Y)] \geq \mathrm{tr}\,[\Delta_- M(Y)]$ (otherwise interchange $\rho \leftrightarrow \rho'$). Then using positivity of $M(Y)$ we obtain (by Born's rule, Exercise 1.7a and Hölder's inequality):

$$\begin{aligned} \big|p(Y|\rho, M) - p(Y|\rho', M)\big| &= \big|\mathrm{tr}\,[\Delta_+ M(Y)] - \mathrm{tr}\,[\Delta_- M(Y)]\,\big| \leq \mathrm{tr}\,[\Delta_+ M(Y)] \\ &\leq \|\Delta_+\|_1 \left\|M(Y)\right\|_\infty \leq \frac{1}{2} \left\|\rho - \rho'\right\|_1 , \end{aligned}$$

where we have used $\left\|M(Y)\right\|_\infty \leq 1$, which is a consequence of $0 \leq M(Y) \leq \mathbb{1}$ (cf. Exercise 1.10). Equality in all the involved inequalities is achieved for $M(Y) = P_+$. The operators $(P_+, \mathbb{1} - P_+)$ then form a suitable POVM.

In order to arrive at Eq.(1.25), first note that Hölder's inequality together with $\|\rho\|_1 = 1$ leads to the upper bound

$$\big|\mathrm{tr}\,\big[\rho\big(M(Y) - M'(Y)\big)\big]\,\big| \leq \left\|M(Y) - M'(Y)\right\|_\infty .$$

That this equals the supremum follows from the fact that the operator norm of the Hermitian operator $M(Y) - M'(Y)$ can already be obtained by taking the supremum over all pure states $\rho = |\psi\rangle\langle\psi|$ on the l.h.s. (cf. Exercise 1.10d). □

The fact that Eq.(1.24) is tight provides an operational interpretation for the trace-norm distance between two density operators as a means of quantifying the extent to which the two corresponding preparations can be distinguished in a statistical experiment.

Observables and expectation values So far we have treated the measurement outcome merely as a label without further meaning. In practice, there is often a numerical value assigned to every $x \in X$. We will denote this value by $m(x) \in \mathbb{R}$ and assume that the function m is $\mathbb{B}$-measurable. Two frequently used quantities are the *expectation value* $\langle m \rangle := \int_X m(x) dp(x)$ and the *variance* $\mathrm{var}(m) := \int_X m(x)^2 dp(x) - \langle m \rangle^2$.

If the probability measure p is represented according to Born's rule, we can write the expectation value as

$$\langle m \rangle = \mathrm{tr}\left[\rho \hat{M}\right], \quad \hat{M} := \int_X m(x)\, dM(x), \tag{1.26}$$

which in the discrete case reduces to $\hat{M} = \sum_x m(x) M_x$. We will also use the common notation $\langle \hat{M} \rangle := \mathrm{tr}\big[\rho \hat{M}\big]$. So far, $\hat{M}$ is a formal expression that is not guaranteed to be meaningful if m is not bounded. For simplicity, we will leave the discussion of the unbounded case aside.

If the underlying POVM M is sharp, then $\hat{M} = \sum_x m(x) M_x$ becomes a spectral decomposition. In this case, we call $\hat{M}$ an *observable*[9] and notice that each $m(x)$ is then an eigenvalue of $\hat{M}$ with corresponding spectral projection M_x. That is, $\hat{M}$ determines both m and M. In this way, any Hermitian operator is a mathematically valid observable whose spectral decomposition determines the set of possible measurement values and the POVM. Furthermore, since spectral projections of a Hermitian operator that correspond to different eigenvalues are mutually orthogonal (i.e. $M_x M_y = \delta_{x,y} M_x$, cf. Exercise1.17) we can express the variance as

$$\mathrm{var}(m) = \mathrm{tr}\left[\rho \hat{M}^2\right] - \mathrm{tr}\left[\rho \hat{M}\right]^2 =: \mathrm{var}(\hat{M}). \tag{1.27}$$

Notice that this does not hold in general, i.e. when M is not sharp.

Textbook descriptions of quantities like position, momentum, energy, angular momentum and spin are usually in terms of observables (albeit in the more general framework of not necessarily bounded self-adjoint operators). For instance, the Pauli matrices, when divided by two, are the observables that correspond to the three spin directions of a spin-$\frac{1}{2}$ particle.

Exercise 1.13. Show that every trace-class operator can be written as a linear combination of four density operators.

Exercise 1.14. Let $V \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ be such that for every density operator $\rho \in \mathcal{B}_1(\mathcal{H}_1)$ the operator $V \rho V^*$ is again a density operator. What can be said about V?

Exercise 1.15. Prove the Bloch ball representation in Eq.(1.19). (Hint: use the determinant). For a given density operator on $\mathbb{C}^2$, how can the vector r be obtained?

Exercise 1.16. For any $\mathcal{H}$ construct a POVM that implements a 'biased coin' whose outcomes occur independently of the density operator with probabilities $\frac{1}{2}(1 \pm b)$, where $b \in [0, 1]$ is a fixed bias.

[9] Traditionally, the term *observable* is associated to self-adjoint operators. Sometimes, however, it is also used more generally, often synonymous with *measurement*.

Exercise 1.17. Let $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ be a sharp POVM on $(X, \mathbb{B})$. Show that $Y \cap J = \emptyset$ implies that $M(Y)M(J) = 0$. From here, prove that the number of pairwise disjoint elements in $\mathbb{B}$ on which M is non-zero is at most d if $\mathcal{H} = \mathbb{C}^d$.

Exercise 1.18. Show that two preparations described by density operators $\rho_1, \rho_2 \in \mathcal{B}_1(\mathcal{H})$ can be distinguished with certainty in a statistical experiment iff $\rho_1\rho_2 = 0$.

Exercise 1.19. Construct a pair of density operators ρ, ρ' on a common Hilbert space with the properties that: (i) their spectra coincide and each eigenvalue has multiplicity one, (ii) there is no unitary U such that $U\rho U^* = \rho'$.

Exercise 1.20. Let $\rho_1, \rho_2 \in \mathcal{B}(\mathcal{H})$ be density operators on a finite-dimensional Hilbert space $\mathcal{H}$ and $\lambda \in \mathbb{R}$. Show that $\rho_1 \geq \lambda\rho_2$ holds iff the following are both satisfied:

(i) $\mathrm{range}(\rho_2) \subseteq \mathrm{range}(\rho_2)$,

(ii) $\lambda \leq \|\rho_1^{-1/2}\rho_2\rho_1^{-1/2}\|_\infty^{-1}$, where the inverse is taken on the range of ρ_1.

Notes and literature The first textbook that covers the mathematical structure of quantum theory including density operators is von Neumann's *Mathematische Grundlagen der Quantenmechanik* [10]. Convergence result for density operators as in Thm.1.7 can be found in [11].
The operational approach of describing measurements via POVMs was introduced in the books of Ludwig [12] and Davies [13].
Simple bounds as the one in Prop.1.12 for the information-carrying capacity of quantum systems can be found in [14]. The tight bound in Eq.(1.24) delineates the limitations of quantum hypothesis testing, which we will discuss in greater detail in Sec.3.1. Historically, this field originates in the works of Helstrom [15], Holevo [16], Belavkin [17, 18], and Yuen, Kennedy, and Lax [19].

1.4 Convexity

Convex sets and extreme points

Definition 1.14. *Let V be a real vector space.*[10]

- *A subset $C \subseteq V$ is called* convex, *if $x, y \in C$ implies that $\lambda x + (1-\lambda)y \in C$ for all $\lambda \in [0,1]$.*
- *For a subset $S \subseteq V$ define the* convex hull $\mathrm{conv}(S)$ *as the set of all finite linear combinations of the form $\sum_{i=1}^n \lambda_i x_i$ with $\lambda_i \geq 0$, $\sum_{i=1}^n \lambda_i = 1$, $x_i \in S$ and $n \in \mathbb{N}$.*
- *The* dimension *of a convex set is the dimension of the affine space generated by it.*
- *An* extreme point *of a convex set C is an element $e \in C$ with the property that $e = \lambda x + (1-\lambda)y$ with $x, y \in C$, $\lambda \in [0,1]$ implies that $e \in \{x, y\}$. We denote the set of extreme points of C by $\mathcal{E}(C)$.*

Example 1.10 (Extreme points of Schatten unit balls)**.** As a consequence of the Schmidt decomposition in Eq.(1.8), the set of extreme points of the unit ball of $\mathcal{B}_1(\mathcal{H})$ is the set of all rank-one operators of the form $|\varphi\rangle\langle\Psi|$, where $\varphi, \psi \in \mathcal{H}$ are unit vectors. For $p \in (1, \infty)$ the unit ball of $\mathcal{B}_p(\mathcal{H})$ turns out to be strictly convex (cf. [20]) so that every element $X \in \mathcal{B}_p(\mathcal{H})$ with $\|X\|_p = 1$ is an extreme point.

Theorem 1.15 (Caratheodory)**.** *Let V be a real vector space, $C \subseteq V$ a compact convex set that spans a real affine space of dimension $d < \infty$, and $x \in C$. There is a set of extreme points $E \subseteq \mathcal{E}(C)$ of size $|E| \leq (d+1)$ so that $x \in \mathrm{conv}(E)$. In particular, $C = \mathrm{conv}(\mathcal{E}(C))$.*

Here, the decomposition into extreme points is unique for all $x \in C$ iff the convex set is a *simplex*, i.e., it has exactly $d+1$ extreme points. The set of probability distributions over a finite set, for instance, forms a simplex.

The infinite-dimensional analog of Caratheodory's theorem requires taking the closure of the set of extreme points. Then the analogous statement is true for all topologies that are 'locally convex'. This means that the topology arises from (semi-)norms, as it is the case for all topologies discussed so far, in particular, for the weak-* topology on $\mathcal{B}(\mathcal{H})$.

Theorem 1.16 (Krein-Milman)**.** *Let V be a locally convex topological vector space and $C \subseteq V$ compact and convex. Then C is the closure of the convex hull of its extreme points, i.e. $\overline{\mathrm{conv}}(\mathcal{E}(C)) = C$.*

By Alaoglu's theorem, in the weak-* topology a set $C \subseteq \mathcal{B}(\mathcal{H})$ is compact iff it is closed and norm-bounded. Hence, Krein-Milman applies especially to the unit ball in $\mathcal{B}(\mathcal{H})$. For that particular case, however, there is a stronger result that holds in the topology of the operator norm:

[10] Note that every complex vector space is in particular a real vector space.

Theorem 1.17 (Russo-Dye, Kadison-Pedersen)**.** *In the operator-norm topology, the unit ball* $\{A \in \mathcal{B}(\mathcal{H}) |\ \|A\| \leq 1\}$ *is the norm-closed convex hull of the set of unitary operators. Specifically, if* $\|A\| \leq 1 - \frac{2}{n}$ *for some* $n \in \mathbb{N}$*, then there are unitaries* $(U_i)_{i=1}^n$ *so that* $\frac{1}{n}(U_1 + \ldots + U_n) = A$*.*

The second part of this theorem (due to Kadison and Pedersen) implies that every element of the unit ball can be approximated up to $2/n$ in operator norm by an equal-weight convex combination of n unitaries. This is reminiscent of the following result that holds for inner product spaces. It has a very elegant proof that exploits the probabilistic method—so we have to state it:

Theorem 1.18 (Maurey)**.** *Let* C *be a subset of an inner product space,* $\phi \in \mathrm{conv}(C)$ *and* $b := \sup_{\xi \in C} \|\xi\|$*. For any* $n \in \mathbb{N}$ *there are elements* $\psi_1, \ldots, \psi_n \in C$ *so that*

$$\left\| \phi - \frac{1}{n} \sum_{i=1}^{n} \psi_i \right\|^2 \leq \frac{b^2}{n}, \tag{1.28}$$

where the norm is the one induced by the inner product.

Proof. As ϕ is in the convex hull of C, there is a finite subset $\Xi \subseteq C$ so that $\phi = \sum_{z \in \Xi} \lambda_z z$, where λ forms a probability distribution over Ξ. Let $Z_1, \ldots, Z_n$ be i.i.d. random variables with values in Ξ, distributed according to λ. Hence, by construction, the expectation values are $\mathbb{E}[Z_i] = \phi$. Using this and the i.i.d. property, it is straightforward to show that

$$\mathbb{E}\left[\left\| \phi - \frac{1}{n} \sum_{i=1}^{n} Z_i \right\|^2 \right] = \frac{1}{n} \big(\mathbb{E}\left[\|Z_i\|^2 \right] - \|\phi\|^2 \big).$$

Here, the r.h.s. can be bounded from above by $\frac{b^2}{n}$. Since the resulting inequality holds for the expectation value, there has to be at least one realization of the random variables for which it is true as well. □

Mixtures of states On any given Hilbert space $\mathcal{H}$, the set of density operators $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{B}_1(\mathcal{H}) | \rho \geq 0, \mathrm{tr}[\rho] = 1\}$ is a convex set: the trace is obviously preserved by convex combinations and the sum of two positive operators is again positive. In fact, slightly more is true: if $(\rho_n)_{n \in \mathbb{N}}$ is any sequence of density operators and $(\lambda_n)_{n \in \mathbb{N}}$ is any sequence of positive numbers that sum up to one, then

$$\sum_{n=1}^{\infty} \lambda_n \rho_n \in \mathcal{S}(\mathcal{H}),$$

where the sequence of partial sums converges in trace norm. In order to see this, realize that it is a Cauchy sequence (as $\left\| \sum_{n=k}^{l} \lambda_n \rho_n \right\|_1 \leq \sum_{n=k}^{l} \lambda_n$ and $\lambda \in l_1(\mathbb{N})$) and that $\mathcal{B}_1(\mathcal{H})$ is a Banach space.

Conversely, every single density operator can be convexly decomposed into pure state density operators via its spectral decomposition, which in this case

coincides with the Schmidt decomposition

$$\rho = \sum_n \lambda_n |\psi_n\rangle\langle\psi_n|,$$

where the λ_n's are the eigenvalues and the ψ_n's the corresponding orthonormal eigenvectors. Pure state density operators can not be convexly decomposed further (Exercise 1.22). Consequently, the pure state density operators are exactly the extreme points of $\mathcal{S}(\mathcal{H})$. If ρ is not pure, there are infinitely many ways of decomposing it convexly into pure states—the spectral decomposition is one of them and distinguishes itself by the fact that the ψ_n's are mutually orthogonal.

Example 1.11 (Decompositions into pure states)**.** For any density operator $\rho \in \mathcal{B}_1(\mathcal{H})$ convex decompositions into pure states can be constructed from any orthonormal basis $\{e_k\}$ via the corresponding resolution of identity in Eq.(1.3): if we multiply Eq.(1.3) from both sides with $\sqrt{\rho}$, we obtain

$$\rho = \sum_k \sqrt{\rho}|e_k\rangle\langle e_k|\sqrt{\rho} = \sum_k p_k |\varphi_k\rangle\langle\varphi_k|, \tag{1.29}$$

with $\varphi_k := \sqrt{\rho}e_k / \left\|\sqrt{\rho}e_k\right\|$ and $p_k := \langle e_k, \rho e_k\rangle$. Since every subspace that has dimension greater than one admits an infinite number of inequivalent orthonormal bases, this construction leads to an infinite number of different decompositions unless ρ is pure. In Cor. 1.23 we will see, that the resulting probability distribution p is always at least as mixed as the distribution of eigenvalues of ρ. Moreover, one can show that all countable convex decompositions into pure states can be obtained in the described way if one allows in addition to first embed isometrically into a larger Hilbert space and then follows the described construction starting from an orthonormal basis of the larger space.

Convex combinations of density operators have a simple operational meaning. To understand this, assume that an experimentalist has two preparation devices at hand, which are described by density operators $\rho_0, \rho_1 \in \mathcal{B}_1(\mathcal{H})$. Assume further, that for every single preparation of the system, she first flips a coin and then uses one of the two devices depending on the outcome, say ρ_1 with probability λ and ρ_0 with probability $1-\lambda$. If eventually a measurement is performed that is described by a POVM M, then the probability of measuring an outcome in Y is given by

$$\lambda p(Y|\rho_1, M) + (1-\lambda)p(Y|\rho_0, M) = \operatorname{tr}\left[\big(\lambda\rho_1 + (1-\lambda)\rho_0\big)M(Y)\right],$$

where Born's rule was used together with the linearity of the trace. Hence, the overall preparation, which now includes the random choice of the experimentalist, is described by the convex combination $\lambda\rho_1 + (1-\lambda)\rho_0$.

Majorization In Prop. 1.6 we saw that the functional $\operatorname{tr}\left[\rho^2\right]$ can be used to quantify how pure or mixed a density operator is. Using functional calculus this can be express as $\operatorname{tr}\left[\rho^2\right] = \operatorname{tr}\left[f(\rho)\right]$ with $f(x) = x^2$. This choice is somewhat

arbitrary since we could instead have used e.g. $f(x) = x^3$, which also orders the set of density operators from the maximally mixed state to the pure states. If $\dim(\mathcal{H}) > 2$, however, the two orders turn out to be inequivalent, i.e. we can find ρ_1, ρ_2 with $\operatorname{tr}\left[\rho_1^2\right] > \operatorname{tr}\left[\rho_2^2\right]$ but $\operatorname{tr}\left[\rho_1^3\right] < \operatorname{tr}\left[\rho_2^3\right]$. So is there any reasonable way of saying that ρ_1 is more mixed (or pure) than ρ_2? The answer to this question is given by a preorder[11] that is based on the notion of *majorization*.

Definition 1.19 (Majorization)**.** *Let λ, μ be two finite (and equal-length) or infinite sequences of non-negative real numbers with $\|\lambda\|_1 = \|\mu\|_1 = 1$. By $\lambda^\downarrow, \mu^\downarrow$ we denote the corresponding sequences rearranged in non-increasing order. We say that λ is* majorized *by μ and we write $\lambda \prec \mu$ if*

$$\sum_{i=1}^{k} \lambda_i^\downarrow \leq \sum_{i=1}^{k} \mu_i^\downarrow \quad \forall k. \tag{1.30}$$

For a pair of density operators $\rho_1, \rho_2 \in \mathcal{B}(\mathcal{H})$ we write $\rho_1 \prec \rho_2$ if the sequence of eigenvalues of ρ_1 is majorized by the one of ρ_2.

We will see that this is closely related to the following concept:

Definition 1.20 (Doubly stochastic matrices)**.** *Let $d \in \mathbb{N} \cup \infty$. A $d \times d$ matrix with non-negative entries M_{ij} is called* doubly stochastic *if for all i:*

$$\sum_{j=1}^{d} M_{ij} = \sum_{j=1}^{d} M_{ji} = 1. \tag{1.31}$$

Example 1.12 (Permutation matrices)**.** Let N be either $\mathbb{N}$ or $\{1, \ldots, d\}$ for $d \in \mathbb{N}$. Then any bijection $\pi : N \to N$ leads to a doubly stochastic matrix via $M_{ij} := \delta_{i,\pi(j)}$ with $i, j \in N$. These are called *permutation matrices*. In the finite-dimensional case, *Birkhoff's theorem* states that permutation matrices form the extreme points of the convex set of doubly stochastic matrices.

Example 1.13 (Unistochastic matrices)**.** Let $U \in \mathcal{B}(\mathcal{H})$ be unitary and $\{e_k\} \subset \mathcal{H}$ an orthonormal basis. Then the matrix with elements $M_{ij} := |\langle e_i, U e_j\rangle|^2$ is called *unistochastic*. This is an example of a doubly stochastic matrix, since

$$\sum_j |\langle e_i, U e_j\rangle|^2 = \sum_j \langle e_i, U^* e_j\rangle\langle e_j, U e_i\rangle = \langle e_i, U^* U e_i\rangle = 1,$$

and similarly for the transposed matrix. Note that in particular, every permutation matrix is unistochastic as it can be obtained by choosing U to be the corresponding permutation of basis elements.

The following relates the concepts discussed so far in this paragraph:

Theorem 1.21. *Let λ, μ be two finite (and equal-length) or infinite sequences of non-negative real numbers with $\|\lambda\|_1 = \|\mu\|_1 = 1$. Then the following are equivalent:*

[11] A *preorder* is a binary relation that is transitive and reflexive.

(i) $\lambda \prec \mu$.

(ii) *There is a doubly stochastic matrix* M *so that* $\lambda = M\mu$.

(iii) *For all continuous convex functions* $f : [0,1] \to \mathbb{R}$ *that satisfy* $f(0) = 0$*:*

$$\sum_k f(\lambda_k) \leq \sum_k f(\mu_k). \tag{1.32}$$

When applied to density operators, this gives:

Corollary 1.22. *Let* $\rho_1, \rho_2 \in \mathcal{B}_1(\mathcal{H})$ *be two density matrices. Then* $\rho_1 \prec \rho_2$ *iff for all continuous convex functions* $f : [0,1] \to \mathbb{R}$ *with* $f(0) = 0$*:* $\operatorname{tr}[f(\rho_1)] \leq \operatorname{tr}[f(\rho_2)]$.

Consequently, majorization is a meaningful way of saying that one density operator is more mixed than another. Note in particular that $\rho \prec |\psi\rangle\langle\psi|$ and for d-dimensional quantum systems $\mathbb{1}/d \prec \rho$ holds for any density operator ρ.

Corollary 1.23. *Let* $\{e_k\} \subset \mathcal{H}$ *be an orthonormal basis,* $\rho \in \mathcal{B}_1(\mathcal{H})$ *a density operator with eigenvalues* (λ_k) *and* $p_k := \langle e_k, \rho e_k\rangle$*. Then*

$$\lambda \succ p. \tag{1.33}$$

Proof. Inserting the spectral decomposition $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ into $p_k = \langle e_k, \rho e_k\rangle$, we obtain $p = M\lambda$ with $M_{ki} := |\langle e_k, \psi_i\rangle|^2$. Since we can express $\psi_i = Ue_i$ for a suitable unitary U, we get that M is an unistochastic matrix, so that by Thm.1.21 $\lambda \succ p$. □

In the context of Example 1.11, this result implies that among all the decompositions into pure states to which Eq.(1.29) gives rise, the spectral decomposition is the least mixed.

Convex functionals In this paragraph we have a closer look at convex functionals that are defined on sets of Hermitian operators and constructed from convex functions in a single real variable by means of functional calculus (cf. p. 13).

Theorem 1.24. *Let* $f : [a,b] \subset \mathbb{R} \to \mathbb{R}$ *be a continuous convex function and* $A \in \mathcal{B}_\infty(\mathcal{H})$ *Hermitian with* $\operatorname{spec}(A) \subseteq [a,b]$*. Then, for every unit vector* $\psi \in \mathcal{H}$*:*

$$f\big(\langle\psi, A\psi\rangle\big) \leq \langle\psi, f(A)\psi\rangle. \tag{1.34}$$

Proof. First observe that $c := \langle\psi, A\psi\rangle \in [a,b]$ since $a\mathbb{1} \leq A \leq b\mathbb{1}$. Assume for the moment that $c \in (a,b)$. By convexity of f we can find an affine function $l : [a,b] \to \mathbb{R}$ such that $f \geq l$ and $f(c) = l(c)$. Then $f(A) \geq l(A)$ and therefore

$$\langle\psi, f(A)\psi\rangle \geq \langle\psi, l(A)\psi\rangle = l(c) = f(c) = f\big(\langle\psi, A\psi\rangle\big),$$

where we have used that $l(A) = \alpha\mathbb{1} + \beta A$ if l is of the form $l(x) = \alpha + \beta x$. It remains to discuss the case $c \in \{a, b\}$. In this case, a linear function with the stated properties might not exist if f is 'infinitely steep' at the boundary. However, for any $\epsilon > 0$ we can still find a linear function l with $l \leq f$ so that $f(c) - l(c) \leq \epsilon$. Following the same argument and using that we can choose any $\epsilon > 0$ then completes the proof. □

Corollary 1.25 (Convex trace functions)**.** *Let $f : [0, 1] \to \mathbb{R}_+$ be convex, continuous and so that $f(0) = 0$. Define $\mathcal{C}(\mathcal{H}) := \{A \in \mathcal{B}_\infty(\mathcal{H}) | 0 \leq A \leq \mathbb{1}\}$ and $F : \mathcal{C}(\mathcal{H}) \to \mathbb{R}$, $F(A) := \operatorname{tr}[f(A)] \in [0, \infty]$. Then:*

(i) F is convex on $\mathcal{C}(\mathcal{H})$.

(ii) For any $A \in \mathcal{C}(\mathcal{H})$ and any orthonormal basis $\{e_k\}$ of $\mathcal{H}$:

$$F(A) \geq \sum_k f\big(\langle e_k, Ae_k\rangle\big). \tag{1.35}$$

Proof. (i) Let $A_\lambda := \lambda A_1 + (1-\lambda)A_0$ be a convex combination of $A_0, A_1 \in \mathcal{C}(\mathcal{H})$ and $\{\psi_k\} \subset \mathcal{H}$ an orthonormal basis of eigenvectors of A_λ. Then

$$\begin{aligned}
\lambda F(A_1) + (1-\lambda)F(A_0) &= \lambda \sum_k \langle \psi_k, f(A_1)\psi_k\rangle + (1-\lambda)\sum_k \langle \psi_k, f(A_0)\psi_k\rangle \\
&\geq \lambda \sum_k f\big(\langle \psi_k, A_1\psi_k\rangle\big) + (1-\lambda)\sum_k f\big(\langle \psi_k, A_0\psi_k\rangle\big) \\
&\geq \sum_k f\big(\langle \psi_k, A_\lambda\psi_k\rangle\big) = F(A_\lambda).
\end{aligned}$$

Here, the first inequality is due to Eq.(1.34), the second inequality uses convexity of f and the last step uses that $(\langle \psi_k, A_\lambda\psi_k\rangle)$ is the sequence of eigenvalues of A_λ.
(ii) follows from Eq.(1.34) applied to each term in $F(A) = \sum_k \langle e_k, f(A)e_k\rangle$. □

The following useful observation also enables to lift inequalities of scalar functions to inequalities of functions of operators under the trace:

Lemma 1.26. *Let $I \subseteq \mathbb{R}$ be an open interval. If $f_i, g_i : I \to \mathbb{R}$ and $\alpha_i \in \mathbb{R}$ for $i \in \{i, \ldots, n\}$ satisfy*

$$\begin{aligned}
\sum_{i=1}^n \alpha_i f_i(a) g_i(b) &\geq 0 \quad \forall a, b \in I, \text{ then} \\
\sum_{i=1}^n \alpha_i \operatorname{tr}[f_i(A) g_i(B)] &\geq 0
\end{aligned} \tag{1.36}$$

holds for all Hermitian $A, B \in \mathcal{B}(\mathbb{C}^d)$ whose spectra are contained in I.

Proof. Inserting spectral decompositions $A = \sum_k \lambda_k |e_k\rangle\langle e_k|$ and $B = \sum_l \mu_l |f_l\rangle\langle f_l|$ we obtain

$$\sum_{i=1}^{n} \alpha_i \operatorname{tr}\left[f_i(A) g_i(B)\right] = \sum_{k,l} |\langle e_k, f_l\rangle|^2 \sum_{i=1}^{n} \alpha_i f_i(\lambda_k) g_i(\mu_l) \geq 0.$$

□

Corollary 1.27 (Klein inequalities)**.** *Let $I \subseteq \mathbb{R}$ be an open interval, and $A, B \in \mathcal{B}(\mathbb{C}^d)$ Hermitian with spectra in I. If $f : I \to \mathbb{R}$ is convex and differentiable, then*

$$\operatorname{tr}\left[f(A) - f(B)\right] \geq \operatorname{tr}\left[(A - B) f'(B)\right]. \tag{1.37}$$

If f is twice differentiable and strongly convex, i.e. $\inf_{x \in I} f''(x) =: c > 0$, then

$$\operatorname{tr}\left[f(A) - f(B)\right] - \operatorname{tr}\left[(A - B) f'(B)\right] \geq \frac{c}{2} \|A - B\|_2^2. \tag{1.38}$$

Proof. Both inequalities exploit Lemma 1.26. Eq.(1.37) then follows from the fact that every convex function satisfies $f(a) - f(b) \geq (a-b) f'(b)$ and Eq.(1.38) uses the mean-value version of Taylor's theorem, which states that there is a $z \in [a, b]$ such that

$$f(a) = f(b) + (a - b) f'(b) + \frac{1}{2}(a - b)^2 f''(z).$$

□

Entropy An important example of a convex trace function is the negative *entropy*. Its classical manifestations are ubiquitous in information theory, statistical physics, probability theory, and thermodynamics.

Definition 1.28 (Entropy)**.** *The* von Neumann entropy *(short* entropy*) of a density operator $\rho \in \mathcal{B}_1(\mathcal{H})$ is defined as $S(\rho) := \operatorname{tr}\left[h(\rho)\right]$, where $h(x) := -x \log x$ with $h(0) := 0$.*

Remarks: 1. Occasionally, it is useful to have the entropy functional defined outside the set of density operators: if $\sigma \in \mathcal{B}_1(\mathcal{H})$ is positive semidefinite and $\lambda := \operatorname{tr}[\sigma]$, then we define $S(\sigma) := \lambda S(\sigma/\lambda)$. In this way, S becomes positive homogeneous in the sense that $S(\mu\rho) = \mu S(\rho)$, for all $\mu \in [0, \infty)$.

2. The *Shannon entropy* of discrete probability distribution (p_x) is defined as $S(p) := -\sum_x p_x \log p_x$. Therefore, the von Neumann entropy of a density operator is the Shannon entropy of its spectrum (regarded as a multiset).

Depending on the field, different bases of the logarithm are used: the natural choice in information theory is $\log_2$, whereas in thermodynamics and statistical physics, the natural logarithm ln is used. An advantage of the latter is, of course, that it occurs naturally and behaves nicely in the presence of differentiation. For instance, we get that for every density operator ρ and with $\log = \ln$ that

$$S(\rho) = -\frac{d}{dp}\bigg|_{p=1} \operatorname{tr}\left[\rho^p\right]. \tag{1.39}$$

On the relevant interval $[0,1]$ the function h is non-negative, continuous and concave. By Cor.1.25 (i) this implies that the von Neumann entropy S is a non-negative, concave functional on the set of density operators. From Cor. 1.22 we get

$$\rho_1 \prec \rho_2 \quad \Rightarrow \quad S(\rho_1) \geq S(\rho_2).$$

Concavity of the entropy means that the entropy of a convex combination is at least as large as the convex combination of the individual entropies. The following shows that the difference between the two is at most the 'classical' entropy of the involved probability distribution that is formed by the convex weights:

Proposition 1.29. *Let $\rho := \sum_k \lambda_k \rho_k$ be a convex combination of density operators $\rho_k \in \mathcal{B}(\mathcal{H})$. Then*

$$0 \;\leq\; S(\rho) - \sum_k \lambda_k S(\rho_k) \;\leq\; S(\lambda) := -\sum_k \lambda_k \log \lambda_k. \tag{1.40}$$

Moreover, equality holds in the upper bound if the ρ_k's are pairwise orthogonal, i.e., if $k \neq l \Rightarrow \rho_k \rho_l = 0$.

Proof. The lower bound is just concavity. The upper bound follows from

$$\begin{aligned} S(\lambda) + \sum_k \lambda_k S(\rho_k) &= -\sum_k \operatorname{tr}\left[\lambda_k \rho_k \log \lambda_k \mathbb{1}\right] + \operatorname{tr}\left[\lambda_k \rho_k \log \rho_k\right] \\ &= -\sum_k \operatorname{tr}\left[\lambda_k \rho_k \log \lambda_k \rho_k\right] \\ &\geq -\sum_k \operatorname{tr}\left[\lambda_k \rho_k \log \rho\right] \;=\; S(\rho), \end{aligned} \tag{1.41}$$

where the step to Eq.(1.41) exploits that $\lambda_k \rho_k \leq \rho$ and thus, by the operator monotonicity of the logarithm, $\log \lambda_k \rho_k \leq \log \rho$. If the ρ_k's are pairwise orthogonal, then $\operatorname{tr}\left[\lambda_k \rho_k \log \sum_l \lambda_l \rho_l\right] = \operatorname{tr}\left[\lambda_k \rho_k \log \lambda_k \rho_k\right]$, so that equality holds. □

For finite-dimensional Hilbert spaces the von Neumann entropy is continuous, which is implied by the continuity of the eigenvalues. In infinite dimensions, continuity has to be relaxed to lower semicontinuity. This means $\liminf_{\rho \to \rho_0} S(\rho) \geq S(\rho_0)$ (cf. Example 1.14 and Exercise 1.25).

Since $h(x) = 0$ iff $x \in \{0,1\}$ we get that $S(\rho) = 0$ iff ρ is pure. On $\mathbb{C}^d$ the maximum $S(\rho) = \log d$ is attained iff $\rho = \mathbb{1}/d$ is maximally mixed. The infinite-dimensional case is elucidated by the following example:

Example 1.14 (Infinite entropy)**.** Consider a sequence $p_n := c/\big(n(\log n)^\gamma\big)$ for $n > 2$, $\gamma \in (1,2)$ and c a positive constant to be chosen shortly. From $\int 1/(x(\log x)^\gamma)dx = (\log x)^{1-\gamma}/(1-\gamma)$ it follows that $p \in l_1(\mathbb{N})$ so that we can choose c in a way that $\sum_n p_n = 1$. However, $-\sum_n p_n \log p_n = \infty$ due to the divergence of the integral $\int 1/(x(\log x)^{\gamma-1})dx$. Hence, if σ is a density operator with eigenvalues (p_n), then $S(\sigma) = \infty$. Moreover, if ρ is any density operator, then $S\big((1-\epsilon)\rho + \epsilon\sigma\big) \geq (1-\epsilon)S(\rho) + \epsilon S(\sigma) = \infty$ for any $\epsilon > 0$.

Consequently, on an infinite dimensional Hilbert space, the density operators with infinite entropy are trace-norm dense in the set of all density operators.

A useful Lemma, in particular for extending properties of the entropy of finite-dimensional systems to infinite-dimensional ones, is the following:

Lemma 1.30. *Let $\rho \in \mathcal{B}(\mathcal{H})$ be a density operator and let $P_n \in \mathcal{B}(\mathcal{H})$ form a sequence of orthogonal projectors converging strongly to the identity, i.e.,* $\lim_{n\to\infty} P_n = \mathbb{1}$. *Then*

$$\lim_{n\to\infty} S\big(P_n \rho P_n\big) = S(\rho). \tag{1.42}$$

Exercise 1.21. Show that every $A \in \mathcal{B}(\mathbb{C}^d)$ with operator norm $\|A\| \leq 1$ is an equal weight convex combination of a pair of unitaries $U_1, U_2 \in \mathcal{B}(\mathbb{C}^d)$, i.e., $A = (U_1 + U_2)/2$. (Hint: use the polar decomposition.)

Exercise 1.22. Show that pure states are extreme points of the convex set of density operators.

Exercise 1.23. Let $\rho_1, \rho_2 \in \mathcal{B}(\mathbb{C}^d)$ be two density operators. Prove that $\rho_1 \prec \rho_2$ iff there exist a finite set of unitaries $U_i \in \mathcal{B}(\mathbb{C}^d)$ and corresponding probabilities $p_i > 0$, $\sum_i p_i = 1$ so that $\rho_1 = \sum_i p_i U_i \rho_2 U_i^*$.

Exercise 1.24. Denote by $\mathcal{U}_n$ all maps from $\mathcal{B}(\mathbb{C}^d)$ to itself that are of the form $\mathcal{B}(\mathbb{C}^d) \ni \rho \mapsto \sum_{i=1}^n p_i U_i \rho U_i^*$, for some $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$ and unitaries $U_i \in \mathcal{B}(\mathbb{C}^d)$. Determine an $m \in \mathbb{N}$ (as a function of d) such that $\mathcal{U}_m = \bigcup_{n\in\mathbb{N}} \mathcal{U}_n$.

Exercise 1.25. Construct a sequence of density operators of finite entropy that converges in trace-norm to a pure state but has entropy diverging to ∞.

Notes and literature Useful general references on convexity are the books by Simon [21] and Rockafellar [22]. The so-called *Caratheodory number* of maximally required extreme points that appears in Caratheodory's theorem Thm. 1.15 can often be bound more tightly by the length n of the longest chain

$$\mathcal{F}_1 \subsetneq \mathcal{F}_2 \subsetneq \cdots \subsetneq \mathcal{F}_n = C$$

of faces of C [23]. For instance, if C is strictly convex, the longest such chain is $\{x\} \subseteq C$ for any $x \in \mathcal{E}(C)$ so that the $d+1$ in Caratheodory's theorem can be replaced by 2 in this case.
Thm.1.17 is due to Kadison and Pedersen [24]. A simple consequence thereof is that every element of $\mathcal{B}(\mathcal{H})$ is a positive multiple of a sum of three unitaries (or two in the finite-dimensional case, as observed in Ex.1.21). Another result of [24] is that every convex combination of n unitaries is also a mean (i.e., equal-weight convex combination) of n unitaries. The approximate Caratheodory result of Thm.1.18 is mentioned in [25] and extended to other Banach spaces in [26].
Lemma 1.30 is from B. Simon's appendix in [27]. More general results in this direction can be found in [28].

1.5 Composite systems and tensor products

For all kinds of mathematical spaces there are three basic ways of constructing new spaces from old ones: quotients, sums and products. In the case of Hilbert spaces, we have essentially discussed quotients already since the quotient of a Hilbert space $\mathcal{H}$ by a subspace V can be identified with the orthogonal complement $V^\perp$ in $\mathcal{H}$. In this section, we will have a closer look at the two remaining constructions: direct sums and, in particular, tensor products.

Direct sums We begin with the simpler construction:

Definition 1.31 (Direct sum)**.** *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces. Their* direct sum *is the Hilbert space $\mathcal{H}_1 \oplus \mathcal{H}_2 := \{(\psi, \varphi) \in \mathcal{H}_1 \times \mathcal{H}_2\}$ with inner product*

$$\langle (\psi_1, \varphi_1), (\psi_2, \varphi_2) \rangle := \langle \psi_1, \psi_2 \rangle + \langle \varphi_1, \varphi_2 \rangle.$$

Instead of (ψ, φ) we also write $\psi \oplus \varphi$ for the elements of $\mathcal{H}_1 \oplus \mathcal{H}_2$.

This construction leads to a Hilbert space of dimension $\dim(\mathcal{H}_1 \oplus \mathcal{H}_2) = \dim(\mathcal{H}_1) + \dim(\mathcal{H}_2)$. $\mathcal{H}_1$ and $\mathcal{H}_2$ can be regarded as embedded mutually orthogonal subspaces $\mathcal{H}_1 \oplus 0$ and $0 \oplus \mathcal{H}_2$ of $\mathcal{H}_1 \oplus \mathcal{H}_2$. For a finite number of Hilbert spaces the definition of $\bigoplus_n \mathcal{H}_n$ extends immediately and it is associative, i.e. $(\mathcal{H}_1 \oplus \mathcal{H}_2) \oplus \mathcal{H}_3 = \mathcal{H}_1 \oplus (\mathcal{H}_2 \oplus \mathcal{H}_3)$. For an infinite sequence $(\mathcal{H}_n)_{n \in \mathbb{N}}$ of Hilbert spaces, one defines the corresponding infinite direct sum Hilbert space as

$$\bigoplus_{n \in \mathbb{N}} \mathcal{H}_n := \big\{ (\varphi_n)_{n \in \mathbb{N}} \big| \varphi_n \in \mathcal{H}_n, \sum_{n \in \mathbb{N}} \|\varphi_n\|^2 < \infty \big\},$$

with inner product $\langle (\varphi_n)_{n \in \mathbb{N}}, (\psi_n)_{n \in \mathbb{N}} \rangle := \sum_{n \in \mathbb{N}} \langle \varphi_n, \psi_n \rangle$.

For $A \in \mathcal{B}(\mathcal{H}_1)$, $B \in \mathcal{B}(\mathcal{H}_2)$ we can define $(A \oplus B) \in \mathcal{B}(\mathcal{H}_1 \oplus \mathcal{H}_2)$ via $(A \oplus B)\varphi \oplus \psi := A\varphi \oplus B\psi$. It is then straightforward to show that $\|A \oplus B\| = \max\{\|A\|, \|B\|\}$ and that $A, B \geq 0$ implies $A \oplus B \geq 0$. When expressed as a matrix $A \oplus B$ simply becomes the block diagonal matrix

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

Tensor products

Definition 1.32 (Tensor product Hilbert space)**.** *For any pair $\psi_1 \in \mathcal{H}_1, \psi_2 \in \mathcal{H}_2$ define a conjugate-bilinear functional $\psi_1 \otimes \psi_2 : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ by $(\alpha, \beta) \mapsto \langle \alpha, \psi_1 \rangle \langle \beta, \psi_2 \rangle$. The* algebraic tensor product *of $\mathcal{H}_1$ and $\mathcal{H}_2$ is defined as the space of all finite linear combinations of maps of the form $\psi_1 \otimes \psi_2$. The* tensor product Hilbert space *$\mathcal{H}_1 \otimes \mathcal{H}_2$ of $\mathcal{H}_1$ and $\mathcal{H}_2$ is defined as the completion of the algebraic tensor product w.r.t. the inner product*

$$\langle \varphi_1 \otimes \varphi_2, \psi_1 \otimes \psi_2 \rangle := \langle \varphi_1, \psi_1 \rangle \langle \varphi_2, \psi_2 \rangle, \tag{1.43}$$

extended by linearity and continuity to the whole space.

If several Hilbert spaces are combined via the tensor product or direct sum construction, then the following Hilbert space isomorphisms hold:

$$\begin{aligned} \mathcal{H}_1 \otimes \mathcal{H}_2 &\simeq \mathcal{H}_2 \otimes \mathcal{H}_1, \\ (\mathcal{H}_1 \otimes \mathcal{H}_2) \otimes \mathcal{H}_3 &\simeq \mathcal{H}_1 \otimes (\mathcal{H}_2 \otimes \mathcal{H}_3), \\ \mathcal{H}_1 \otimes (\mathcal{H}_2 \oplus \mathcal{H}_3) &\simeq (\mathcal{H}_1 \otimes \mathcal{H}_2) \oplus (\mathcal{H}_1 \otimes \mathcal{H}_3). \end{aligned} \tag{1.44}$$

It should be noted that the concrete construction of $\mathcal{H}_1 \otimes \mathcal{H}_2$, which appears in terms of conjugate-bilinear maps in the above definition, is usually not used. What is used a lot, however, are the resulting properties. In particular linearity:

$$\left(\sum_{i=1}^{k} \psi_i\right) \otimes \left(\sum_{j=1}^{l} \varphi_j\right) = \sum_{i=1}^{k}\sum_{j=1}^{l} \psi_i \otimes \varphi_j, \tag{1.45}$$

$$(c\psi) \otimes \varphi = c(\psi \otimes \varphi) = \psi \otimes (c\varphi), \quad \text{for } c \in \mathbb{C}. \tag{1.46}$$

The constructed Hilbert space has $\dim(\mathcal{H}_1 \otimes \mathcal{H}_2) = \dim(\mathcal{H}_1)\dim(\mathcal{H}_2)$. In fact, every pair of orthonormal bases $\{e_k\} \subset \mathcal{H}_1$, $\{f_l\} \subset \mathcal{H}_2$ gives rise to an orthonormal basis $\{e_k \otimes f_l\} \subset \mathcal{H}_1 \otimes \mathcal{H}_2$. Such a basis is called a *product basis* as all its elements are simple products. Expanding an element $\Psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ in this basis leads to $\Psi = \sum_{k,l} \Psi_{k,l} e_k \otimes f_l$, where $\Psi_{k,l} := \langle e_k \otimes f_l, \Psi\rangle$ satisfies $\|\Psi\|^2 = \sum_{k,l} |\Psi_{k,l}|^2$ by Parseval's identity. The right-hand side of this identity looks like the square of the Hilbert-Schmidt-norm of the 'matrix' $(\Psi_{k,l})$. Hence, the expansion suggests an isomorphism between elements of the tensor product Hilbert space and elements of the space of Hilbert-Schmidt class operators. This is formalized in the following:

Theorem 1.33 (Hilbert-Schmidt isomorphism)**.** *The tensor product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is isomorphic to the space of Hilbert-Schmidt-class operators $\mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$. That is, there is a linear bijection $\mathcal{I} : \mathcal{H}_1 \otimes \mathcal{H}_2 \to \mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ so that for all $\Psi, \Phi \in \mathcal{H}_1 \otimes \mathcal{H}_2$:*

$$\langle \Phi, \Psi\rangle = \operatorname{tr}\left[\mathcal{I}(\Phi)^* \mathcal{I}(\Psi)\right]. \tag{1.47}$$

Proof. We could simply argue that the respective orthonormal bases have the same cardinality and thus there has to be an isomorphism. For later use, however, we follow a more explicit route. For that, it is convenient to introduce the complex conjugate $\overline{\psi} := \sum_k \langle \psi, e_k\rangle e_k$ of an arbitrary element $\psi \in \mathcal{H}_1$ w.r.t. a fixed orthonormal basis $\{e_k\} \subset \mathcal{H}_1$. Note that the operation $\psi \mapsto \overline{\psi}$ is an involution that preserves the norm as well as orthogonality. Now we define

$$\mathcal{I} : |\psi\rangle \otimes |\varphi\rangle \mapsto |\varphi\rangle\langle\overline{\psi}| \tag{1.48}$$

and extend it by linearity and continuity to the entire space. Then $\mathcal{I}$ is the sought Hilbert space isomorphism since it is a bijection between orthonormal bases: a product basis $|e_k\rangle \otimes |f_l\rangle$ of $\mathcal{H}_1 \otimes \mathcal{H}_2$ and a basis of rank-one operators $|f_l\rangle\langle e_k|$ of $\mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$. □

An important application of this isomorphism is a normal form for elements of a tensor product Hilbert space:

Theorem 1.34 (Schmidt decomposition for tensor products)**.** *For every $\Psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ there is an $r \in \mathbb{N} \cup \{\infty\}$, a sequence of strictly positive numbers $(s_i)_{i=1}^r$ and orthonormal bases $\{e_k\} \subset \mathcal{H}_1$, $\{f_l\} \subset \mathcal{H}_2$ such that*

$$\Psi = \sum_{i=1}^{r} s_i \, e_i \otimes f_i. \tag{1.49}$$

Moreover, the s_i's (called Schmidt coefficients*) are as a multiset uniquely determined by Ψ and satisfy $\sum_{i=1}^r s_i^2 = \|\Psi\|^2$.*

Proof. We exploit the isomorphism from Thm.1.33 together with the fact that $\mathcal{I}(\Psi)$ is a compact operator for which there is a Schmidt decomposition $\mathcal{I}(\Psi) = \sum_i s_i |f_i\rangle\langle\overline{e_i}|$. Applying the inverse $\mathcal{I}^{-1}$ and using Eq.(1.48) then proves the decomposition in Eq.(1.49). Uniqueness of the s_i's follows from the uniqueness of the multiset of singular values of compact operators and $\sum_{i=1}^r s_i^2 = \|\Psi\|^2$ is an application of Parseval's identity. □

Since the Schmidt coefficients are uniquely determined by Ψ, the same is true for their number r, which is called the *Schmidt rank* of Ψ. Obviously, $r \leq \min\{\dim(\mathcal{H}_1), \dim(\mathcal{H}_2)\}$ and $r = 1$ iff Ψ is a *simple tensor*, i.e. of the form $\Psi = \varphi_1 \otimes \varphi_2$ for some $\varphi_i \in \mathcal{H}_i$.

Example 1.15 (Maximally entangled states)**.** A pure state represented by a unit vector $\Psi \in \mathbb{C}^d \otimes \mathbb{C}^d$ is called a d-dimensional *maximally entangled state* if all its Schmidt coefficients are equal to $1/\sqrt{d}$ (and thus $r = d$). The isomorphism in Thm.1.33 then yields a bijection between the set of d-dimensional maximally mixed states and the projective unitary group $PU(d)$ (i.e., the quotient of $U(d)$ by $U(1)$, which corresponds to the phases that lead to equivalent states). In particular, the Hilbert-Schmidt-orthogonal basis of unitaries from Eq.(1.11) then leads to an orthonormal basis $\Psi_{k,l} := \mathcal{I}^{-1}(U_{k,l})/\sqrt{d}$ in $\mathbb{C}^d \otimes \mathbb{C}^d$ that consists of d^2 maximally entangled states.

Before we discuss further properties of the Hilbert-Schmidt isomorphism, we need to introduce the tensor product of operators. For $A_i \in \mathcal{B}(\mathcal{H}_i)$ one defines the tensor product $A_1 \otimes A_2$ as an operator on $\mathcal{H}_1 \otimes \mathcal{H}_2$ via $(A_1 \otimes A_2)(\psi_1 \otimes \psi_2) := (A_1\psi_1) \otimes (A_2\psi_2)$ and its extension by linearity. Then $(A_1 \otimes A_2)^* = A_1^* \otimes A_2^*$ and if $B_i \in \mathcal{B}(\mathcal{H}_i)$ then

$$(A_1 \otimes A_2)(B_1 \otimes B_2) = (A_1B_1) \otimes (A_2B_2). \tag{1.50}$$

The tensor product can be shown to preserve properties like unitarity, positivity, Hermiticity, normality, boundedness, compactness, trace-class or Hilbert-Schmidt-class. That is, if both A_1 and A_2 have one of these properties, then so does $A_1 \otimes A_2$. More specifically, $\|A_1 \otimes A_2\|_p = \|A_1\|_p \|A_2\|_p$ holds for all $p \in [1, \infty]$ and if A_1, A_2 are trace-class, then $\operatorname{tr}[A_1 \otimes A_2] = \operatorname{tr}[A_1] \operatorname{tr}[A_2]$.

A useful representation of the tensor product in the finite-dimensional case is the *Kronecker product* of matrices: if A and B are finite matrices, then $A \otimes B$ can be represented as a block matrix

$$\begin{pmatrix} A_{11}B & A_{12}B & \cdots \\ A_{21}B & A_{22}B & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Now, let us have a closer look at properties of the particular Hilbert-Schmidt isomorphism that we used in the proof of Thm.1.33 and see how it treats tensor products of operators:

Corollary 1.35. *Let $\mathcal{I} : \mathcal{H}_1 \otimes \mathcal{H}_2 \to \mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ be the Hilbert-Schmidt isomorphism constructed via Eq.(1.48), and consider any $\Psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$.*

(i) For any $A \in \mathcal{B}(\mathcal{H}_1)$, $B \in \mathcal{B}(\mathcal{H}_2)$ we have $\mathcal{I} : (A \otimes B)\Psi \mapsto B\mathcal{I}(\Psi)A^T$, where A^T is the transpose of A in the basis used to define $\mathcal{I}$.

(ii) If $\mathcal{H}_1 \simeq \mathcal{H}_2 \simeq \mathbb{C}^d$ and $\mathcal{I}(\Psi)$ is invertible, then for any $A \in \mathcal{B}(\mathcal{H}_1)$ there is a $B \in \mathcal{B}(\mathcal{H}_2)$, which can be obtained from A via a similarity transformation, so that

$$(A \otimes \mathbb{1})\Psi = (\mathbb{1} \otimes B)\Psi. \tag{1.51}$$

If Ψ is maximally entangled, then B has the same singular values as A. In particular, if $\mathcal{I}(\Psi) = \mathbb{1}/\sqrt{d}$, then $B = A^T$.

Proof. (i) follows from the defining equation of the isomorphism, Eq.(1.48), via $(A \otimes B)|\psi\rangle \otimes |\varphi\rangle = |A\psi\rangle \otimes |B\varphi\rangle \mapsto |B\varphi\rangle\langle\overline{A\psi}| = B|\varphi\rangle\langle\overline{\psi}|A^T$.

Eq.(1.51) in (ii) follows from (i) by setting $B := \mathcal{I}(\Psi)A^T\mathcal{I}(\Psi)^{-1}$. Since A is similar to A^T, B is similar to A. If in addition Ψ is maximally entangled, then $\sqrt{d}\mathcal{I}(\Psi)$ is a unitary, so that the claim follows by inserting the singular value decomposition of A. □

Eq.(1.51), especially for maximally entangled Ψ, will play a crucial role in applications such as quantum teleportation or quantum super-dense coding.

Let us finally have a closer look at tensor products of more than two spaces and start with some popular examples:

Example 1.16 (GHZ and W-states)**.** As a shorthand for $e_k \otimes f_l \otimes g_m$, where k, l, m each label elements of an orthonormal basis, it is sometimes convenient to write $|k\ l\ m\rangle$. Using this notation, two prominent examples of states in $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ are the *Greenberger-Horne-Zeilinger (GHZ) state* $(|000\rangle + |111\rangle)/\sqrt{2}$ and the *W-state* $(|100\rangle + |010\rangle + |001\rangle)/\sqrt{3}$.

Definition 1.36 (Tensor rank)**.** *The* tensor rank *of an element $\Psi \in \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_m$, is defined as $\mathcal{R}(\Psi) := \min\left\{r \in \mathbb{N} |\ \Psi = \sum_{i=1}^r \psi_i^{(1)} \otimes \ldots \otimes \psi_i^{(m)},\ \psi_i^{(k)} \in \mathcal{H}_k\right\}$.*

The case $m = 2$ turns out to be significantly simpler and more well-behaved than $m > 2$. For instance:

Proposition 1.37. *Let $\mathcal{H} := \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_m$ be a tensor product of spaces that satisfy $2 \leq \dim(\mathcal{H}_i) < \infty$ and let $\mathcal{R} : \mathcal{H} \to \mathbb{N}$ be the tensor rank. For $m = 2$ the tensor rank is lower semi-continuous on $\mathcal{H}$ and $\mathcal{R}(\Psi)$ equals the Schmidt rank of Ψ. For $m \geq 3$ there are converging sequences $\Psi_n \to \Psi$ for $n \to \infty$ with $\mathcal{R}(\Psi_n) < \mathcal{R}(\Psi)$.*

Proof. For $m = 2$ we can exploit the Hilbert-Schmidt isomorphism from Thm.1.33, which then relates the tensor rank of Ψ to the rank of the operator $\mathcal{I}(\Psi)$. The latter is equal to the Schmidt rank and known to be lower semi-continuous. One way of showing that the rank of a matrix is lower semi-continuous is to argue that the rank of a matrix is at most k iff all $(k+1) \times (k+1)$ minors vanish. As the zero-set of a finite number of polynomials, this forms a closed set so that $\Psi_n \to \Psi$ implies $\liminf \mathcal{R}(\Psi_n) \geq \mathcal{R}(\Psi)$.

For $m > 2$ consider the simplest case $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$, which can be embedded into all larger spaces. Denote by $e, f \in \mathbb{C}^2$ two orthogonal unit vectors. The unnormalized W-state $\Psi = e \otimes e \otimes f + e \otimes f \otimes e + f \otimes e \otimes e$ can be shown to have tensor rank three. However, it can be obtained as a limit of

$$\Psi_n = n\Big(e + \frac{1}{n} f\Big) \otimes \Big(e + \frac{1}{n} f\Big) \otimes \Big(e + \frac{1}{n} f\Big) - n\, e \otimes e \otimes e. \tag{1.52}$$

Consequently, for $m > 2$ the set $\{\Psi \in \mathcal{H} | \mathcal{R}(\Psi) \leq k\}$ is not closed in general. □

Example 1.17 (Matrix-multiplication tensor)**.** Consider $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ where all three tensor factors are matrix spaces of the form $\mathbb{C}^{d \times d}$. Denoting the matrix units by $(e_{kl})_{ij} := \delta_{k,i}\delta_{l,j}$ the *matrix-multiplication tensor* is defined as $T := \sum_{k,l,m=1}^{d} e_{kl} \otimes e_{lm} \otimes e_{mk}$. With its help, the matrix-product of two matrices $A, B \in \mathbb{C}^{d \times d}$ can be expressed as $(AB)_{\alpha\beta} = \mathrm{tr}\,[T(B \otimes A \otimes e_{\beta\alpha})]$. If T has tensor rank $\mathcal{R}(T) = r$, then there are linear maps $a, b : \mathbb{C}^{d \times d} \to \mathbb{C}^r$ and matrices $(C_i)_{i=1}^r \subset \mathbb{C}^{d \times d}$ so that for every $A, B \in \mathbb{C}^{d \times d}$ we have

$$(AB)_{\alpha\beta} = \sum_{i=1}^{r} C_{i,\alpha\beta}\, a(A)_i\, b(B)_i. \tag{1.53}$$

This can be seen by inserting the assumed form $T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$ and taking traces. Eq.(1.53) means that the elements of AB can be obtained as linear combinations of the r products $a(A)_i b(B_i)$. In this way, and by using recursion, the (so far unkonwn) tensor rank of T provides an upper bound on the (so far unknown) complexity of matrix multiplication. Note that naive matrix multiplication would require d^3 products but, as Strassen has observed, $\mathcal{R}(T) < d^3$. Specifically, for $d = 2$ he found $\mathcal{R}(T) = 7$.

Partial trace In classical probability theory, if we have a pair of random variables with a given joint distribution, then there is a well-defined way of assigning a *marginal distribution* to each of the random variables individually. In the following theorem, we construct the quantum analog of this marginalizing map. The analogy will then be made clearer in the subsequent paragraph.

Theorem 1.38 (Partial trace)**.** *There is a unique map (called* partial trace*)* $\mathrm{tr}_2 : \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}_1(\mathcal{H}_1)$ *for which*

$$\mathrm{tr}\,[B(A \otimes \mathbb{1})] = \mathrm{tr}\,[\mathrm{tr}_2[B]A]\,, \quad \forall A \in \mathcal{B}(\mathcal{H}_1), \tag{1.54}$$

holds for all $B \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$*. Moreover,* tr_2 *is trace-norm continuous and*

$$\begin{aligned} B \geq 0 \quad &\Rightarrow \quad \mathrm{tr}_2[B] \geq 0, \\ B = B_1 \otimes B_2 \quad &\Rightarrow \quad \mathrm{tr}_2[B] = B_1 \mathrm{tr}\,[B_2]\,, \\ \mathrm{tr}\,[\mathrm{tr}_2[B]] \quad &= \quad \mathrm{tr}\,[B]\,. \end{aligned} \tag{1.55}$$

Proof. For any unit vector $\psi \in \mathcal{H}_2$ define a bounded linear map $\mathbb{1} \otimes \langle\psi| : \mathcal{H}_1 \otimes \mathcal{H}_2 \to \mathcal{H}_1$ via $\varphi_1 \otimes \varphi_2 \mapsto \varphi_1 \langle\psi, \varphi_2\rangle$ and extension by linearity and continuity (which is possible since the map has operator norm one). Choose an orthonormal basis $\{e_k\} \subset \mathcal{H}_2$ an consider the ansatz

$$\mathrm{tr}_2[B] := \sum_k \big(\mathbb{1} \otimes \langle e_k|\big) B \big(\mathbb{1} \otimes |e_k\rangle\big). \tag{1.56}$$

According to the subsequent Lemma 1.39, the r.h.s. of this equation converges in trace-norm to a trace-class operator. Hence, tr_2 is well-defined and Eq.(1.54) can be verified by insertion. Uniqueness of the map is implied by the fact that specifying $\mathrm{tr}\,[XA]$ for all $A \in \mathcal{B}(\mathcal{H}_1)$ determines X. In particular, the construction in Eq.(1.56) is basis-independent.

The properties summarized in Eq.(1.55) follow immediately from Eq.(1.54). For instance, positivity of $\langle\psi, \mathrm{tr}_2[B]\psi\rangle = \mathrm{tr}\,[B(|\psi\rangle\langle\psi| \otimes \mathbb{1})]$ is implied by positivity of B together with $|\psi\rangle\langle\psi| \otimes \mathbb{1} \geq 0$ (cf. Exercise 1.7). □

Finally, we prove the missing Lemma that shows trace-norm convergence of the ansatz in Eq.(1.56). For later use, the formulation is slightly more general.

Lemma 1.39. *Let* $(A_k)_{k\in\mathbb{N}} \subset \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ *be a sequence of operators for which* $\lim_{n\to\infty} \sum_{k=1}^n A_k^* A_k = X \in \mathcal{B}(\mathcal{H}_1)$ *converges weakly. Then for every* $B \in \mathcal{B}_1(\mathcal{H}_1)$ *there is a* $B' \in \mathcal{B}_1(\mathcal{H}_2)$ *so that*

$$\left\| B' - \sum_{k=1}^n A_k B A_k^* \right\|_1 \to 0\,, \tag{1.57}$$

and $\mathrm{tr}\,[B'] = \mathrm{tr}\,[BX]$*. The map* $B \mapsto B'$ *is linear, it commutes with the adjoint map (i.e.,* $(B')^* = (B^*)'$*) and if* $B = B^*$ *then* $\|B'\|_1 \leq \|X\|_\infty \|B\|_1$*.*

Proof. By assumption, $\sum_{k\geq n} A_k^* A_k$ converges weakly to zero for $n \to \infty$. This implies weak-* convergence since we deal with a uniformly bounded subset of operators. W.l.o.g. we assume that $B \geq 0$ as we can always write it as a linear combination of four positive trace-class operators. Then

$$\left\| \sum_{k\geq n} A_k B A_k^* \right\|_1 = \mathrm{tr}\left[B \sum_{k\geq n} A_k^* A_k \right] \to 0.$$

This implies that $\sum_{k=1}^{n} A_k B A_k^*$ is a Cauchy sequence in $\mathcal{B}_1(\mathcal{H}_2)$ and thus convergent in trace-norm to some element B'. $\mathrm{tr}\,[B'] = \mathrm{tr}\,[BX]$ then follows from the cyclic properties of the trace together with dominated convergence (or Fubini-Tonelli).

Linearity of the map $B \mapsto B'$ follows from linearity of $B \mapsto A_k B A_k^*$ and the commutation with the adjoint map from $\left(A_k B A_k^*\right)^* = A_k B^* A_k^*$. Finally, assume that $B = B^*$ so that we can decompose $B = B_+ - B_-$ into orthogonal positive and negative parts. Then $\|B'\|_1 \leq \|(B_+)'\|_1 + \|(B_-)'\|_1 = \mathrm{tr}\,[(B_+ + B_-)X] \leq \|X\|_\infty \|B\|_1$. □

By interchanging the labels $1 \leftrightarrow 2$ and using the isomorphism $\mathcal{H}_1 \otimes \mathcal{H}_2 \simeq \mathcal{H}_2 \otimes \mathcal{H}_1$ we can define a partial trace $\mathrm{tr}_1 : \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}_1(\mathcal{H}_2)$ in complete analogy to Thm.1.38. The defining equation in this case would be $\mathrm{tr}\,[\mathrm{tr}_1[B]A] = \mathrm{tr}\,[B(\mathbb{1} \otimes A)]$ imposed for all $A \in \mathcal{B}(\mathcal{H}_2)$. More generally, for $\mathcal{H} := \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n$ we can define a partial trace for any non-empty subset $\Lambda \subseteq \{1, \ldots, n\}$, which then equals the composition of all individual partial traces, i.e. $\mathrm{tr}_\Lambda = \prod_{i \in \Lambda} \mathrm{tr}_i$.

Finally, a useful Lemma when dealing with partial traces:

Lemma 1.40. *For $X \in \mathcal{B}_2(\mathcal{H}_B \otimes \mathcal{H}_E, \mathcal{H}_A)$ and $Y \in \mathcal{B}_2(\mathcal{H}_D, \mathcal{H}_E \otimes \mathcal{H}_C)$ we have*

$$\|(X \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes Y)\|_\infty^2 \leq \|\mathrm{tr}_E[X^*X]\|_\infty \|\mathrm{tr}_E[YY^*]\|_\infty \tag{1.58}$$

Proof. We make implicit use of the Hilbert-Schmidt isomorphism of Thm.1.33. To this end, we fix orthonormal bases of all involved Hilbert spaces and define an operator $V \in \mathcal{B}_2(\mathcal{H}_B, \mathcal{H}_A \otimes \mathcal{H}_E)$ by expressing its matrix elements in terms of the ones of X as $\langle a, e|V|b\rangle := \langle a|X|b, e\rangle$. Similarly, we define $W : \mathcal{B}_2(\mathcal{H}_D \otimes \mathcal{H}_E, \mathcal{H}_C)$ via $\langle c|W|d, e\rangle := \langle e, c|Y|d\rangle$. This implies that $V^*V = \mathrm{tr}_E[X^*X]$, $WW^* = \mathrm{tr}_E[YY^*]$ and $(\mathbb{1}_A \otimes W)(V \otimes \mathbb{1}_D) = (X \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes Y)$. Hence,

$$\begin{aligned} \|(X \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes Y)\|_\infty^2 &= \|(\mathbb{1}_A \otimes W)(V \otimes \mathbb{1}_D)\|_\infty^2 \\ &\leq \|W\|_\infty^2 \|V\|_\infty^2 \\ &= \|WW^*\|_\infty \|V^*V\|_\infty = \|\mathrm{tr}_E[X^*X]\|_\infty \|\mathrm{tr}_E[YY^*]\|_\infty. \end{aligned}$$

□

Corollary 1.41. *Let $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B), \sigma_{BC} \in \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_C)$ be two positive definite density operators on finite-dimensional Hilbert spaces, and $\rho_A := \mathrm{tr}_B[\rho_{AB}]$, $\sigma_C := \mathrm{tr}_B[\sigma_{BC}]$. Then*

$$\rho_{AB} \otimes \sigma_C^{-1} \leq \rho_A \otimes \sigma_{BC}^{-1}, \qquad \textit{and} \tag{1.59}$$

$$\ln \rho_{AB} + \ln \sigma_{BC} \leq \ln \rho_A + \ln \sigma_C, \tag{1.60}$$

where we have omitted identity operators in the last line.[12]

Proof. We define $X := \rho_{AB}^{1/2}\left(\rho_A^{-1/2} \otimes \mathbb{1}_B\right)$ and $Y := \left(\mathbb{1}_B \otimes \sigma_C^{-1/2}\right)\sigma_{BC}^{1/2}$ to which we apply Lem.1.40 (carefully renaming Hilbert spaces—in particular $\mathcal{H}_B$ now plays

[12] That is, e.g. ρ_A should be understood as $\rho_A \otimes \mathbb{1}_B \otimes \mathbb{1}_C$.

the role of $\mathcal{H}_E$ in Lem.1.40). By construction, $\mathrm{tr}_B[X^*X] = \mathbb{1}_A$ and $\mathrm{tr}_B[YY^*] = \mathbb{1}_C$ so that the r.h.s. of Eq.(1.58) is equal to 1. Therefore

$$\begin{aligned} 1 &\geq \|(X \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes Y)\|_\infty^2 = \|(\mathbb{1}_A \otimes Y^*)(X^*X \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes Y)\|_\infty \\ &= \left\|\left(\rho_A^{-1/2} \otimes \sigma_{BC}^{1/2}\right)\left(\rho_{AB} \otimes \sigma_C^{-1}\right)\left(\rho_A^{-1/2} \otimes \sigma_{BC}^{1/2}\right)\right\|_\infty. \end{aligned} \tag{1.61}$$

As the operator inside the norm of Eq.(1.61) is positive, it has norm bounded by one iff it is itself bounded by the identity operator. From here Eq.(1.59) follows from rearranging terms. Finally, applying the logarithm to Eq.(1.59) and exploiting that ln is an operator monotone function (i.e., $\forall x, y \in \mathcal{B}(\mathcal{H})$: $x \geq y \geq 0 \Rightarrow \ln x \geq \ln y$) yields Eq.(1.60). □

Composite and reduced systems Within quantum theory, tensor products are used to describe composite systems. If a system is composed of distinguishable subsystems that are individually assigned to Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, then the description of the composite system is based on the tensor product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. Here 'distinguishable subsystems' might refer to spatially separated parts of a system or to different degrees of freedom of one system, such as the spin and the position of a single electron. In this case, one would describe the spin within $\mathbb{C}^2$ and the position within $L_2(\mathbb{R}^3)$. Hence $\mathbb{C}^2 \otimes L_2(\mathbb{R}^3)$ would be the Hilbert space underlying the description that covers both degrees of freedom. Aspects of a system that exclude each other, on the other hand, are reflected by a direct sum. Consider for instance a neutron n, which can decay into a proton p, an electron e^- and an electron-anti-neutrino $\bar{\nu}_e$, i.e. $n \to p + e^- + \bar{\nu}_e$. This would be modeled using $\mathcal{H}_n \oplus \mathcal{H}_p \otimes \mathcal{H}_{e^-} \otimes \mathcal{H}_{\bar{\nu}_e}$ as the overall Hilbert space since there is either the neutron *or* its decay products. However, if a composite system would consist out of a neutron *and* a proton, an electron and an electron anti-neutrino, then we would use $\mathcal{H}_n \otimes \mathcal{H}_p \otimes \mathcal{H}_{e^-} \otimes \mathcal{H}_{\bar{\nu}_e}$.

Suppose $\rho \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is a density operator that describes the preparation of a composite system composed of two subsystems. If we disregard say the second system and consider only the first part, the corresponding density operator is given by $\rho_1 := \mathrm{tr}_2[\rho]$. This is then called a *reduced density operator*. Similarly, if we discard the first subsystem, the reduced density operator that describes the remaining part is $\rho_2 := \mathrm{tr}_1[\rho]$. If ρ is a pure state, the reduced density operators can be read off its Schmidt decomposition:

Corollary 1.42. *Let $|\Psi\rangle\langle\Psi| \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$ be a pure density operator with Schmidt decomposition $|\Psi\rangle = \sum_{i=1}^r \sqrt{\lambda_i}|e_i\rangle \otimes |f_i\rangle$ with $r \in \mathbb{N} \cup \{\infty\}$. Then its reduced density operators are given by*

$$\rho_1 = \sum_{i=1}^r \lambda_i |e_i\rangle\langle e_i| \quad \textit{and} \quad \rho_2 = \sum_{i=1}^r \lambda_i |f_i\rangle\langle f_i|. \tag{1.62}$$

Proof. The statement follows from inserting the Schmidt decomposition into the explicit form of the partial trace in Eq.(1.56). The calculation simplifies if we use the basis of the Schmidt decomposition in the respective partial trace. □

Cor.1.42 leads to some simple but useful observations: the spectra of the two reduced density operators coincide as multisets and, more qualitatively, the rank of each reduced density operator equals the Schmidt rank. In particular, Ψ is a simple tensor product ($r = 1$) iff the reduced states are pure.

Another simple but useful observation is that the above corollary can be read in reverse, and we can (at least mathematically) regard every mixed state as the reduced state of some larger system that is described by a pure state:

Corollary 1.43 (Purification)**.** *Let $\rho_1 \in \mathcal{B}_1(\mathcal{H}_1)$ be a density operator of rank $r \in \mathbb{N} \cup \{\infty\}$. Then there is a Hilbert space $\mathcal{H}_2$ of dimension $\dim(\mathcal{H}_2) = r$ and a pure state $|\Psi\rangle\langle\Psi| \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$ so that $\rho_1 = \mathrm{tr}_2\big[|\Psi\rangle\langle\Psi|\big]$.*

Proof. We start with the spectral decomposition of ρ_1, which we interpret as the l.h.s. of Eq.(1.62), and construct a pure state Ψ via its Schmidt decomposition with Schmidt coefficients $\sqrt{\lambda_i}$ and the eigenvectors of ρ_1 as orthonormal family on the first tensor factor. Cor. 1.42 then guarantees that we recover ρ_1 as the partial trace of $|\Psi\rangle\langle\Psi|$. □

Clearly, such a *purification* is not unique. Any state vector of the form $(\mathbb{1} \otimes V)\Psi$ with V an isometry would also be a working purification.

Cor.1.43 and Cor.1.42 together imply that if two density operators ρ_1, ρ_2 have the same spectrum, then they can arise as the two partial traces of one pure state. The following is a quantitative extension of this fact and shows that states with similar spectra can arise as partial traces of a state that is almost pure:

Lemma 1.44 (Quantum coupling)**.** *For any pair of density operators $\rho_1, \rho_2 \in \mathcal{B}_1(\mathcal{H})$ there exists a density operator $\rho \in \mathcal{B}_1(\mathcal{H} \otimes \mathcal{H})$ s.t. $\mathrm{tr}_2[\rho] = \rho_1$, $\mathrm{tr}_1[\rho] = \rho_2$, and*

$$\|\rho\|_\infty \geq 1 - \frac{1}{2}\|\rho_1 - \rho_2\|_1. \tag{1.63}$$

Proof. Let λ, μ be the decreasingly ordered sequences of eigenvalues of ρ_1 and ρ_2, respectively. Their l_1-distance satisfies $\|\lambda - \mu\|_1 =: 2\epsilon \leq \|\rho_1 - \rho_2\|_1$ (cf. Eq.(1.22) in [29]). Denoting by $\{|\lambda_i\rangle\}$ and $\{|\mu_i\rangle\}$ the orthonormal bases of eigenvectors of ρ_1 and ρ_2, respectively, we define a vector $|\Psi\rangle := \sum_i \sqrt{\min\{\lambda_i, \mu_i\}}|\lambda_i\rangle \otimes |\mu_i\rangle$. This satisfies $\mathrm{tr}\,[|\Psi\rangle\langle\Psi|] = 1 - \epsilon$ and $\mathrm{tr}_1[|\Psi\rangle\langle\Psi|] \leq \rho_2$, $\mathrm{tr}_2[|\Psi\rangle\langle\Psi|] \leq \rho_1$. Therefore, we can define density operators σ_1, σ_2 by imposing

$$\rho_1 = \mathrm{tr}_2[|\Psi\rangle\langle\Psi|] + \epsilon\sigma_1, \qquad \rho_2 = \mathrm{tr}_1[|\Psi\rangle\langle\Psi|] + \epsilon\sigma_2.$$

The state $\rho := |\Psi\rangle\langle\Psi| + \epsilon\sigma_1 \otimes \sigma_2$ then fulfills the requirements of the theorem. □

Let us finally have a closer look at how the machinery of reduced and composite systems works on the side of the measurements. Suppose there are two independent measurement devices acting on the two parts of a composite system, individually described by POVMs M_1 and M_2. If $Y_1 \subseteq X_1$ and $Y_2 \subseteq X_2$ are corresponding measurable sets of measurement outcomes, then the overall measurement that now has outcomes in $X_1 \times X_2$, equipped with the product sigma-algebra, is described by a POVM that satisfies $M(Y_1 \times Y_2) = M_1(Y_1) \otimes M_2(Y_2)$.

Taking disjoint unions and complements (as in Lemma 1.9) this defines M on the entire product sigma-algebra. The marginal probabilities are then given by

$$\begin{aligned} p_1(Y_1) &= p(Y_1 \times X_2) = \mathrm{tr}\left[\rho M(Y_1 \times X_2)\right] = \mathrm{tr}\left[\rho\big(M_1(Y_1) \otimes M_2(X_2)\big)\right] \\ &= \mathrm{tr}\left[\rho\big(M_1(Y_1) \otimes \mathbb{1}\big)\right] = \mathrm{tr}\left[\rho_1 M_1(Y_1)\right], \end{aligned}$$

consistent with the definition and interpretation of the reduced density operator $\rho_1 = \mathrm{tr}_2[\rho]$.

If the overall state is described by a simple tensor product $\rho = \rho_1 \otimes \rho_2$, which is then called a *product state*, we obtain

$$\begin{aligned} p(Y_1 \times Y_2) &= \mathrm{tr}\left[(\rho_1 \otimes \rho_2)\big(M_1(Y_1) \otimes M_2(Y_2)\big)\right] = \mathrm{tr}\left[\rho_1 M_1(Y_1)\right] \mathrm{tr}\left[\rho_2 M_2(Y_2)\right] \\ &= p_1(Y_1)\, p_2(Y_2). \end{aligned}$$

This means that the measurement outcomes are independent. In other words, there are no correlations between the subsystems if the preparation is described by a product state.

Entropic quantities

Definition 1.45 (Relative entropy & mutual information)**.**

- *Let $\rho, \sigma \in \mathcal{B}_1(\mathcal{H})$ be positive. If $\ker(\rho) \supseteq \ker(\sigma)$, the* relative entropy *between ρ and σ is defined as $S(\rho\|\sigma) := \mathrm{tr}\left[\rho\big(\log(\rho) - \log(\sigma)\big)\right]$ where the trace is taken in an eigenbasis of ρ. If $\ker(\rho) \not\supseteq \ker(\sigma)$ then $S(\rho\|\sigma) := \infty$.*
- *Let $\rho_{AB} \in \mathcal{B}_1(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a density operator with reduced density operators $\rho_A := \mathrm{tr}_B[\rho_{AB}]$ and $\rho_B := \mathrm{tr}_A[\rho_{AB}]$. The* mutual information *between the subsystems A and B in ρ is defined as $I(A:B) := S(\rho\|\rho_A \otimes \rho_B)$.*

Remark: There are many ways of characterizing the relative entropy (and thus also the mutual information) as the derivative or limit of some quantity. Two of them, which can be easily verified, are the following: let $\rho_\lambda := \lambda\rho_0 + (1-\lambda)\rho_1$ be a convex combination of two density operators ρ_0, ρ_1. Then (taking the natural logarithm in Eq.(1.65))

$$S(\rho_0\|\rho_1) = \frac{d}{d\lambda}\bigg|_{\lambda=0} S(\rho_\lambda) \qquad \text{and} \tag{1.64}$$

$$S(\rho_0\|\rho_1) = \frac{d}{dt}\bigg|_{t=1} \mathrm{tr}\left[\rho_0^t\, \rho_1^{(1-t)}\right]. \tag{1.65}$$

A crucial property of both, the relative entropy and the mutual information, is positivity together with the fact that they are zero only in the obvious case:

Proposition 1.46 (Pinsker's inequality)**.** *The relative entropy and the mutual information as defined in Def.1.45 with* $\log = \ln$ *satisfy:*

$$S(\rho\|\sigma) \geq \frac{1}{2}\,\|\rho - \sigma\|_1^2\,, \tag{1.66}$$

$$I(A:B) \geq \frac{1}{2}\,\|\rho_{AB} - \rho_A \otimes \rho_B\|_1^2\,. \tag{1.67}$$

In particular, $S(\rho\|\sigma) = 0$ *and* $I(A:B) = 0$ *iff* $\rho = \sigma$ *and* $\rho_{AB} = \rho_A \otimes \rho_B$*, respectively.*

Proof. For ease of the argument, we are going to cheat a little bit and prove Eqs. (1.66,1.67) for $\|\cdot\|_2$ instead of for $\|\cdot\|_1$. Clearly, the trace-norm bound is the stronger result and we refer to [30] for its proof.

By definition of the mutual information, Eq.(1.67) is a consequence of Eq.(1.66). In order to arrive at Eq.(1.66) (resp. its analog with $\|\cdot\|_2$), we use the fact that $f(x) := x \ln x$ is strongly convex on $[0,1]$ with $f''(x) = 1/x \geq 1$. So we can apply Eq.(1.38) from which the result then follows instantly. □

Corollary 1.47 (Strong concavity of entropy)**.** *Let* $\rho_\lambda := \lambda\rho_1 + (1-\lambda)\rho_0$ *be a convex combination of density operators with* $\lambda \in [0,1]$*. Then, using* $\log = \ln$*,*

$$S(\rho_\lambda) - \lambda S(\rho_1) - (1-\lambda)S(\rho_0) \geq \frac{1}{2}\lambda(1-\lambda)\,\|\rho_1 - \rho_0\|_1^2\,. \tag{1.68}$$

Proof. The statement becomes an immediate consequence of Pinsker's inequality Eq.(1.66) when realizing that the definitions of the entropy and relative entropy lead to the identity

$$S(\rho_\lambda) - \lambda S(\rho_1) - (1-\lambda)S(\rho_0) = \lambda S(\rho_1\|\rho_\lambda) + (1-\lambda)S(\rho_0\|\rho_\lambda).$$

Applying Eq.(1.66) to the r.h.s. then leads to the lower bound by

$$\frac{1}{2}\lambda\,\|\rho_1 - \rho_\lambda\|_1^2 + \frac{1}{2}(1-\lambda)\,\|\rho_0 - \rho_\lambda\|_1^2 = \frac{1}{2}\lambda(1-\lambda)\,\|\rho_1 - \rho_0\|_1^2\,.$$

□

The mutual information is an information-theoretic tool for quantifying correlations between two systems. The following connects it to a different, operationally more direct way, of measuring correlations:

Corollary 1.48 (Mutual information vs. connected correlation function)**.** *Let* $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *be a density operator with reduced states* ρ_A *and* ρ_B*, and mutual information* $I(A:B)$ *(using* $\ln$*). If* $A \in \mathcal{B}(\mathcal{H})$ *and* $B \in \mathcal{B}(\mathcal{H}_B)$ *are Hermitian with operator norms not exceeding* 1 *and* $\langle A \otimes B\rangle := \operatorname{tr}[\rho_{AB} A \otimes B]$*,* $\langle A\rangle\langle B\rangle := \operatorname{tr}[\rho_A A]\operatorname{tr}[\rho_B B]$*, then*

$$I(A:B) \geq \frac{1}{2}\big(\langle A \otimes B\rangle - \langle A\rangle\langle B\rangle\big)^2. \tag{1.69}$$

Proof. Using Pinsker's inequality in the form of Eq.(1.67), we obtain the stated inequality when applying the fact that $\|X\|_1 = \max\{|\operatorname{tr}[XC]| \mid C = C^*, \|C\|_\infty \leq 1\}$ to $X = \rho_{AB} - \rho_A \otimes \rho_B$ with $A \otimes B$ instead of C. □

A simple but insightful relation involving the mutual information is a form of the *second law of thermodynamics*: building up correlations increases the overall entropy, when looking at subsystems:

Lemma 1.49 (2^{nd} law)**.** *Let $\rho'_{AB} := U(\rho_A \otimes \rho_B)U^*$ be a unitarily evolved product state with reduced density operators ρ'_A and ρ'_B and mutual information $I(A' : B')$. The entropy-changes $\Delta S_A := S(\rho'_A) - S(\rho_A)$, $\Delta S_B := S(\rho'_B) - S(\rho_B)$ satisfy:*

$$\Delta S_A + \Delta S_B = I(A' : B'). \tag{1.70}$$

Proof. Using additivity of the von Neumann entropy together with its unitary invariance, which implies that $S(\rho'_{AB}) = S(\rho_A \otimes \rho_B)$, we get

$$\Delta S_A + \Delta S_B = S(\rho'_A) + S(\rho'_B) - S(\rho_A \otimes \rho_B) = I(A' : B').$$

□

Theorem 1.50 (Weak monotonicity & strong subadditivity)**.** *The entropies of the reduced states of a density operator $\rho_{ABC} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ satisfy*

$$S(\rho_{AB}) + S(\rho_{BC}) \geq S(\rho_A) + S(\rho_B) \quad \text{'weak monotonicity'} \tag{1.71}$$
$$S(\rho_{AC}) + S(\rho_{BC}) \geq S(\rho_{ABC}) + S(\rho_C) \quad \text{'strong subadditivity'} \tag{1.72}$$

Proof. Assume for the moment that ρ_{ABC} is positive definite and acts on a finite-dimensional Hilbert space. Under this assumption, the *weak monotonicity* inequality of Eq.(1.71) follows from Eq.(1.60) by setting $\sigma_{BC} := \rho_{BC}$ and taking the trace of the resulting inequality after multiplying with ρ_{ABC} or, more precisely, left- and right-multiplying with $\rho_{ABC}^{1/2}$. By continuity of the entropy, the result extends to positive semidefinite states, and with the help of Lemma 1.30 it can be lifted to infinite-dimensional Hilbert spaces.

In order to obtain the *strong subadditivity* inequality of Eq.(1.72) we begin with weak monotonicity and consider a purification $\rho_{A'ABC}$ of ρ_{ABC} on $\mathcal{H}_{A'} \otimes \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. Since complementary reduced states of a pure state have the same non-zero spectrum (cf. Cor.1.42)), we can rewrite Eq.(1.71) as

$$S(\rho_{A'C}) + S(\rho_{BC}) \geq S(\rho_{A'BC}) + S(\rho_C).$$

Realizing that this inequality has to hold for arbitrary Hilbert spaces and states then completes the proof. □

We will discuss two immediate consequences of the strong subadditivity inequality. For the first one, let us introduce the *conditional entropy* $S(A|B)_\rho := S(\rho_{AB}) - S\rho_B$ of a density operator $\rho = \rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Unlike its classical counterpart, $S(A|B)_\rho$ can be positive or negative. However, it turns out to be a concave quantity:

Corollary 1.51 (Concavity of conditional entropy)**.** *Let $\rho := \sum_k \lambda_k \rho_k$ be a convex combination of density operators $\rho_k \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then*

$$S(A|B)_\rho \geq \sum_k \lambda_k S(A|B)_{\rho_k}. \tag{1.73}$$

Proof. Define $\rho_{ABC} := \sum_k \lambda_k \rho_k \otimes |k\rangle\langle k| \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, with $\{|k\rangle\}$ an orthonormal basis of $\mathcal{H}_C$. Using the strong subadditivity inequality of Eq.(1.72) (with labels $B \leftrightarrow C$ interchanged) together with the case of equality in the right inequality of Eq.(1.40), we obtain

$$\begin{aligned} S(A|B)_\rho &= S(\rho_{AB}) - S(\rho_B) \;\geq\; S(\rho_{ABC}) - S(\rho_{BC}) \\ &= \sum_k \lambda_k S(A|B)_{\rho_k}. \end{aligned}$$

□

The second consequence is the *monotonicity of the relative entropy*:

Corollary 1.52 (Monotonicity of relative entropy)**.** *Let $\rho_{AB}, \sigma_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be two density operators. Then*

$$S(\rho_{AB}\|\sigma_{AB}) \;\geq\; S(\rho_B\|\sigma_B). \tag{1.74}$$

Proof. On the positive cone generated by the density operators on $\mathcal{H}_A \otimes \mathcal{H}_B$ we define the functional $F(\rho) := S(A|B)_\rho$. F satisfies $F(tx) = tF(x)$ for all $t \in \mathbb{R}_+$ and, according to Cor.1.51, is a concave function. Taken together, these facts imply that $F(x+ty) \geq F(x) + tF(y)$, so that in particular $\lim_{t\searrow 0}[F(x+ty) - F(x)]/t \geq F(y)$. Now we apply this inequality with $x = \sigma_{AB}$ and $y = \rho_{AB}$ and obtain

$$\begin{aligned} \lim_{t\searrow 0} \frac{F(\sigma_{AB} - t\rho_{AB}) - F(\sigma_{AB})}{t} &= \operatorname{tr}\left[\rho_B \log \sigma_B\right] - \operatorname{tr}\left[\rho_{AB} \log \sigma_{AB}\right] \\ \geq\; F(\rho_{AB}) &= \operatorname{tr}\left[\rho_B \log \rho_B\right] - \operatorname{tr}\left[\rho_{AB} \log \rho_{AB}\right]. \end{aligned}$$

Rearranging terms then leads to Eq.(1.74). □

Exercise 1.26. For $i \in \{1,2\}$ consider $A_i \in \mathcal{B}(\mathcal{H}_i)$. Show that if A_1, A_2 are positive or unitary then the same holds true for $A_1 \otimes A_2$.

Exercise 1.27 (Flip). Let $\mathcal{H}_1 \simeq \mathcal{H}_2 \simeq \mathbb{C}^d$. By identifying bases of the two spaces we can define a *flip operator* $\mathbb{F} \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ via $\mathbb{F}(\varphi \otimes \psi) = \psi \otimes \varphi$.

(a) Determine the eigenvalues and eigenvectors of $\mathbb{F}$.

(b) Prove that $\mathbb{F}$ is the unique operator satisfying $\operatorname{tr}[\mathbb{F}(A \otimes B)] = \operatorname{tr}[AB]$ $\forall A, B \in \mathcal{B}(\mathbb{C}^d)$.

(c) Let $(G_i)_{i=1}^{d^2} \subset \mathcal{B}(\mathbb{C}^d)$ be any Hilbert-Schmidt-orthonormal basis of Hermitian operators. Show that $\mathbb{F} = \sum_{i=1}^{d^2} G_i \otimes G_i$.

Exercise 1.28 (Partial trace). Consider an element of $\mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^n)$ in block matrix representation. How can the partial traces be understood in this picture?

Exercise 1.29 (Monogamy). Alice, Bob and Charlie share a quantum system described by a density operator $\rho \in \mathcal{B}_1(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ where $\mathcal{H}_B \simeq \mathcal{H}_C$. Suppose the reduced density operator ρ_{AB} is pure. Show that $\rho_{AC} = \rho_{AB}$ is not possible unless both are simple products (i.e. their Schmidt rank is one).

Exercise 1.30 (Entropy inequalities). Let $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a density operator with partial traces ρ_A and ρ_B.

(a) The Shannon entropy for classical discrete probability distributions satisfies the monotonicity inequality $S_{AB} \geq \max\{S_A, S_B\}$. Show that the naive quantum analogue fails, i.e., that there exist states for which $S(\rho_{AB}) \ngeq \max\{S(\rho_A), S(\rho_B)\}$.

(b) Provide three proofs for the *subadditivity inequality* $S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$: an elementary one, one based on mutual information and one that uses Thm.1.50. Show that equality holds iff $\rho_{AB} = \rho_A \otimes \rho_B$.

(c) Prove the *Araki-Lieb triangle inequality* $|S(\rho_A) - S(\rho_B)| \leq S(\rho_{AB})$. (Hint: use (b) together with purification.)

Notes and literature Lemma 1.44 appeared in [31]. It is inspired by the vast applications of *couplings* of classical random variables.
Cor.1.41 and the idea for Lemma 1.40 are from [32].
Strong subadditivity of the von Neumann entropy as well as monotonicity of the relative entropy and concavity of the conditional von Neumann entropy were proven by Lieb and Ruskai [27]. Carlen and Lieb observed in [33] that strong subadditivity implies the following improved version of itself:

$$S_{AC} + S_{BC} - S_{ABC} - S_C \geq 2\max\{0, S_A - S_{AB}, S_B - S_{AB}\}, \tag{1.75}$$

where $S_{AB} := S(\rho_{AB})$, etc. . The factor 2 is thereby best possible.
Different remainder terms in entropic inequalities can be obtained from variants of Pinsker's inequality [34]. In particular, it holds that

$$S(\rho\|\sigma) \geq -2\log \operatorname{tr}\left[\sqrt{\rho}\sqrt{\sigma}\right], \tag{1.76}$$

which is neither generally stronger nor weaker than Pinsker's inequality in Eq.(1.66).
The subadditivity inequality (cf. exercise 1.30) is a crucial ingredient in axiomatic characterizations of entropy. For instance, suppose S is a functional from the set of all density operators into the reals such that the following hold:

(i) *Subadditivity.* For all states on a tensor product space: $S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$.

(ii) *Additivity.* $S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$.

(iii) *Expansibility.* $S(\rho \oplus 0) = S(\rho)$.

(iv) *Symmetry.* $S(\rho) = S(U\rho U^*)$ for all unitaries U.

It follows from [35] that (i)-(iv) imply that there are non-negative constants a, b such that for all density operators ρ:

$$S(\rho) = aS_0(\rho) + bS_1(\rho), \tag{1.77}$$

where $S_0(\rho) := \log \operatorname{rank}(\rho)$ is the *Hartley entropy* and S_1 is the von Neumann entropy. Conversely, every functional of the form in Eq.(1.77) has these properties and the von Neumann entropy gets singled out if we impose continuity (which implies $a = 0$) and a suitable normalization (s.t. $b = 1$).

1.6 Tensor-powers

Theorem 1.53 (Asymptotic equipartition property (AEP)). *For $\mathcal{H} \simeq \mathbb{C}^d$, let $\rho \in \mathcal{B}(\mathcal{H})$ be a density operator, $S(\rho)$ its entropy, and* $\operatorname{Var}[\log(1/\rho))] := \operatorname{tr}\left[\rho(\log\rho)^2\right] - S(\rho)^2$ *(all w.r.t. $\log_2$). For every $\epsilon > 0$ and $n \in \mathbb{N}$ there is a subspace $V \subseteq \mathcal{H}^{\otimes n}$ spanned by eigenvectors of $\rho^{\otimes n}$ s.t. if Q is the projection*

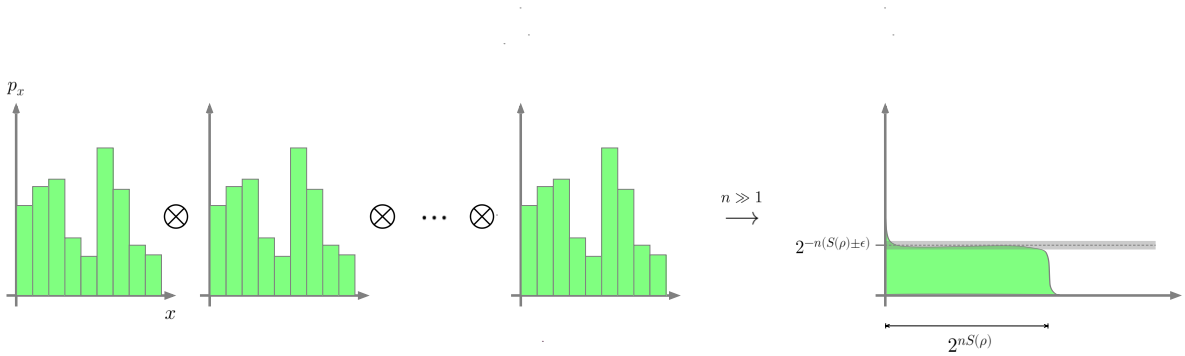

Figure 1.1: *Asymptotic equipartition property.* A tensor power $\rho^{\otimes n}$ of a finite-dimensional density matrix ρ corresponds to a product of n identical and independent probability distributions. For sufficiently large n, the product distribution is essentially supported on a space of size $2^{nS(\rho)}$ on which the probabilities/eigenvalues are (very roughly) equal.

onto V *and* $\sigma := Q\rho^{\otimes n}Q$*, then*

$$\|\rho^{\otimes n} - \sigma\|_1 \;<\; \frac{\mathrm{Var}\left[\log(1/\rho)\right]}{\epsilon^2 n}, \tag{1.78}$$

$$\dim(V) \;=\; \mathrm{rank}(\sigma) \;\leq\; 2^{n(S(\rho)+\epsilon)}, \tag{1.79}$$

$$\mathrm{spec}(\sigma)\setminus\{0\} \;\subset\; \left[2^{-n(S(\rho)+\epsilon)}, 2^{-n(S(\rho)-\epsilon)}\right]. \tag{1.80}$$

Remark: From the fact that $a(\log a)^2 \leq -\log(e^2)/(e^2) =: c \simeq 1.127$ for $a \in [0,1]$, we get the rough bound $\mathrm{Var}\left[\log(1/\rho)\right] \leq cd$.

Proof. (sketch): Expressing everything in the eigenbasis of $\rho^{\otimes n}$, we realize that the statement is really one about the classical discrete probability distribution that is given by the eigenvalues of $\rho^{\otimes n}$. The main idea is to apply the law of large numbers to the random variable $X(x) := \frac{1}{n}\log(1/p(x))$, $x \in \{1,\ldots,d\}^n$ with distribution $p(x) := \prod_{i=1}^n p_{x_i}$, where $p_1,\ldots,p_d$ are the eigenvalues of ρ. This means that each x labels an element of the orthonormal basis of eigenvectors of $\rho^{\otimes n}$ and $p(x)$ is the corresponding eigenvalue. The random variable X is an average of n i.i.d. random variables each of which has mean $S(\rho)$ and variance $\mathrm{Var}\left[\log(1/\rho)\right]$. By the weak law of large numbers (in the version quantified by Chebyshev's inequality), the probability for X to deviate by more than ϵ from its mean can be bounded by:

$$\mathrm{Prob}\left[\left|X - S(\rho)\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}\left[\log(1/\rho)\right]}{\epsilon^2 n}. \tag{1.81}$$

If we define the subspace V as the one spanned by all eigenvectors whose label x is such that $|X(x) - S(\rho)| < \epsilon$, then Eq.(1.80) holds by construction

and Eq.(1.78) becomes a reformulation of Eq.(1.81). Finally, Eq.(1.79) follows from the fact, that a sum of eigenvalues of a density operator is bounded by one. Hence, since all $\dim(V)$ non-zero eigenvalues of σ are at least $2^{-n(S(\rho)+\epsilon)}$, Eq.(1.79) follows. □

Notes and literature The asymptotic equipartition theorem is the core of Shannon's source coding theorem [36] and of many other results in information theory. There exist various extensions to beyond i.i.d. sources, often under the name *Shannon-McMillan-Breiman theorem*.

1.7 Quantum channels and operations

So far, we have introduced and discussed aspects of preparation and measurement. In this section, we will analyze the mathematical objects that are used to describe anything that could happen to a quantum system between preparation and measurement. This could mean active operations performed by an experimentalist, interactions either between parts of the system or with an environment or plain time evolution.

Since quantum theory divides the description of every statistical experiment into preparation and measurement, there are two natural ways to describe intermediate operations or evolutions: either by incorporating them into the preparation or into the measurement description. These two viewpoints are called *Schrödinger picture* and *Heisenberg picture*, respectively. While the Schrödinger picture updates the density operator, the Heisenberg picture updates the POVM.

Schrödinger & Heisenberg picture The mathematical maps that are to describe the evolution/operation in either Schrödinger or Heisenberg picture have to be consistent with the probabilistic interpretation. In particular, they have to preserve convex combinations, which implies that they have to be affine maps. These, however, can always be extended to linear maps: for instance, the affine map $\rho \mapsto \rho' = L(\rho) + C$, where L is a linear map and C a constant, has a linear extension from the trace-one-hyperplane to the entire space of trace-class operators that is obtained by simply replacing C with $C\mathrm{tr}\,[\rho]$. In this way, we can without loss of generality restrict ourselves to linear maps. Elementary properties of such maps are introduced in the following:

Definition 1.54. *Let $\mathcal{L} \subseteq \mathcal{B}(\mathcal{H}_1)$ be a linear subspace. A linear map $T : \mathcal{L} \to \mathcal{B}(\mathcal{H}_2)$ is called*

- trace-preserving *if the image of any $A \in \mathcal{L} \cap \mathcal{B}_1(\mathcal{H}_1)$ under T is trace-class and* $\mathrm{tr}\,[T(A)] = \mathrm{tr}\,[A]$,
- unital *if $T(\mathbb{1}) = \mathbb{1}$ (assuming $\mathbb{1} \in \mathcal{L}$),*
- positive *if $T(A) \geq 0$ for all positive $A \in \mathcal{L}$,*
- completely positive *if $T \otimes \mathrm{id}_n$ is positive for all $n \in \mathbb{N}$, where id_n is the identity map on $\mathcal{B}(\mathbb{C}^n)$.*

Remark: here we have tacitly introduced a third-level tensor product, namely the tensor product of linear maps on spaces of operators. $T \otimes \mathrm{id}_n$ is defined as $T \otimes \mathrm{id}_n : A \otimes B \mapsto T(A) \otimes B$ and linear extension to finite linear combinations.

Let us see how these properties come into play. If $T : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ is a trace-preserving and positive linear map, then $T(\rho)$ is a density operator whenever ρ is one. Recalling that ρ might describe a part of a larger system whose other parts are left untouched by T, it is necessary to impose that not only T maps density operators to density operators but $(T \otimes \mathrm{id})$ does so as

well. This is captured by the notion of complete positivity. In principle, this should hold not only for a finite-dimensional 'innocent bystander'. We will see later though, from the representation theory of completely positive maps, that considering finite-dimensional systems is sufficient in this context.

Example 1.18 (Transposition)**.** The paradigm of a map that is positive but not completely positive is matrix transposition. Let $\Theta : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}), \Theta(A) := A^T$ be the transposition map w.r.t. a fixed basis $\{|k\rangle\} \subset \mathcal{H}$. This is a positive map since it preserves Hermiticity as well as the spectrum. However, for $|\psi\rangle = |00\rangle + |11\rangle \in \mathcal{H} \otimes \mathbb{C}^2$ we get $(\Theta \otimes \mathrm{id}_2)\big(|\psi\rangle\langle\psi|\big) = \sum_{i,j=0}^{1} \Theta\big(|i\rangle\langle j|\big) \otimes |i\rangle\langle j| = \sum_{i,j=0}^{1} |j\rangle\langle i| \otimes |i\rangle\langle j|$, for which -1 is an element of the spectrum (cf. Exercise 1.27).

Let us turn to the Heisenberg picture. Assume that $T^* : \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_1)$ is a continuous, unital and positive linear map.[13] If $M : \mathbb{B} \to \mathcal{B}(\mathcal{H}_2)$ is a POVM, then $M' := T^* \circ M : \mathbb{B} \to \mathcal{B}(\mathcal{H}_1)$ is a POVM as well. To see this, note that positivity of T^* implies positivity of $M'(Y)$ for all $Y \in \mathbb{B}$ and if $X = \cup_k X_k$ is countable disjoint partition of the set X of all possible outcomes into measurable subsets X_k, then

$$\sum_k M'(X_k) = T^*\left(\sum_k M(X_k)\right) = T^*(\mathbb{1}) = \mathbb{1},$$

where we used continuity of T^* in the first step and unitality in the last step.

Since Schrödinger picture and Heisenberg picture describe the same thing from different viewpoints, they should lead to consistent predictions. As the predictions are in the end probabilities expressed through Born's rule, the equivalence of the two viewpoints should be expressible on this level. This equivalence is established in the following theorem. For any map T in the Schrödinger picture it proves the existence of an equivalent description via a map T^* in the Heisenberg picture. We will comment on the more subtle converse direction below.

Theorem 1.55 (Schrödinger picture to Heisenberg picture)**.** *Let $T : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ be a bounded linear map. Then there is a unique linear map $T^* : \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_1)$ (called the* dual map*) that satisfies $\forall A \in \mathcal{B}(\mathcal{H}_2), \rho \in \mathcal{B}_1(\mathcal{H}_1)$:*

$$\mathrm{tr}\left[T(\rho)A\right] = \mathrm{tr}\left[\rho\, T^*(A)\right]. \tag{1.82}$$

Moreover, the following equivalences hold:

(i) T is positive iff T^* is positive,

(ii) T is completely positive iff T^* is completely positive,

(iii) T is trace-preserving iff T^* is unital.

[13] The meaning of the '*' will become clear below. For now, read 'T^*' just as an arbitrary symbol that we assign as a name to the map.

Proof. Consider the map $f : \mathcal{B}_1(\mathcal{H}_1) \to \mathbb{C}$ defined by $f(B) := \operatorname{tr}[T(B)A]$ for fixed $A \in \mathcal{B}(\mathcal{H}_2)$. Due to the linearity of T, f is linear. It is also bounded since Hölder's inequality and boundedness of T lead to $|f(B)| \leq \|T(B)\|_1 \|A\|_\infty \leq c\|B\|_1$ for some constant $c < \infty$. Hence f is a continuous linear functional on $\mathcal{B}_1(\mathcal{H}_1)$. The duality $\mathcal{B}_1(\mathcal{H}_1)' = \mathcal{B}(\mathcal{H}_1)$ then implies the existence of a $T^*(A) \in \mathcal{B}(\mathcal{H}_1)$ so that $f(B) = \operatorname{tr}[BT^*(A)]$, which verifies Eq.(1.82). As the l.h.s. of Eq.(1.82) depends linearly on A, $T^*(A)$ has to depend linearly on A as well so that T^* is a linear map. Uniqueness is guaranteed by the fact that specifying $\operatorname{tr}[\rho T^*(A)]$ for all density operators ρ determines the operator $T^*(A)$.

As for positivity, we use the defining relation between T and T^* in the form

$$\operatorname{tr}[T(|\psi\rangle\langle\psi|)A] = \langle\psi, T^*(A)\psi\rangle. \tag{1.83}$$

Imposing positivity of the l.h.s. for all $\psi \in \mathcal{H}_1$ and all positive $A \in \mathcal{B}(\mathcal{H}_2)$ is equivalent to positivity of T. Imposing the same for the r.h.s. is equivalent to positivity of T^*. So these conditions are equivalent. The same argument applies to complete positivity by replacing T with $T \otimes \mathrm{id}_n$ and realizing that $(T \otimes \mathrm{id}_n)^* = T^* \otimes \mathrm{id}_n$.

Similarly, from Eq.(1.82) we derive the equation

$$\operatorname{tr}[T(B) - B] = \operatorname{tr}\left[B\big(T^*(\mathbb{1}) - \mathbb{1}\big)\right]. \tag{1.84}$$

Here the l.h.s. is zero for all $B \in \mathcal{B}_1(\mathcal{H}_1)$ iff T is trace-preserving, whereas the r.h.s. is zero for all $B \in \mathcal{B}_1(\mathcal{H}_1)$ iff T^* is unital. □

One important property of the dual map has been left aside and will be covered in the following corollary: continuity. Before proving this in a quantitative way, some remarks on the involved norms are in order.

Both T and T^* are maps between Banach spaces. If not specified otherwise, their norms are the corresponding Banach space operator norms. That is, $\|T\| = \sup\{\|T(B)\|_1 \,|\, \|B\|_1 \leq 1\}$ and $\|T^*\| = \sup\{\|T^*(A)\|_\infty \,|\, \|A\|_\infty \leq 1\}$. The involved trace-norm and the operator norm in $\mathcal{B}(\mathcal{H})$ are dual to each other in the sense that

$$\|B\|_1 = \sup_{\|A\|_\infty = 1} \big|\operatorname{tr}[AB]\big|, \quad \text{and} \quad \|A\|_\infty = \sup_{\|B\|_1 = 1} \big|\operatorname{tr}[AB]\big|. \tag{1.85}$$

These equations can for instance be proven by means of the polar decomposition and the Schmidt decomposition, respectively.

Corollary 1.56. *Let $T : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ be a bounded linear map and T^* the corresponding dual map. Then $\|T^*\| = \|T\|$. Moreover, if T is positive, these norms are equal to $\|T^*(\mathbb{1})\|_\infty$. In particular, if T is positive and trace-preserving, then for all $B \in \mathcal{B}_1(\mathcal{H}_1), A \in \mathcal{B}(\mathcal{H}_2)$:*

$$\|T(B)\|_1 \leq \|B\|_1 \quad \textit{and} \quad \|T^*(A)\|_\infty \leq \|A\|_\infty. \tag{1.86}$$

Proof. Using the defining relation between T and T^* and Eq.(1.85) we obtain

$$\|T^*\| = \sup_{\|A\|_\infty = 1} \sup_{\|B\|_1 = 1} \Big|\underbrace{\operatorname{tr}[BT^*(A)]}_{=\operatorname{tr}[T(B)A]}\Big| = \|T\|. \tag{1.87}$$

To proceed, we exploit the convex structure of the unit balls in $\mathcal{B}_1(\mathcal{H}_1)$ and $\mathcal{B}(\mathcal{H}_2)$ by which it suffices to take the suprema over all rank-one elements in the trace-class and all unitaries in $\mathcal{B}(\mathcal{H}_2)$. The latter is justified by the Russo-Dye theorem (Thm.1.17) and the former by the Schmidt-decomposition (Eq.(1.8)). Thus

$$\|T^*\| = \sup_U \sup_{\psi,\varphi} \big|\langle\varphi, T^*(U)\psi\rangle\big|, \tag{1.88}$$

where the suprema are taken over all unitaries $U \in \mathcal{B}(\mathcal{H}_2)$ and unit vectors $\varphi, \psi \in \mathcal{H}_1$. Let us for the moment assume that $\mathcal{H}_2$ is finite-dimensional. This enables a spectral decomposition of the form $U = \sum_k \exp[i\alpha_k]|e_k\rangle\langle e_k|$ with $\alpha_k \in \mathbb{R}$ and $\{e_k\} =: E \subset \mathcal{H}_2$ an orthonormal basis. Inserting this into Eq.(1.88) leads to

$$\begin{aligned} \|T^*\| &\leq \sup_E \sup_{\psi,\varphi} \sum_k \big|\langle\varphi, T^*\big(|e_k\rangle\langle e_k|\big)\psi\rangle\big|, &(1.89)\\ &= \sup_E \sup_{\psi} \sum_k \langle\psi, T^*\big(|e_k\rangle\langle e_k|\big)\psi\rangle = \|T^*(\mathbb{1})\|_\infty . &(1.90) \end{aligned}$$

Here, in the step from the first to the second line we have used positivity of T^* together with two applications of Cauchy-Schwarz. Note that equality has to hold in the inequality since $U = \mathbb{1}$ was a valid choice in the first place. Eq.(1.86) then follows from unitality of T^*, which for positive maps now implies $\|T\| = \|T^*\| = 1$.

Finally, we have to come back to the assumption $\dim(\mathcal{H}_2) < \infty$. Suppose this is not the case. Then note that the core expression in Eq.(1.88) can also be written as $\mathrm{tr}\big[UT\big(|\psi\rangle\langle\varphi|\big)\big]$. Since $T\big(|\psi\rangle\langle\varphi|\big)$ is a trace-class operator on $\mathcal{H}_2$ it can be approximated arbitrarily well in trace-norm by a finite rank operator F. So we may restrict ourselves to unitaries that act non-trivial only on the finite-dimensional subspace $\mathrm{supp}(F)+\mathrm{ran}(F)$ and continue with the finite-dimensional argument. □

Thm.1.55 constructs a map in the Heisenberg picture for any map in the Schrödinger picture. What about the converse? In finite dimensions the situation is symmetric. There we can interpret the expression in Born's rule as Hilbert-Schmidt inner product w.r.t. which T^* is the adjoint operator corresponding to T. In infinite dimensions, the proof of Thm.1.55 relied on the duality relation $\mathcal{B}_1(\mathcal{H}_1)' = \mathcal{B}(\mathcal{H}_1)$, which does not hold in the other direction. In other words, there are maps $\Phi : \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_1)$ in the Heisenberg picture that have no predual that maps density operators to density operators. A map Φ is called *normal* if there exists such a predual. Equivalently, Φ is normal if it is continuous as a map from $\mathcal{B}(\mathcal{H}_2)$ to $\mathcal{B}(\mathcal{H}_1)$ when both spaces are equipped with the weak-* topology.

Kraus representation and environment We already know three elementary classes of linear maps that are completely positive and trace-preserving:

(i) Addition of an ancillary density operator σ via $B \mapsto B \otimes \sigma$.

(ii) Partial trace $B \mapsto \mathrm{tr}_2[B]$ in a composite system.

(iii) Unitary evolution of the form $B \mapsto UBU^*$, where U is a unitary.

Since complete positivity as well as the trace-preserving property is preserved under composition of maps, any composition of the three elementary building blocks is again completely positive and trace-preserving. In fact, we will see later that this construction is exhaustive.

Theorem 1.57 (Kraus/environment representation)**.** *For $T : \mathcal{B}_1(\mathcal{H}) \to \mathcal{B}_1(\mathcal{H})$ the following are equivalent:*

(1) There is a Hilbert space $\mathcal{K}$, a unitary $U \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ and a density operator $\sigma \in \mathcal{B}(\mathcal{K})$ s.t.

$$T(\rho) = \mathrm{tr}_{\mathcal{K}}\big[U(\rho \otimes \sigma)U^*\big], \tag{1.91}$$

(2) There is a Hilbert space $\mathcal{K}$, a unitary $W \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ and a unit vector $\psi \in \mathcal{K}$ s.t.

$$T(\rho) = \mathrm{tr}_{\mathcal{K}}\big[W(\rho \otimes |\psi\rangle\langle\psi|)W^*\big], \tag{1.92}$$

(3) There is a Hilbert space $\mathcal{K}$ and an isometry $V : \mathcal{H} \to \mathcal{H} \otimes \mathcal{K}$ s.t.

$$T(\rho) = \mathrm{tr}_{\mathcal{K}}\big[V\rho V^*\big], \tag{1.93}$$

(4) There is a finite or infinite sequence $(A_k)_{k=1}^r \subset \mathcal{B}(\mathcal{H})$, $r \in \mathbb{N} \cup \{\infty\}$ for which $\sum_{k=1}^r A_k^* A_k = \mathbb{1}$ converges weakly and

$$T(\rho) = \sum_{k=1}^{r} A_k \rho A_k^*. \tag{1.94}$$

Remark: The A_k's are called *Kraus-operators* and Eq.(1.94) the *Kraus representation* of T. As we have seen in Lemma 1.39, weak convergence of $\sum_k A_k^* A_k$ to a bounded operator(which in this case is equivalent to strong convergence) implies trace-norm convergence in Eq.(1.94).

Proof. To distinguish the auxiliary Hilbert spaces of the first three points, we denote them by $\mathcal{K}_1, \mathcal{K}_2$ and $\mathcal{K}_3$. We will show (1)⇔(2)⇒(4)⇒(3)⇒(2).

Assume (1) holds. Then we can use a purification $\psi \in \mathcal{K}_1 \otimes \mathcal{K}_1 := \mathcal{K}_2$ of $\sigma \in \mathcal{B}_1(\mathcal{K}_1)$, as derived in Cor.1.43, and we obtain (2) by choosing $W = U \otimes \mathbb{1}$. Conversely, (2)⇒(1) since (2) is a special case of (1).

Now suppose (2) holds. In order to show that (2)⇒(4), we set $A_k := \big(\mathbb{1} \otimes \langle e_k|\big)W\big(\mathbb{1} \otimes |\psi\rangle\big)$ for an orthonormal basis $\{e_k\} \subset \mathcal{K}_2$. Using the explicit construction of the partial trace in Eq.(1.56), we see that Eq.(1.94), after insertion of the A_k's, becomes Eq.(1.92). Strong convergence of $\sum_k |e_k\rangle\langle e_k| = \mathbb{1}$ then implies strong convergence in

$$\sum_k A_k^* A_k = \sum_k \big(\langle\psi| \otimes \mathbb{1}\big)W^*\big(\mathbb{1} \otimes |e_k\rangle\langle e_k|\big)W\big(|\psi\rangle \otimes \mathbb{1}\big) = \big(\langle\psi| \otimes \mathbb{1}\big)\underbrace{W^*W}_{=\mathbb{1}}\big(|\psi\rangle \otimes \mathbb{1}\big).$$

If (4) holds, then we can construct an isometry $V : \mathcal{H} \to \mathcal{H} \otimes \mathcal{K}_3$ with $\mathcal{K}_3 = l_2(\mathbb{N})$ if $r = \infty$ or otherwise $\mathcal{K}_3 = \mathbb{C}^r$ via $V : \varphi \mapsto \sum_k (A_k\varphi) \otimes e_k$ where $\{e_k\} \subset \mathcal{K}_3$ is any orthonormal basis. This is indeed an isometry, since

$$\langle \varphi, V^*V\varphi\rangle = \langle \varphi, \sum_k A_k^* A_k\, \varphi\rangle = \langle \varphi, \varphi\rangle.$$

Finally, assuming (3), we want extend the isometry V to a unitary in order to arrive at (2). To this end, take any unit vector $\psi \in \mathcal{K}_2 := \mathcal{K}_3$ and suppose the spaces $\mathcal{H} \otimes \big(\mathcal{K}_2 \ominus \mathbb{C}\psi\big) \simeq (\mathrm{ran}V)^\perp$ are isomorphic, which is certainly true if $\mathcal{H}$ has finite dimensions. Then there is a unitary $V' : \mathcal{H} \otimes \big(\mathcal{K}_2 \ominus \mathbb{C}\psi\big) \to (\mathrm{ran}V)^\perp$, which extends $V : \mathcal{H} \simeq \mathcal{H}\otimes\mathbb{C}\psi \to \mathcal{H}\otimes\mathcal{K}_2$ to a unitary $W := V \oplus V'$. If $(\mathrm{ran}V)^\perp$ is too small so that the assumed isomorphism does not hold, we first compose V with a canonical embedding of $\mathcal{K}_3$ into $\mathcal{K}_3 \oplus \mathbb{C} =: \mathcal{K}_2$. Then $(\mathrm{ran}V)^\perp$ with the orthogonal complement taken in $\mathcal{H} \otimes \mathcal{K}_2$ is infinite-dimensional and the desired isomorphism holds. □

Eq.(1.91) has a simple physical interpretation: we may think of T as describing an interaction, which is characterized by U, with an *environment* that is initially uncorrelated with the systems, described by a density operator σ and traced out after the interaction.

The Kraus representation of a completely positive linear map is not unique. This is, in fact, closely related to the non-uniqueness of the convex decomposition of a density operator into rank-one projections (cf. Example 1.11) and, in a similar vein, one can show the following:

Proposition 1.58 (Ambiguity in the Kraus representation)**.** *Let $T : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ have a Kraus representation of the form $T(\rho) = \sum_{i\in N} K_i\rho K_i^*$ with $N \subseteq \mathbb{N}$. If u_{ij} are the entries of a unitary matrix with index set $N \ni i, j$, then $B_i := \sum_{j\in N} u_{ij}K_j$ defines a set of Kraus operators that represent the same map via $T(\rho) = \sum_{i\in N} B_i\rho B_i^*$.*

Conversely, if $\{A_i\}_{i\in N}$ and $\{B_i\}_{i\in N}$ are two sets of Kraus-operators that represent the same trace-preserving map and if either N is finite or both sets contain an infinite number of zeros, then there is a unitary u s.t. $B_i := \sum_{j\in N} u_{ij}A_j$.

Definition 1.59 (Quantum channels)**.** *A linear map $T : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ is called a* quantum channel *if it is trace-preserving and completely positive.*

We will see later that every quantum channel can be represented in the ways specified by Thm.1.57.

Example 1.19 (Phase damping channel)**.** Let $\{|0\rangle, |1\rangle\}$ denote an orthonormal basis of $\mathbb{C}^2$ and define $\rho_{ij} := \langle i|\rho|j\rangle$. A simple model of a 'decoherence process' is given by the *phase damping channel* that is parametrized by $\lambda \in [0, 1]$ and

can be represented in the following ways:

$$\rho \quad \mapsto \quad \begin{pmatrix} \rho_{00} & (1-\lambda)\rho_{01} \\ (1-\lambda)\rho_{10} & \rho_{11} \end{pmatrix} = \sum_{k=1}^{3} A_k \rho A_k^* \tag{1.95}$$
$$\text{with} \quad A_1 := \sqrt{1-\lambda}\,\mathbb{1}, \quad A_2 := \sqrt{\lambda}\,|0\rangle\langle 0|, \quad A_2 := \sqrt{\lambda}\,|1\rangle\langle 1|.$$

In order to give an environment representation of this quantum channel, we specify an orthonormal basis $\{|i\rangle_{\mathcal{K}}\}_{i=0}^{2}$ of the ancillary space $\mathcal{K} \simeq \mathbb{C}^3$ and define the isometry

$$\begin{aligned} V : |0\rangle \quad &\mapsto \quad \sqrt{1-\lambda}\,|0\rangle \otimes |0\rangle_{\mathcal{K}} + \sqrt{\lambda}\,|0\rangle \otimes |1\rangle_{\mathcal{K}}, \\ V : |1\rangle \quad &\mapsto \quad \sqrt{1-\lambda}\,|1\rangle \otimes |0\rangle_{\mathcal{K}} + \sqrt{\lambda}\,|1\rangle \otimes |2\rangle_{\mathcal{K}}. \end{aligned}$$

Example 1.20 (Hadamard channels)**.** The phase damping channel is a particular instance of a *Hadamard channel*. Let $H \in \mathbb{C}^{d\times d}$ be a positive matrix whose diagonal entries are all equal to 1. Then

$$\rho \mapsto H * \rho$$

defines a quantum channel, where '$*$' denotes the entry-wise product (a.k.a. *Hadamard product*), i.e. $(H * \rho)_{ij} = H_{ij}\rho_{ij}$, where the matrix elements are w.r.t. a fixed orthonormal basis $\{|i\rangle\}_{i=1}^{d}$. Showing that Hadamard channels are indeed quantum channels is most easily done by observing that the set of Hadamard channels coincides with the set of quantum channels with diagonal Kraus operators. Consider a quantum channel $\rho \mapsto \rho' := \sum_k A_k \rho A_k^*$ with $\langle i|A_k|j\rangle = \delta_{ij}a_{ki}$. This is a Hadamard channel since $\langle i|\rho'|j\rangle = \langle i|\rho|j\rangle H_{ij}$ with $H_{ij} = \sum_k a_{ki}\bar{a}_{kj}$. For the converse direction, observe that the last equation can be seen as a decomposition of H into positive rank-one operators. In this way, we can construct diagonal Kraus operators from H, and so prove that Hadamard channels are indeed completely positive.

Choi-matrices If a quantum channel, or a more general linear map, acts on a finite-dimensional input space (with possibly infinite-dimensional output space), the following will turn out to be a useful representation tool:

Definition 1.60 (Choi matrix)**.** *For finite-dimensional $\mathcal{H}_1 \simeq \mathbb{C}^{d_1}$ define $|\Omega\rangle := \sum_{i=1}^{d_1} |ii\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_1$ where each i labels an element of a fixed orthonormal basis*[14]*. The* Choi matrix *$C \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$ of a linear map $T : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ is defined as*

$$C := (\mathrm{id} \otimes T)\big(|\Omega\rangle\langle\Omega|\big).$$

Note that $|\Omega\rangle/\sqrt{d}$ is a unit vector corresponding to a maximally entangled state. The usefulness of the Choi matrix stems from a simple Lemma:

[14]The notation $\mathcal{H}_1 \otimes \mathcal{H}_1$ should be read as $\mathcal{H}_0 \otimes \mathcal{H}_1$ where $\mathcal{H}_0$ is isomorphic to $\mathcal{H}_1$ and in addition we identify two orthonormal bases.

Lemma 1.61 (Cyclicity of maximally entangled state vectors)**.** *Let $\mathcal{H}_1 \simeq \mathbb{C}^{d_1}$ be finite-dimensional and $|\Omega\rangle := \sum_{i=1}^{d_1} |ii\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_1$. For any $\psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ define $A := \mathcal{I}(\psi) \in \mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$, where $\mathcal{I}$ is the Hilbert-Schmidt isomorphism constructed via Eq.(1.48) (w.r.t. the same basis that defines Ω). Then*

$$|\psi\rangle = (\mathbb{1} \otimes A)|\Omega\rangle. \tag{1.96}$$

Proof. Expanding in a product basis like $|\psi\rangle = \sum_{i=1}^{d_1} \sum_k A_{ik}|i\rangle \otimes |e_k\rangle$ we obtain $\mathcal{I}(\psi) = \sum_{i=1}^{d_1} \sum_k A_{ik}|e_k\rangle\langle i|$ so that Eq.(1.96) follows by insertion. □

Clearly, the statement of the Lemma holds similarly for interchanged tensor factors. In particular, for any $\psi \in \mathcal{H}_2 \otimes \mathcal{H}_1$ there is an $A \in \mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ so that $|\psi\rangle = (A \otimes \mathbb{1})|\Omega\rangle$.

Theorem 1.62 (Choi)**.** *Let $\mathcal{H}_1 \simeq \mathbb{C}^{d_1}$ be finite-dimensional and $T : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ be a linear map with Choi matrix $C \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$. Then*

(i) The map $T \mapsto C$ is a bijection whose inverse ($C \mapsto T$) is characterized by

$$\operatorname{tr}[T(A)B] = \operatorname{tr}\left[C(A^T \otimes B)\right], \quad \forall A \in \mathcal{B}(\mathcal{H}_1), B \in \mathcal{B}(\mathcal{H}_2), \tag{1.97}$$

where the transpose is w.r.t. the basis that is used in the definition of C.

(ii) $C = C^*$ iff $T(A)^* = T(A^*)$ for all $A \in \mathcal{B}(\mathcal{H}_1)$.

(iii) C is positive iff T is completely positive.

(iv) $\operatorname{tr}_2[C] = \mathbb{1}$ iff T is trace-preserving.

(v) $\operatorname{tr}_1[C] = \mathbb{1}$ iff T is unital.

Proof. (i) Note that via Eq.(1.97) T and C mutually determine each other so that Eq.(1.97) specifies a bijection if we regard C as an unconstrained element in $\mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$. That this C is indeed the Choi matrix is verified by

$$\begin{aligned} \operatorname{tr}[T(A)B] &= \operatorname{tr}[AT^*(B)] = \operatorname{tr}[\mathbb{F}(\mathrm{id} \otimes T^*)(A \otimes B)], \\ &= \operatorname{tr}[|\Omega\rangle\langle\Omega|(\Theta \otimes T^*)(A \otimes B)], \\ &= \operatorname{tr}\left[(\mathrm{id} \otimes T)\big(|\Omega\rangle\langle\Omega|\big)(A^T \otimes B)\right]. \end{aligned}$$

Here we have used the property of the flip operator from Exercise 1.27 (b) together with $\mathbb{F} = (\Theta \otimes \mathrm{id})\big(|\Omega\rangle\langle\Omega|\big)$, where Θ denotes the matrix transposition.

(ii) Since $C^* = \sum_{i,j} |j\rangle\langle i| \otimes T\big(|i\rangle\langle j|\big)^*$ with mutually orthogonal $|i\rangle\langle j|$, we have that this equals $C = \sum_{i,j} |j\rangle\langle i| \otimes T\big(|j\rangle\langle i|\big)$ iff $T\big(|i\rangle\langle j|\big)^* = T\big(|j\rangle\langle i|\big)$ holds for all i, j. In other words, $C = C^*$ iff $T(A)^* = T(A^*)$ holds for all $A = |i\rangle\langle j|$. By expanding an arbitrary A in that basis, the general statement follows.

(iii) The requirements in the definition of complete positivity of T imply positivity of the Choi matrix as a special case. In order to prove the converse, realize that it suffices to show $(\mathrm{id}_n \otimes T)\big(|\psi\rangle\langle\psi|\big) \geq 0$ for all $\psi \in \mathbb{C}^n \otimes \mathcal{H}_1$

and all $n \in \mathbb{N}$ since the spectral decomposition of an arbitrary positive trace-class operator allows us to restrict to rank-one operators. Lemma 1.61, with interchanged tensor factors, now enables us to write $|\psi\rangle = (A \otimes \mathbb{1})|\Omega\rangle$ for some $A \in \mathcal{B}(\mathcal{H}_1, \mathbb{C}^n)$. Then

$$(\mathrm{id}_n \otimes T)\big(|\psi\rangle\langle\psi|\big) = (A \otimes \mathbb{1}) \underbrace{(\mathrm{id}_{d_1} \otimes T)\big(|\Omega\rangle\langle\Omega|\big)}_{=C \geq 0} (A \otimes \mathbb{1})^* \geq 0.$$

(iv) Using that $\mathrm{tr}\left[T\big(|i\rangle\langle j|\big)\right] = \langle j|T^*(\mathbb{1})|i\rangle$ the claim follows from $\mathrm{tr}_2[C] = \sum_{ij} |i\rangle\langle j| \mathrm{tr}\left[T\big(|i\rangle\langle j|\big)\right] = T^*(\mathbb{1})^T$.

(v) Using that $\mathrm{tr}\left[|i\rangle\langle j|\right] = \delta_{i,j}$ we get $\mathrm{tr}_1[C] = \sum_{ij} \mathrm{tr}\left[|i\rangle\langle j|\right] T\big(|i\rangle\langle j|\big) = T(\mathbb{1})$, which completes the proof. □

Part (iii) of Thm.1.62 should be particularly emphasized: while the definition of complete positivity requires positivity of $(\mathrm{id}_n \otimes T)(A)$ for all $n \in \mathbb{N}$ and all positive A, Choi's theorem shows that $n = d_1$ and the choice $A = |\Omega\rangle\langle\Omega|$ is sufficient. Note that, by using that T is completely positivity iff T^* is (Thm.1.55), we can equivalently apply Choi's theorem to T^*, then with $n = d_2$. In both cases, we are left with a square matrix of dimension $d_1 d_2$.

We will now return to Kraus decompositions and in particular use the Choi matrix to prove the existence of a structured Kraus decomposition for every completely positive map with finite-dimensional input space.

Corollary 1.63 (Kraus decomposition)**.** *Let $T : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ be a linear map and $d_i := \dim(\mathcal{H}_i)$ with $d_1 < \infty$. Then there are two Hilbert-Schmidt orthogonal families of operators $\{A_k\}_{k=1}^r, \{B_k\}_{k=1}^r$ in $\mathcal{B}_2(\mathcal{H}_1, \mathcal{H}_2)$ with $r \leq d_1 d_2$ such that*

$$T(\cdot) = \sum_{k=1}^{r} A_k \cdot B_k^*. \tag{1.98}$$

Moreover, if T is completely positive, we can in addition choose $B_k = A_k$ for all k.

Proof. We will construct the Kraus decomposition from the Choi matrix $C \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$ of T using Lemma 1.61. Since the Choi matrix is trace-class, we can invoke the Schmidt-decomposition for compact operators and write $C = \sum_{k=1}^r |\psi_k\rangle\langle\varphi_k|$, where $\{\psi_k\}, \{\varphi_k\}$ are two orthogonal families in $\mathcal{H}_1 \otimes \mathcal{H}_2$. Using Lemma 1.61 and defining $A_k := \mathcal{I}(\psi_k)$, $B_k := \mathcal{I}(\varphi_k)$ we can express $|\psi_k\rangle = (\mathbb{1} \otimes A_k)|\Omega\rangle$ and $|\varphi_k\rangle = (\mathbb{1} \otimes B_k)|\Omega\rangle$. As $\mathcal{I}$ is an isomorphism onto the Hilbert-Schmidt class, the A_k's are orthogonal w.r.t. the Hilbert-Schmidt inner product, and so are the B_k's. The Choi matrix now reads

$$C = \sum_{k=1}^{r} (\mathbb{1} \otimes A_k)|\Omega\rangle\langle\Omega|(\mathbb{1} \otimes B_k)^*.$$

The representation claimed in Eq.(1.98) then follows from the fact that there is a unique T corresponding to C (Thm.1.62 (i)). If T is completely positive, then C is positive (Thm.1.62(iii)) so that we can choose $\varphi_k = \psi_k$ and thus $B_k = A_k$. □

Instruments For describing processes that output classical information in the form of a measurement outcome *and* a post-measurement quantum system, it is useful to introduce *instruments*. In a way, instruments generalize quantum channels and POVMs by merging them. We begin with the formal definition:

Definition 1.64 (Instrument (in Schrödinger picture))**.** *Let $(X, \mathbb{B})$ be a measurable space and denote by $CP(\mathcal{H}_1, \mathcal{H}_2)$ the set of completely positive maps from $\mathcal{B}_1(\mathcal{H}_1)$ to $\mathcal{B}_1(\mathcal{H}_2)$. A map $I : \mathbb{B} \to CP(\mathcal{H}_1, \mathcal{H}_2), Y \mapsto I_Y$ is called an* instrument *if (i) I_X is trace-preserving, and (ii) for all countable disjoint partitions $X = \cup_k X_k$ with $X_k \in \mathbb{B}$ it holds that $I_X(\rho) = \sum_k I_{X_k}(\rho)$ with convergence in trace-norm for all $\rho \in \mathcal{B}_1(\mathcal{H}_1)$.*

Note that the definition implies that $I_J + I_Y = I_{J \cup Y}$ for all disjoint $J, Y \in \mathbb{B}$. The interpretation of an instrument is as follows. Upon input of a quantum system characterized by a density operator $\rho \in \mathcal{B}_1(\mathcal{H}_1)$, the instrument yields two outputs: (i) a measurement result that is contained in Y with probability $p(Y) := \mathrm{tr}\,[I_Y(\rho)]$ and (ii) a quantum system described by a density operator in $\mathcal{B}_1(\mathcal{H}_2)$. Conditioned on having received a measurement outcome in Y, the quantum system at the output is described by the density operator $I_Y(\rho)/p(Y)$. That is, if one ignores the measurement outcome, the instrument gives rise to a quantum channel I_X, and if one ignores (i.e., traces out) the quantum output, the instrument gives rise to a POVM $Y \mapsto I_Y^*(\mathbb{1})$.

One way to arrive at an instrument is to use a quantum channel $T : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2 \otimes \mathcal{H}_3)$ that outputs a composite system of which one part undergoes a measurement that is described by a POVM $M : \mathbb{B} \to \mathcal{B}(\mathcal{H}_3)$. This results in an instrument of the form

$$I_Y(\rho) = \mathrm{tr}_3\big[\big(\mathbb{1} \otimes M(Y)\big)T(\rho)\big].$$

In fact, one can show that every instrument can be obtained in this way.

For any quantum channel and any discrete POVM there are simple ways of constructing an instrument that implements the channel or the POVM, respectively.

On the one side, given a quantum channel T with Kraus representation $T(\cdot) = \sum_{i \in X} K_i \cdot K_i^*$ where $X \subseteq \mathbb{N}$ is any index set, we can construct an instrument via $I_Y(\cdot) := \sum_{i \in Y} K_i \cdot K_i^*$. Here $\mathbb{B}$ would simply be the set of all subsets of X. This instrument 'implements' T in the sense that $I_X = T$.

On the other side, given a POVM M on a discrete measurable space $(X, \mathbb{B})$ with $\mathbb{B}$ the powerset of X, we can construct an instrument

$$I_Y(\rho) := \sum_{i \in Y} M(Y)^{1/2} \rho M(Y)^{1/2}. \tag{1.99}$$

This is called the *Lüders instrument* corresponding to the POVM M. The instrument implements M in the sense that $M(Y) = I_Y^*(\mathbb{1})$ for all $Y \in \mathbb{B}$. If the POVM M is in addition projection valued, then Eq.(1.99) is said to be an *ideal measurement* or an *ideal instrument*. Traditionally, these are the ones that

are used in quantum mechanics textbooks to describe measurements and their effect on the quantum system.

Note that one property of ideal measurements is *repeatability*. Physically, this means that if we repeat the measurement (with the same ideal instrument), then the outcome of the second measurement will be identical to the outcome of the first measurement. Mathematically, this is reflected by the fact that $I_Y \circ I_Y = I_Y$ for any $Y \in \mathbb{B}$.

Naimark's theorem One of the recurrent mantras of quantum information theory is the use of larger Hilbert spaces for simplifying mathematical representations. We have already seen two incarnations of this: the purification of mixed state density operators and the representation of a quantum channel by a unitary evolution acting on system plus environment. In this section, we apply the same mantra first to POVMs and later to sets of operators and represent them, in a larger space, by PVMs and sets of commuting operators, respectively. The core result is the following:

Theorem 1.65 (Naimark's dilation theorem)**.** *Let $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ be a POVM on a measurable space $(X, \mathbb{B})$. There exists a Hilbert space $\mathcal{K}$, an isometry $V : \mathcal{H} \to \mathcal{K}$ and a PVM $M' : \mathbb{B} \to \mathcal{B}(\mathcal{K})$ s.t. for all $Y \in \mathbb{B}$:*

$$V^* M'(Y) V = M(Y). \tag{1.100}$$

If the set X of measurement outcomes is finite, one can choose $\dim(\mathcal{K}) = \sum_{x\in X} \operatorname{rank}\big(M_x\big)$, where $M_x := M(\{x\})$ corresponds to the measurement outcome $x \in X$.

Remark: Since $\dim(\mathcal{K}) \geq \dim(\mathcal{H})$ we can regard $\mathcal{H}$ as a subspace of $\mathcal{K}$. Denoting the corresponding orthogonal projection from $\mathcal{K}$ onto $\mathcal{H}$ by $P_{\mathcal{H}}$, Eq.(1.100) then becomes

$$P_{\mathcal{H}} M'(Y)\big|_{\mathcal{H}} = M(Y). \tag{1.101}$$

We will provide an elementary proof for the case of finitely many measurement outcomes. The general case follows from Stinespring's dilation theorem.

Proof. We define $\tilde{\mathcal{K}} := \bigoplus_{x\in X} \mathcal{K}_x$ with $\mathcal{K}_x := \ker(M_x)^\perp$ and equip it with an inner product

$$\langle \varphi, \phi \rangle_{\mathcal{K}} := \sum_{x\in X} \langle \varphi_x, M_x \phi_x \rangle,$$

where $\varphi = \oplus_x \varphi_x$ and $\phi = \oplus_x \phi_x$. The space $\mathcal{K}$ is then chosen to be the completion of $\tilde{\mathcal{K}}$ w.r.t. to this inner product. Therefore, $\dim(\mathcal{K}) = \sum_{x\in X} \operatorname{rank}\big(M_x\big)$. $\mathcal{H}$ is isometrically embedded in $\tilde{\mathcal{K}}$, and thus in $\mathcal{K}$, as follows: for any $\psi \in \mathcal{H}$ let ψ_x be the projection of ψ to $\mathcal{K}_x$. Then $\Psi := \oplus_x \psi_x$ satisfies

$$\langle \Psi, \Psi \rangle_{\mathcal{K}} = \sum_{x\in X} \langle \psi_x, M_x \psi_x \rangle = \langle \psi, \sum_x M_x \psi \rangle = \langle \psi, \psi \rangle.$$

So $V : \psi \mapsto \Psi$ is an isometry. Defining $\mathbb{1}_x$ the identity operator on $\mathcal{K}_x$ we construct a PVM M' by setting $M'_x := \mathbb{1}_x$. Clearly, $M'_x \geq 0$, $(M'_x)^2 = M'_x$ and $\sum_x M'_x = \mathbb{1}$ so that M' is indeed a PVM. Moreover, as desired

$$\langle \psi, V^* M'_x V \psi \rangle = \langle V\psi, M'_x V\psi \rangle_{\mathcal{K}} = \langle \psi, M_x \psi \rangle.$$

□

One of the consequences of Naimark's dilation theorem is that we can regard every POVM as arising from a sharp measurement that is performed on system plus environment:

Corollary 1.66 (Environment representation of POVMs)**.** *Let $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ be a POVM on a measurable space $(X, \mathbb{B})$. There is a Hilbert space $\mathcal{K}_0$, a unit vector $\psi \in \mathcal{K}_0$ and a PVM $M' : \mathbb{B} \to \mathcal{B}(\mathcal{H} \otimes \mathcal{K}_0)$ so that for all $Y \in \mathbb{B}$:*

$$\operatorname{tr}[\rho M(Y)] = \operatorname{tr}\left[\big(\rho \otimes |\psi\rangle\langle\psi|\big) M'(Y)\right] \qquad \forall \rho \in \mathcal{B}_1(\mathcal{H}). \tag{1.102}$$

Conversely, if ψ and M' are as specified, then Eq.(1.102) uniquely defines a POVM $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$.

Proof. W.l.o.g. we can assume that the space $\mathcal{K}$ appearing in Naimark's theorem (Thm.1.65) is isomorphic to $\mathcal{H} \otimes \mathcal{K}_0$ for some Hilbert space $\mathcal{K}_0$. This can be achieved by isometrically embedding $\mathcal{K}$, if necessary, into a larger space, since this does not change the main result of Naimark's theorem. For the same reason, we can assume that the isometry $V : \mathcal{H} \to \mathcal{K}$ of Naimark's theorem is such that $V(\mathcal{H})$ is not dense in $\mathcal{K}$. Under these assumptions, by copying the argument of the proof of Thm.1.57, we can extend the isometry to a unitary $U \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K}_0)$ so that $V = U\big(\mathbb{1} \otimes |\psi\rangle\big)$ for some unit vector $\psi \in \mathcal{K}_0$. Naimark's theorem then leads to Eq.(1.102) after absorbing the unitary U into the PVM M'.

For the converse direction note that $M(Y) := \big(\mathbb{1} \otimes \langle\psi|\big) M'(Y) \big(\mathbb{1} \otimes |\psi\rangle\big)$ inherits all necessary properties for becoming a POVM from M'. □

Corollary 1.67 (Density of PVMs in infinite dimensions)**.** *Let $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ be a POVM on an infinite dimensional Hilbert space $\mathcal{H}$. There exists a sequence of PVMs $M^{(n)} : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ that converge to M in the sense that for every $\rho \in \mathcal{B}_1(\mathcal{H})$:*

$$\lim_{n\to\infty} \sup_{Y \in \mathbb{B}} \left| \operatorname{tr}\left[\rho M^{(n)}(Y)\right] - \operatorname{tr}[\rho M(Y)] \right| = 0.$$

Proof. Let $P_n \in \mathcal{B}(\mathcal{H})$ be a sequence of orthogonal projectors onto n-dimensional subspaces $\mathcal{H}_n$ of $\mathcal{H}$ such that $P_n \to \mathbb{1}$ converges strongly for $n \to \infty$. Then $\mathbb{B} \ni Y \mapsto P_n M(Y) P_n$ forms a POVM on $\mathcal{H}_n$, which by Naimark's theorem can be dilated to a PVM $M^{(n)}$ on a larger space $\mathcal{K} \supseteq \mathcal{H}_n$. However, since $\dim(\mathcal{H}) = \infty$, we can choose $\mathcal{K} = \mathcal{H}$. Then $P_n M^{(n)}(Y) P_n = P_n M(Y) P_n$ holds for every $Y \in \mathbb{B}$. Therefore,

$$\begin{aligned} \left| \operatorname{tr}\left[\rho\big(M^{(n)}(Y) - M(Y)\big)\right] \right| &= \left| \operatorname{tr}\left[(\rho - P_n\rho P_n + P_n\rho P_n)\big(M^{(n)}(Y) - M(Y)\big)\right] \right| \\ &\leq \|\rho - P_n\rho P_n\|_1 \end{aligned} \tag{1.103}$$

follows from Hölder's inequality together with $\|M^{(n)}(Y) - M(Y)\|_\infty \leq 1$. Finally, the remaining expression in Eq.(1.103) is a null sequence due to Thm. 1.7. □

Commuting dilations A simple but central aspect of Naimark's theorem is that operators that are in general not commuting are represented by commuting ones in a larger space. This point is emphasized in the following corollary:

Corollary 1.68 (Commuting Hermitian dilations)**.** *Let $H_1, \ldots, H_n \in \mathcal{B}(\mathcal{H})$ be Hermitian operators. There is a Hilbert space $\mathcal{K}$ of dimension $\dim(\mathcal{K}) \leq (n+1)\dim(\mathcal{H})$, an isometry $V : \mathcal{H} \to \mathcal{K}$ and pairwise commuting Hermitian operators $K_1, \ldots, K_n \in \mathcal{B}(\mathcal{K})$ s.t. $H_i = V^* K_i V$ for all i.*

Proof. Let $H_i = B_i - B_{n+i}$ be the decomposition of H_i into its orthogonal positive and negative parts. With $c := \big\| \sum_{i=1}^{2n} B_i \big\|_\infty$ define $A_i := B_i/c$ and $A_0 := \mathbb{1} - \sum_{i=1}^{2n} A_i$. Then $A_0, \ldots, A_{2n}$ are positive operators that sum up to one, and therefore can be regarded as forming a POVM. To this POVM we can apply Naimark's dilation theorem. The dimension of the dilation space $\mathcal{K}$ can then be bounded by $\dim(\mathcal{K}) \leq \sum_{i=0}^{2n} \operatorname{rank}(A_i) \leq (n+1)\dim(\mathcal{H})$ where the last inequality follows from $\operatorname{rank}(A_i) + \operatorname{rank}(A_{n+i}) \leq \dim(\mathcal{H})$ for all $i \in \{1, \ldots, n\}$. If we denote by $P_k \in \mathcal{B}(\mathcal{K})$ the orthogonal projection that Naimark's theorem assigns to A_k via the relation $A_k = V^* P_k V_k$, then we can express

$$H_i = V^* K_i V, \qquad \text{with} \quad K_i := c\big(P_i - P_{n+i}\big).$$

Commutativity of the P_i's then implies that all K_i's commute as well. □

If we do not insist on Hermiticity of the commuting dilations, there is an even simpler construction whose proof does not resort to Naimark's theorem:

Proposition 1.69 (Commuting dilations)**.** *For any finite sequence of operators $A_0, \ldots, A_{n-1} \in \mathcal{B}(\mathcal{H})$ there exist pairwise commuting operators $K_0, \ldots, K_{n-1} \in \mathbb{B}(\mathbb{C}^n \otimes \mathcal{H})$ and a unit vector $|0\rangle \in \mathbb{C}^n$ s.t.*

$$A_k = \big(\langle 0| \otimes \mathbb{1}\big) K_k \big(|0\rangle \otimes \mathbb{1}\big) \qquad \forall k. \tag{1.104}$$

This means that we can regard K_k as a (possibly infinite) 'block matrix' that contains A_k in its north-west block.

Proof. Regarding the range of all indices as $\mathbb{Z}_n$ with addition modulo n we set

$$K_k := \sum_{i,j} |i\rangle\langle j| \otimes A_{i-j+k},$$

for a fixed orthonormal basis $\{|i\rangle\}_{i=0}^{n-1} \subset \mathbb{C}^n$. This construction clearly satisfies Eq.(1.104). To see that this leads to a commuting set of operators note that

$$K_{k_1} K_{k_2} = \sum_{i_1, j_2} |i_1\rangle\langle j_2| \otimes \sum_{j_1} A_{i_1 - j_1 + k_1} A_{j_1 - j_2 + k_2}. \tag{1.105}$$

Replacing j_1 with $j_1 - k_2 + k_1$ does not change this expression (as we sum over all j_1 anyhow) but it effectively interchanges $k_1 \leftrightarrow k_2$. Hence, $K_{k_1} K_{k_2} = K_{k_2} K_{k_1}$. □

Exercise 1.31 (Complete positivity). Consider finite-dimensional Hilbert spaces.

(a) Show that any linear map $T : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ can be written as a linear combination of four completely positive maps.

(b) Write matrix transposition $\Theta(A) := A^T$ as a real linear combination of two completely positive maps.

(c) Use the definition of complete positivity to prove that $X \to AXA^*$ is completely positive for any $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$.

(d) Show that if T_1, T_2 are completely positive maps, then $T_1 \circ T_2$, $T_1 + T_2$, $T_1 \otimes T_2$ are completely positive as well.

(e) Show that for the partial trace(s) positivity implies complete positive by using not much more than the definitions of the partial trace and of complete positivity.

Exercise 1.32 (Positive but not completely). Let $K \in \mathbb{C}^{d\times d}$ be such that $K^T = -K$ and $K^*K \leq \mathbb{1}$. Show that the map $T : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ defined as $T(X) := \mathrm{tr}\,[X]\,\mathbb{1} - X - KX^TK^*$ is positive. Is it completely positive?

Exercise 1.33 (Kraus operators).

(a) Which is the minimal number of Kraus operators necessary to represent the *phase damping channel* ?

(b) Decoherence and decay processes can often be described by a map of the form

$$T(\rho) = e^{-t}\rho + \left(1 - e^{-t}\right)\mathrm{tr}\,[\rho]\,\sigma,$$

where $t \in \mathbb{R}_+$ and σ is a density operator. Find a Kraus representation for this map.

Exercise 1.34 (Dual maps). Derive the dual map (i.e., description in the Heisenberg picture) of the following quantum channels:

(a) $T(\rho) := \lambda\rho + (1-\lambda)\mathrm{tr}\,[\rho]\,\sigma$, where σ is a density operator and $\lambda \in [0, 1]$.

(b) The partial trace $\mathrm{tr}_2 : \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}_1(\mathcal{H}_1)$.

(c) $T(\rho) := \rho \otimes \sigma$ where σ is a density operator.

(d) $T(\rho) := (\mathbb{1}\mathrm{tr}\,[\rho] + \rho^T)/(d-1)$, with $d < \infty$ the Hilbert space dimension.

Exercise 1.35 (Commuting dilations).

(a) Let $\sigma_1, \sigma_2 \in \mathbb{C}^{2\times 2}$ be the first two Pauli matrices. Give an explicit construction of two Hermitian, commuting block matrices Σ_1, Σ_2 that are such that σ_i is the north-west block of Σ_i, for both $i \in \{1, 2\}$.

(b) Prove the following statement: there is a Hilbert space $\mathcal{K}$, an isometry $V : \mathbb{C}^d \to \mathcal{K}$ and a Hermiticity-preserving linear map $R : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathcal{K})$ so that (i) $[R(\rho), R(\sigma)] = 0$ for all $\rho, \sigma \in \mathcal{B}(\mathbb{C}^d)$ and (ii) $V^*R(\rho)V = \rho$ for all $\rho \in \mathcal{B}(\mathbb{C}^d)$.

Chapter 2

Basic trade-offs

2.1 Uncertainty relations

Heisenberg's uncertainty relation is one of the most famous consequences of the formalism of quantum theory. It is one out of at least three superficially related consequences that can be traced back to Heisenberg's original paper:

- *Preparation uncertainty relations*: Constraints on individual states regarding how sharp the values of different observables can be in that state.
- *Measurement uncertainty relations*: Constraints on different measurements concerning their simultaneous implementability.
- *Measurement-disturbance relations*: Constraints on the minimal disturbance caused by a quantum measurement.

We will discuss central aspects of these three points in the following two sections. In the case of observables or sharp POVMs, a central property in the discussion of uncertainty relations for both preparation and measurement will be the non-commutativity of operators. So, let us briefly recall some notation and useful mathematical background related to commutators.

The *commutator* of two operators that act on the same space will be written as $[A, B] := AB - BA$. If the operators are Hilbert-Schmidt class, then the commutator is obviously trace-less and if A, B are Hermitian, the commutator is anti-Hermitian (i.e., it becomes Hermitian when multiplied by i). A is said to *commute* with B if $[A, B] = 0$. If a collection of normal, compact operators commute pairwise, then they can be diagonalized simultaneously. That is, there is a basis in which they are all diagonal. An analogous statement is true for arbitrary sets of normal operators. Via continuous functional calculus, this implies that if $[A, B] = 0$ holds for two normal operators, then $[f(A), B] = 0$ holds as well for any continuous function f. In particular, it holds for $\sqrt{A}$ when A is positive.

Variance-based preparation uncertainty relations

Theorem 2.1 (Robertson uncertainty relation)**.** *Let $H_1, \dots, H_n \in \mathcal{B}(\mathcal{H})$ be Hermitian, $\rho \in \mathcal{B}_1(\mathcal{H})$ a density operator and define $\langle X \rangle := \mathrm{tr}\,[\rho X]$ for any $X \in \mathcal{B}(\mathcal{H})$. Then the $n \times n$* covariance matrix[1] *$V_{kl} := \frac{1}{2}\langle \{H_k - \langle H_k \rangle, H_l - \langle H_l \rangle\}_+ \rangle$ and the* commutator matrix *$\sigma_{kl} := \frac{i}{2}\langle [H_k, H_l] \rangle$ satisfy*

$$V \geq i\sigma, \qquad \text{and} \qquad \det(V) \geq \det(\sigma). \tag{2.1}$$

Remark: Positivity of covariance matrices is a well-known and simple to show property for classical random variables. In the quantum context, the new term that leads to a more demanding inequality is the commutator matrix.

Proof. We abbreviate $H_k - \langle H_k \rangle \mathbb{1} =: A_k$ and define an $n \times n$ matrix $R_{kl} := \langle A_k A_l \rangle$. The claim is that $R \geq 0$. In order to see this, note that $R_{kl} = \langle A_k \sqrt{\rho}, A_l \sqrt{\rho} \rangle$ is a Gram matrix w.r.t. the Hilbert-Schmidt inner product and thus positive. Decomposing every matrix element R_{kl} into real and imaginary part and using that $\overline{R}_{kl} = R_{lk}$ together with $[A_k, A_l] = [H_k, H_l]$ we obtain $R = V - i\sigma$. So the l.h.s. of Eq.(2.1) is just a reformulation of $R \geq 0$.

The determinant inequality, in turn, is implied by $V \geq i\sigma$. Here, a central ingredient in the argumentation is the anti-symmetry of σ. First, this implies that $\det(\sigma)$ can be non-zero only in even dimensions. Second, assuming even dimensions, σ can be block-diagonalized to a direct sum of anti-symmetric 2×2 matrices via orthogonal transformations. From here one can use a classical result by Williamson on symplectic normal forms [37]: there exists a matrix $S \in SL(n, \mathbb{R})$ which allows to map $V \mapsto SVS^T$ to the same block-diagonal structure while keeping the transformed σ unchanged. Hence, the sought implication is reduced to the one for 2×2 matrices, where it can be shown by direct computation. For an alternative proof of the determinant inequality see Exercise 2.4. □

For a pair of observables, writing out the determinant inequality immediately leads to the following, better known, uncertainty relation:

Corollary 2.2 (Robertson-Schrödinger uncertainty relation)**.** *Let $A, B \in \mathcal{B}(\mathcal{H})$ be Hermitian, $\rho \in \mathcal{B}_1(\mathcal{H})$ a density operator, $\langle X \rangle := \mathrm{tr}\,[\rho X]$ for any $X \in \mathcal{B}(\mathcal{H})$, and $\mathrm{var}(A) := \langle A^2 \rangle - \langle A \rangle^2$. Then*

$$\mathrm{var}(A)\mathrm{var}(B) \;\geq\; \frac{1}{4}\big|\langle [A, B] \rangle\big|^2 + \frac{1}{4}\big\langle \{A - \langle A \rangle, B - \langle B \rangle\}_+ \big\rangle^2. \tag{2.2}$$

Moreover, equality holds iff $(\alpha A - \beta B)\rho = \gamma\rho$ for some $(\alpha, \beta, \gamma) \in \mathbb{C}^3 \setminus \{0\}$.

Remark: This corollary as well as Robertson's uncertainty relation in Thm. 2.1 also applies to Hermitian operators that are not necessarily bounded. What requires additional care in this case, are the domains of all involved operators.

[1] Here $\{\cdot, \cdot\}_+$ denotes the *anti-commutator*, defined as $\{A, B\}_+ = AB + BA$ and $H_k - \langle H_k \rangle$ should be read as $H_k - \langle H_k \rangle \mathbb{1}$.

For instance, in Robertson's uncertainty relation, if $\rho = |\psi\rangle\langle\psi|$ and if $\mathcal{D}(H_k H_l)$ is the domain of $H_k H_l$, then we need $\psi \in \bigcap_{kl} \mathcal{D}(H_k H_l)$. In this way, Heisenberg's uncertainty relation for position and momentum is obtained from Cor.2.2 by neglecting the covariance term on the r.h.s. and inserting $i\mathbb{1}$ for the commutator of the position and momentum operator.

Proof. As pointed out already, the inequality stated in Eq.(2.2) is just a reformulation of the determinant inequality in Eq.(2.1) for the special case of two observables. In order to characterize cases of equality we will, however, use a different proof. Assume for the moment that $\rho = |\psi\rangle\langle\psi|$ and set $\tilde{A} := A - \langle A\rangle\mathbb{1}$, $\tilde{B} := B - \langle B\rangle\mathbb{1}$. Then Cauchy-Schwarz gives

$$\left\|\tilde{A}\psi\right\|^2 \left\|\tilde{B}\psi\right\|^2 \geq |\langle\psi, \tilde{A}\tilde{B}\psi\rangle|^2 = \left(\mathrm{Re}\langle\psi, \tilde{A}\tilde{B}\psi\rangle\right)^2 + \left(\mathrm{Im}\langle\psi, \tilde{A}\tilde{B}\psi\rangle\right)^2. \tag{2.3}$$

Inserting the expressions defining $\tilde{A}$ and $\tilde{B}$ then leads to the claimed uncertainty relation in Eq.(2.2) for pure states. The advantage of this proof is that we know that equality in the Cauchy-Schwarz inequality, and thus in the uncertainty relation, holds iff $\alpha\tilde{A}\psi = \beta\tilde{B}\psi$ for some $\alpha, \beta \in \mathbb{C}$. This proves the claimed characterization of cases of equality for pure states (with γ necessarily being equal to $\alpha\langle A\rangle - \beta\langle B\rangle$).

The result can be lifted to mixed states by purification (Cor1.43). If a unit vector $\psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ characterizes a purification of $\rho = \mathrm{tr}_2|\psi\rangle\langle\psi|$ and if we use $\tilde{A} \otimes \mathbb{1}$ and $\tilde{B} \otimes \mathbb{1}$ in Eq.(2.3) instead of $\tilde{A}$ and $\tilde{B}$, then we arrive at the general form of the uncertainty relation in Eq.(2.2) for mixed states. Equality is then attained iff ψ is in the kernel of $(\alpha\tilde{A} - \beta\tilde{B}) \otimes \mathbb{1}$ for some $\alpha, \beta \in \mathbb{C}$. Exploiting the Schmidt-decomposition of ψ (1.42) we can see that this is equivalent to the statement that every eigenvector of ρ that corresponds to a non-zero eigenvalue has to be in the kernel of $(\alpha\tilde{A} - \beta\tilde{B})$. This, in turn, is equivalent to the claimed characterization. □

States, in particular pure states, that achieve equality in this uncertainty relation are sometimes called *minimal uncertainty states*. On should keep in mind, however, that they might not minimize the product of the variances ('uncertainties') among all states. Imposing equality only means that the two sides are equal—they are not necessarily small.

Joint measurability

Definition 2.3 (Joint measurability)**.** *Two POVMs $M_i : \mathbb{B}_i \to \mathcal{B}(\mathcal{H})$, $i \in \{1, 2\}$ on measurable spaces $(X_i, \mathbb{B}_i)$ are* jointly measurable *if there exists a POVM $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ defined on the product σ-algebra $\mathbb{B}$ on $X_1 \times X_2$ s.t.*

$$\begin{aligned} M(Y_1, X_2) &= M_1(Y_1) \quad \forall Y_1 \in \mathbb{B}_1, \\ M(X_1, Y_2) &= M_2(Y_2) \quad \forall Y_2 \in \mathbb{B}_2. \end{aligned}$$

Theorem 2.4 (Joint measurability vs. commutativity)**.** *Consider two POVMs $M_i : \mathbb{B}_i \to \mathcal{B}(\mathcal{H})$, $i \in \{1, 2\}$ on measurable spaces $(X_i, \mathbb{B}_i)$ and assume that at*

least one of them is sharp (i.e. projection valued). Then M_1 and M_2 are jointly measurable iff they commute in the sense that $\forall Y_i \in \mathbb{B}_i : [M_1(Y_1), M_2(Y_2)] = 0$. In that case the joint POVM $M : \mathbb{B} \to \mathcal{B}(\mathcal{H})$ is characterized by $M(Y_1 \times Y_2) = M_1(Y_1)M_2(Y_2)$.

Proof. Assume that the two POVMs commute. Since commutativity is a property that extends to the square root, we can use that

$$M_1(Y_1)M_2(Y_2) = \sqrt{M_1(Y_1)}M_2(Y_2)\sqrt{M_1(Y_1)} =: M(Y_1 \times Y_2)$$

defines a proper POVM, which by construction has M_1 and M_2 as its marginals in the sense of Def.2.3. So the two POVMs are jointly measurable. Note that for this direction we haven't used that any of the POVMs is sharp.

Conversely, suppose there is a joint POVM M and that M_1 is projection-valued. The core of the argument will be the fact that if a positive operator A is bounded from above by a projection $P \geq A$, then $A = AP = PA$ (cf. Exercise 1.8). This applies, in particular, to $M_1(Y_1) \geq M(Y_1 \times Y_2)$ and similarly to the case where Y_1 is replaced by $\bar{Y}_1 := X_1 \setminus Y_1$. Since $M_1(Y_1)M_1(\bar{Y}_1) = 0$ this leads to $M(\bar{Y}_1 \times Y_2)M_1(Y_1) = 0$ and with $M(Y_1 \times Y_2)M_1(Y_1) = M(Y_1 \times Y_2)$ we obtain

$$\begin{aligned} M_2(Y_2)M_1(Y_1) &= M(Y_1 \times Y_2)M_1(Y_1) + M(\bar{Y}_1 \times Y_2)M_1(Y_1) \\ &= M(Y_1 \times Y_2). \end{aligned}$$

Following the same steps, we can show that $M_1(Y_1)M_2(Y_2) = M(Y_1 \times Y_2)$. Hence, M_1 commutes with M_2. □

Entropic uncertainty relations Variance-based uncertainty relations depend on the values of the measurement results. While in some scenarios, especially for elementary physical properties such as position or momentum, there is a natural, if not unique, choice of these values, this is not always the case. In computational, information-theoretic, or cryptographic contexts, the measurement result is often just an abstract label, and any permutation of these labels would be equally good. Since the variance can change drastically under relabeling, a permutation-invariant quantity is more appropriate in these cases. Entropy is such a quantity:

Theorem 2.5 (Entropic uncertainty relations)**.** *Let $X := \{|x_i\rangle\}_{i=1}^d$, $Y := \{|y_j\rangle\}_{j=1}^d$ be two orthonormal bases of $\mathcal{H}$, $\rho \in \mathcal{B}(\mathcal{H})$ a density operator and S_X, S_Y the Shannon entropies of the probability distributions obtained by measuring ρ w.r.t X and Y, respectively. Then*

$$S_X + S_Y \geq S(\rho) + \log(1/c), \quad \textit{where} \quad c := \max_{i,j} |\langle x_i|y_j\rangle|^2. \tag{2.4}$$

Proof. Define a quantum channel $T_X(\cdot) := \sum_{i=1}^d |x_i\rangle\langle x_i|\langle x_i| \cdot |x_i\rangle$ and similarly T_Y by using Y instead of X. Then $S_X = -\mathrm{tr}\,[\rho \log T_X(\rho)] = S(\rho\|T_X(\rho)) + S(\rho)$.

Using the data processing inequality for the relative entropy, we obtain

$$\begin{aligned} S_X - S(\rho) &= S\big(\rho \| T_X(\rho)\big) \\ &\geq S\big(T_Y(\rho) \| T_Y \circ T_X(\rho)\big) \\ &= -S_Y - \sum_j \langle y_j|\rho|y_j\rangle \log\Big(\sum_i \underbrace{|\langle x_i|y_j\rangle|^2}_{\leq c} \langle x_i|\rho|x_i\rangle\Big). \\ &\geq -S_Y + \log(1/c), \end{aligned}$$

where the last step exploits the monotonicity of the logarithm, which enables the bound in terms of c. □

Example 2.1. Consider a pure state measured w.r.t. any basis $X = \{|k\rangle\}_{k=0}^{d-1}$ and the corresponding Fourier basis $Y = \{U|k\rangle\}_{k=0}^{d-1}$, where the unitary U implements the discrete Fourier transformation and is given by $U_{kl} := \exp(2\pi i k l)\sqrt{d}$. Then $c = 1/d$ so that

$$S_X + S_Y \geq \log d. \tag{2.5}$$

This means that if $S_X = 0$, i.e., the X-measurement has one outcome that occurs with certainty, then $S_Y = \log d$, which implies that the outcome of the Y-measurement is maximally uncertain.

2.2 Information–disturbance

No information without disturbance

Theorem 2.6 (Knill-Laflamme/no information without disturbance)**.** *Let $T : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ be a quantum channel and $\dim(\mathcal{H}_1) < \infty$. The following are equivalent:*

(i) There exists a quantum channel $D : \mathcal{B}_1(\mathcal{H}_2) \to \mathcal{B}_1(\mathcal{H}_1)$ s.t. $D \circ T = \mathrm{id}$.

(ii) For any Kraus representation $T(\cdot) = \sum_{j=1}^r K_j \cdot K_j^*$ there is a density matrix $\sigma \in \mathbb{C}^{r\times r}$ so that

$$K_i^* K_j = \sigma_{ij} \mathbb{1} \quad \forall i, j. \tag{2.6}$$

(iii) Any instrument $I : \mathbb{B} \to CP(\mathcal{H}_1, \mathcal{H}_2)$ on a measurable space $(X, \mathbb{B})$ that implements the channel (in the sense that $I_X = T$) satisfies:

$$I_Y^*(\mathbb{1}) \propto \mathbb{1} \quad \forall Y \in \mathbb{B}. \tag{2.7}$$

Proof. Assuming (i) and using a Kraus decomposition of $D(\cdot) = \sum_l A_l \cdot A_l^*$ we can exploit the bijective relation between a completely positive map and its Choi matrix (cf. Thm.1.62) to show that $(A_l K_j \otimes \mathbb{1})|\Omega\rangle = c_{lj}|\Omega\rangle$ for some complex number c_{lj} and thus $A_l K_j = \mathbb{1} c_{lj}$. As D^* is unital, this leads to

$K_i^* K_j = \sum_l K_i^* A_l^* A_l K_j = \mathbb{1}\sigma_{ij}$ with $\sigma_{ij} = \sum_l \bar{c}_{li} c_{lj}$. So σ is positive and since $\sum_j K_j^* K_j = \mathbb{1}$ we have to have $\mathrm{tr}\,[\sigma] = 1$, which proves (i)⇒(ii).

If (ii) holds, and I is an instrument that implements T, then the Kraus operators of I_Y have to satisfy Eq.(2.6) as well since they appear in a particular Kraus representation of T. Consequently, $I_Y^*(\mathbb{1})$ is proportional to $\mathbb{1}$. So (ii)⇒(iii).

For the converse direction ((iii)⇒(ii)) note first that for every Kraus operator K of T there exists an instrument with two outcomes, which we may label with K and $\neg K$, whose corresponding completely positive maps are given by $K \cdot K^* =: I_K(\cdot)$ and $I_{\neg K} := T - I_K$, respectively. Then Eq.(2.7) applied to this instrument implies that $K^*K \propto \mathbb{1}$. If K_j and K_i are two Kraus-operators, we know from the ambiguity of the Kraus representation (Prop. 1.58) that a multiple of any linear combination of them is a possible Kraus-operator as well. Consequently, in particular, $(K_i + \gamma K_j)^*(K_i + \gamma K_j) \propto \mathbb{1}$ for any $\gamma \in \mathbb{C}$. An application of the polarization identity of Eq.(1.7) then implies Eq.(2.6). Positivity of σ and $\mathrm{tr}\,[\sigma] = 1$ are then consequences of its definition and of unitality of T^*. So (iii)⇒(ii).

For proving (ii)⇒(i) we exploit the freedom in the Kraus representation (Prop.1.58) again. It allows choosing Kraus-operators for which $\sigma_{ij} = \delta_{ij} s_i$ is diagonal (by using the unitary that diagonalizes σ to construct new Kraus-operators according to Prop.1.58)). Then each $K_i = \sqrt{s_i} V_i$ is a multiple of an isometry $V_i : \mathcal{H}_1 \to \mathcal{K}_i \subseteq \mathcal{H}_2$ where the $\mathcal{K}_i$'s are mutually orthogonal subspaces of $\mathcal{H}_2$. The map $D(\cdot) := \sum_i V_i^* \cdot V_i$ then satisfies that $D \circ T = \mathrm{id}$. Moreover, $\sum_i V_i V_i^* = \sum_i \mathbb{1}_{\mathcal{K}_i} \leq \mathbb{1}$ and if equality does not hold, which is then due to $\mathcal{K}_0 := \mathcal{H}_2 \ominus \oplus_i \mathcal{K}_i$ being non-empty, we can always make D trace-preserving by adding a suitable completely positive map from $\mathcal{B}_1(\mathcal{K}_0)$ to $\mathcal{B}_1(\mathcal{H}_1)$. □

As already suggested by the name given to the theorem, the equivalence of (i) and (iii) has an interpretation that should be emphasized. Point (iii) means that no information about an input state ρ is contained in the measurement outcomes since their probabilities are all proportional to $\mathrm{tr}\,[\rho]$ with proportionality constants that depend on the instrument only, and not on ρ. Point (i) on the other hand, means that any 'disturbance' caused by T can be undone by some channel D. So (i)⇒(iii) means that if there is no (uncorrectable) disturbance, then no information about the state of the system is revealed. Conversely, (iii)⇒(i) means that if no information has leaked into the environment, then the effect of T can be undone.

The equivalent condition (ii) is sometimes called *Knill-Laflamme condition*. The following discussion aims as explaining the appearance of this condition in its natural environment.

Example 2.2 (Quantum error correcting codes)**.** The condition in Eq.(2.6) plays a crucial role in the context of quantum error correction. To see how, we first need to define what is a *quantum error correcting code* (QECC). A QECC is a linear subspace that can be thought of being the image of an isometry $V : (\mathbb{C}^2)^{\otimes k} \to \mathcal{H} := (\mathbb{C}^2)^{\otimes n}$. In this case, k qubits are encoded into n qubits via a quantum channel $E(\rho) := V^* \rho V$. Let $\mathcal{P}_d \subset \mathcal{B}(\mathcal{H})$ be the set of all tensor

products of n Pauli matrices (including $\sigma_0 = \mathbb{1}$) that differ on at most d tensor factors from the identity σ_0. A QECC is called an $[[n, k, d]]$ QECC, if

$$V^* F V \propto \mathbb{1} \tag{2.8}$$

holds for all $F \in \mathcal{P}_{d-1}$ but fails for some $F \in \mathcal{P}_d$. What is the reason behind this definition? Consider a quantum channel $\Phi : \mathcal{B}_1(\mathcal{H}) \to \mathcal{B}_1(\mathcal{H})$, which models the noise/decoherence/errors that affect the n qubits, whose Kraus-operators $\{A_i\}$ are all in the linear span of $\mathcal{P}_t$ with $t := \lfloor \frac{d-1}{2} \rfloor$. Then the Kraus operators $\{K_i\}$ of $T := \Phi \circ E$ satisfy Eq.(2.6) so that Thm.2.6 guarantees the existence of a *decoding* quantum channel D such that $D \circ \Phi \circ E = \mathrm{id}$. d is called the *distance* of the code and t can be interpreted as the number of errors the code corrects.

An important point to note is that a given $[[n, k, 2t + 1]]$-QECC does not only work for one noise-characterizing channel Φ, but for all channels whose Kraus-operators are in the linear span of $\mathcal{P}_t$.

2.3 Time–energy

Mandelstam-Tamm inequalities

Theorem 2.7 (Mandelstam-Tamm inequality)**.** *Let $A, H \in \mathcal{B}(\mathcal{H})$ be Hermitian, $\psi(t) = \exp[-itH]\psi(0)$, $t \in \mathbb{R}_+$ describe the time evolution of pure states in $\mathcal{H}$, $\langle A \rangle := \langle \psi(t), A\psi(t) \rangle$, $\Delta(A) := \left(\langle A^2 \rangle - \langle A \rangle^2 \right)^{1/2}$ and $\Delta(H)$ analogously. Then*

$$\Delta(H)\Delta(A) \geq \frac{1}{2} \left| \frac{d\langle A \rangle}{dt} \right| . \tag{2.9}$$

Moreover, for any Hermitian H, any unit vector $\psi(0)$ and any $\tau \geq 0$ there is a Hermitian $A \in \mathcal{B}(\mathcal{H})$ so that equality holds in Eq.(2.9) when evaluated at $t = \tau$.

Remarks: 1. Note that $\Delta(H)$ is time-independent, while $\Delta(A)$ as well as the right-hand side of Eq.(2.9) do generally depend on time (and are meant to be evaluated at the same time t). 2. With the necessary care concerning the domain, Eq.(2.9) also holds for an unbounded Hamiltonian H.

Proof. With $\frac{d}{dt}\langle A \rangle = i\langle [H, A] \rangle$ Eq.(2.9) is a direct consequence of the Robertson-Schrödinger uncertainty relation (Cor.2.2) when neglecting the anti-commutator term and taking the square root (since $\Delta(A) = \mathrm{var}(A)^{1/2}$).

For showing tightness of the inequality we define $A := B + B^*$ with $B := i(H - \langle H \rangle \mathbb{1})|\psi(\tau)\rangle\langle\psi(\tau)|$. By construction, $\langle A \rangle|_{t=\tau} = 0$ as well as $\langle \{A, H - \langle H \rangle \mathbb{1}\}_+ \rangle|_{t=\tau} = 0$. Moreover, $A\psi(\tau) = i(H - \langle H \rangle \mathbb{1})\psi(\tau)$ so that equality holds in the Robertson-Schrödinger uncertainty relation (with vanishing anti-commutator) and therefore also in Eq.(2.9). □

Corollary 2.8 (Life-time/energy-width uncertainty relation)**.**
Let $\psi(t) = \exp[-itH]\psi(0)$, $t \in \mathbb{R}_+$ describe the time evolution of pure states

and define $p(t) := |\langle\psi(t), \psi(0)\rangle|^2$. *Then*

$$\Delta(H)\, t \quad \geq \quad \arccos\left(\sqrt{p(t)}\right), \quad \textit{so that} \tag{2.10}$$

$$\Delta(H) t_{1/2} \geq \frac{\pi}{4} \quad , \quad \Delta(H) t_0 \geq \frac{\pi}{2}, \tag{2.11}$$

where $t_{1/2}$ *and* t_0 *are the shortest times for* $p(t)$ *to drop to* $1/2$ *and* 0, *respectively (i.e.,* $p(t_{1/2}) = 1/2$, *and* $p(t_0) = 0$*).*

Proof. We apply the Mandelstam-Tamm uncertainty relation in Eq.(2.9) to $A = |\psi(0)\rangle\langle\psi(0)|$. Then $p(t) = \langle A\rangle$ and $\Delta(A) = \left(p(t) - p(t)^2\right)^{1/2}$ so that

$$\Delta(H)\, \tau \leq \int_0^\tau \frac{|\dot{p}(t)|}{2\sqrt{p(t) - p(t)^2}}\, dt = \arccos\left(\sqrt{p(\tau)}\right).$$

The inequalities in Eq.(2.11) then follow from Eq.(2.10) by using $\arccos(\sqrt{1/2}) = \pi/4$ and $\arccos(0) = \pi/2$. □

Note that $p(t)$ can be interpreted as the probability of the system still being in its initial state after time t. That is, if a projective measurement with two outcomes and corresponding projectors $P_0 := |\psi(0)\rangle\langle\psi(0)|$ and $P_1 := \mathbb{1} - P_0$ is performed after time t, then the outcome corresponding to P_0 occurs with probability $p(t)$. For short times, the evolution of $p(t)$ is governed by the variance of the Hamiltonian as shown by the Taylor expansion

$$p(t) = 1 - \Delta(H)^2 t^2 + \mathcal{O}(t^3).$$

Evolution to orthogonal states In Cor.2.8 we have seen that the Mandelstam-Tamm inequality implies a lower bound on the time it takes for a quantum system to evolve to an orthogonal state. We will now discuss alternative bounds of this type and then derive a condition for the feasibility of such an evolution. The following Lemma will be the main ingredient in the proof of the subsequent 'quantum-speed-limit' theorem.

Lemma 2.9 (First zero of a characteristic function)**.** *Let* μ *be a Borel-probability measure on* $[0, \infty)$. *Define its* characteristic function $\chi : \mathbb{R} \to \mathbb{C}$ *by* $\chi(t) := \int_{\mathbb{R}_+} e^{-it\lambda}\, d\mu(\lambda)$ *and its p'th moment for any* $p > 0$ *by* $m_p := \int_{\mathbb{R}_+} \lambda^p\, d\mu(\lambda)$. *Then*

$$t_0 := \inf\left\{t > 0 | \chi(t) = 0\right\} \geq \frac{\pi}{(2m_p)^{1/p}}. \tag{2.12}$$

With this we can prove the following:

Theorem 2.10 (Generalized Margolus-Levitin bound)**.** *Let* $H \geq 0$ *be self-adjoint and* $\psi(t) := \exp[-iHt]\psi$, $t \in \mathbb{R}_+$ *for some unit vector* ψ *in the domain of* H. *If* $\langle\psi, H^p\psi\rangle$ *is defined for some* $p > 0$, *and* $t_0 := \inf\left\{t > 0 | \langle\psi, \psi(t)\rangle = 0\right\}$, *then*

$$t_0\, \langle\psi, H^p\psi\rangle^{1/p} \geq \frac{\pi}{2^{1/p}}. \tag{2.13}$$

Proof. Exploiting positivity of H and the spectral representation $H = \int_{\mathbb{R}_+} \lambda dP(\lambda)$ we can define a Borel-probability measure on $[0,\infty)$ via $\mu(Y) := \langle\psi, P(Y)\psi\rangle$. For $p > 0$, the p'th moment of μ is then given by $m_p = \langle\psi, H^p\psi\rangle$ and its characteristic function by

$$\int_{\mathbb{R}_+} e^{-i\lambda t} d\mu(\lambda) = \int_{\mathbb{R}_+} e^{-i\lambda t} d\langle\psi, P(\lambda)\psi\rangle = \langle\psi, e^{-iHt}\psi\rangle.$$

The claim follows then from Lemma 2.9. □

For $p = 2$ Eq.(2.13) is similar to the consequence that we obtained in Cor.2.8 from the Mandelstam-Tamm inequality. In fact, at first glance, Eq.(2.13) looks even stronger since there is a missing factor $1/\sqrt{2}$. Note, however, that Eq.(2.13) requires an additional assumption, namely positivity of the Hamiltonian.

For $p = 1$ Eq.(2.13) is called the *Margolus-Levitin* bound, which directly relates the energy of a pure state (w.r.t. a positive Hamiltonian) to the minimal time it takes to evolve into an orthogonal state. So far, we do, however, not know under which circumstances a pure state ψ will ever evolve to an orthogonal state under the time-evolution governed by the Hamiltonian H. For obtaining a better understanding of this matter, it is useful to import the following classic result:

Lemma 2.11 (Kronecker-Weyl)**.** *Let $x \in [0,1)^d$ be a point in the unit-cube so that $1, x_1, \ldots, x_d$ are linearly independent over $\mathbb{Q}$. Then the sequence of points $(nx)_{n\in\mathbb{N}} \in [0,1)^d$ where each coordinate is understood* mod 1 *is uniformly distributed (and thus in particular dense) in $[0,1)^d$.*

With this Lemma, we can now show that a necessary and 'generically' also sufficient condition for a pure state to ever evolve to an orthogonal state is that its overlap with any of the eigenvectors of the Hamiltonian is not larger than 1/2:

Theorem 2.12 (Condition for reaching minimal overlap)**.** *Let $\dim(\mathcal{H}) < \infty$, $H \in \mathcal{B}(\mathcal{H})$ Hermitian with an orthonormal basis of eigenvectors $\{\varphi_i\}$ and corresponding eigenvalues $\{\lambda_i\}$. For any $\psi \in \mathcal{H}$ define $p := \max_i |\langle\psi, \varphi_i\rangle|^2$ and $\nu := \inf_{t\in\mathbb{R}_+} \{|\langle\psi, e^{-iHt}\psi\rangle|\}$. Then*

$$\nu \geq \max\{0, 2p-1\}, \tag{2.14}$$

with equality if the eigenvalues $\{\lambda_i\}$ (as a multiset) are linearly independent over $\mathbb{Q}$.

Proof. Using the spectral decomposition of H we can write $|\langle\psi, e^{-iHt}\psi\rangle| = \left|\sum_k p_k e^{-i\lambda_k t}\right|$ where $p_k := |\langle\psi, \varphi_k\rangle|^2$. Since the p_k's are positive and sum up to one, this is a convex combination (i.e. a weighted average) of complex numbers of modulus one. Assume w.l.o.g. that $p = p_1$. From the triangle-inequality and using $\sum_{k>1} p_k = 1 - p$ we obtain

$$\left|\sum_k p_k e^{-i\lambda_k t}\right| \;\geq\; p - \left|\sum_{k>1} p_k e^{i(\lambda_1-\lambda_k)t}\right| \;\geq\; p - \sum_{k>1} p_k = 2p - 1, \tag{2.15}$$

thus proving the inequality in Eq.(2.14).

For proving that equality holds if the λ_k's are independent over $\mathbb{Q}$ we want to use Lemma 2.11. To this end, note that $\{t\lambda_k/2\pi\}_k \cup \{1\}$ are linearly independent iff $\{\lambda_k/2\pi\}_k \cup \{1/t\}$ is. The latter can, however, always be achieved by a suitable choice of $t > 0$ if only the λ_k's are linearly independent: since $\mathbb{R}$ is infinite-dimensional over $\mathbb{Q}$ we can always find a $t > 0$ s.t. $1/t$ is linearly independent of the λ_k's. Consequently, Lemma 2.11 implies that for any $\alpha \in [0, 2\pi)^d$ with $d := \dim(\mathcal{H})$ there is a sequence with elements $t_n \in \mathbb{R}_+$ s.t. $\exp[-it_n\lambda_k] \to \exp[i\alpha_k]$ for all $k \in \{1, \ldots, d\}$. It remains to show that there exists an α so that $\sum_k p_k \exp[i\alpha_k] = \max\{0, 2p-1\}$. If $p \geq 1/2$, a solution is given by $\alpha_1 = 0$ and $\alpha_k = \pi$ for all $k \geq 2$. So consider the case $p < 1/2$. First, we partition $\{1, \ldots, d\} = A_1 \cup A_2 \cup A_3$ into disjoint subsets that are chosen so that $p_{A_i} := \sum_{k \in A_i} p_k$ are all three smaller than $1/2$. For k in A_1, A_2, A_3 we set α_k equal to $0, \beta$ and γ, respectively. β and γ then have to be chosen so that

$$p_{A_1} + p_{A_2} e^{i\beta} = p_{A_3} e^{-i\gamma}.$$

To see that this is feasible, regard the two sides of this equation as parametrizations of two circles in the complex plane (when varying $\beta, \gamma \in [0, 2\pi)$). The circles have radii p_{A_2} and p_{A_3} and their centers are p_{A_1} apart. Assuming w.l.o.g. that p_{A_1} is the largest of the three weights, we see that the circles always intersect, i.e. there is always a solution, since $p_{A_2} + p_{A_3} = 1 - p_{A_1} \geq p_{A_1}$ as $p_{A_1} \leq 1/2$. □

Corollary 2.13 (Condition for evolving to an orthogonal state)**.** $\dim(\mathcal{H}) < \infty$. *Except for a null set in the set of Hermitian operators $H \in \mathcal{B}(\mathcal{H})$, every such H satisfies the following: for any unit vector $\psi \in \mathcal{H}$ the evolved state $\psi(t) := \exp[-iHt]\psi$ eventually satisfies $\inf_{t \in \mathbb{R}_+} |\langle \psi, \psi(t) \rangle| = 0$ iff the maximal overlap $\max_i |\langle \psi, \varphi_i \rangle|^2$ with any of the normalized eigenvectors φ_i of H is at most $1/2$.*

Proof. In order to be able to use Thm.2.12, we have to exclude any H whose eigenvalues $\{\lambda_i\}$ are linearly dependent over $\mathbb{Q}$. With $d := \dim(\mathcal{H})$ this means that there is a $q \in \mathbb{N}^d \setminus \{0\}$ so that $\langle q, \lambda \rangle = 0$. For a fixed q this equation determines a hyperplane $\mathcal{S}_q := \{\lambda \in \mathbb{R}^d | \langle q, \lambda \rangle = 0\}$ that has Lebesgue measure $\mu(\mathcal{S}_q) = 0$. Since the union of all these hyperplanes is countable, we still have $\mu\big(\cup_q \mathcal{S}_q\big) = 0$. Since this is equally true in any basis, the set of Hamiltonians to exclude forms a null set. For the remaining ones, Thm.2.12 proves the claim. □

Exercise 2.1 (Commutator identity for 2×2 matrices). Show that for any $A, B, C \in \mathbb{C}^{2\times 2}$ the relation $\big[[A, B]^2, C\big] = 0$ holds.

Exercise 2.2 (Uncertainty relations). Let $H_1, H_2 \in \mathcal{B}(\mathcal{H})$ be Hermitian, $\rho \in \mathcal{B}_1(\mathcal{H})$ a density operator and $A_i := H_i - \operatorname{tr}[\rho H_i]\,\mathbb{1}$.

(a) Express the inequality $\operatorname{tr}[\rho BB^*] \geq 0$ as an uncertainty relation for ρ, H_1, H_2 by inserting $B := A_1 + i\gamma A_2$ and optimizing over all $\gamma \in \mathbb{R}$.

(b) Apply the derived uncertainty relation for $\mathcal{H} \simeq \mathbb{C}^2$ to a pair of Pauli matrices. Identify 'minimal uncertainty states' that achieve equality in this uncertainty relation. Where are they located in the Bloch ball?

(c) Which uncertainty relation is obtained when optimizing over all $\gamma \in \mathbb{C}$?

Exercise 2.3 (Canonical commutation relation). Let Q, P be operators on a Hilbert space $\mathcal{H}$ that satisfy the 'canonical commutation relation' $[Q, P] = i\mathbb{1}$.

(a) Show that necessarily $\dim(H) = \infty$ and that Q, P cannot be Hilbert-Schmidt class operators.

(b) Prove that for any $n \in \mathbb{N}$: $[Q^n, P] = inQ^{n-1}$.

(c) Use (b) to show that Q and P cannot both be bounded operators.

Exercise 2.4 (Tensor-power trick). We write $A^{\otimes n} := A \otimes \ldots \otimes A$ for the n-fold tenor product of A.

(a) Let $A, B \in \mathcal{B}(\mathcal{H})$ be Hermitian, A invertible and $A \geq \pm B$ (meaning that the inequality holds for both signs). Show that $A \geq 0$ and $A^{\otimes n} \geq \pm B^{\otimes n}$ for all $n \in \mathbb{N}$.

(b) Show that for $\mathcal{H} \simeq \mathbb{C}^d$ there is a $\psi \in \mathcal{H}^{\otimes d}$ so that for any $A \in \mathcal{B}(\mathcal{H})$: $\det(A) = \langle \psi, A^{\otimes d}\psi \rangle$.

(c) Use (a) and (b) to prove that in Eq.(2.1) from Robertson's uncertainty relation $V \geq i\sigma$ implies $\det(V) \geq \det(\sigma)$.

Exercise 2.5 (Quantum error correction).

(a) Why are Pauli matrices used in the definition of an $[[n, k, d]]$-QECC? What if an 'error' occurs that is not described by one of the three Pauli matrices?

(b) Assume you have encoded k qubits into n qubits using an $[[n, k, d]]$ quantum error correcting code. Unfortunately, $d - 1$ of the qubits were completely destroyed (a cat jumped out of a box and knocked over this part of the experiment). The good news is that the remaining qubits are perfectly intact. Show that and how you can perfectly recover the state of the original k qubits.

Exercise 2.6 (Time-energy uncertainty relation). (a) Formulate and prove the Mandelstam-Tamm uncertainty relation for mixed states.

(b) Consider a finite-dimensional Hamiltonian that satisfies $0 \leq H \leq \mathbb{1}$ and that governs the time evolution of a pure state via $\psi(t) = \exp[-iHt]\psi$. Let t_0 be the first time so that $\langle \psi, \psi(t) \rangle = 0$. Provide a lower bound on t_0 that is as good as possible and that does not depend on ψ.

Notes and literature The origin of uncertainty relations in quantum theory is Heisenberg's 1927 paper [38], in which the relation now bearing his name appears very vaguely as '$q_1 p_1 \sim h$'. Soon after, it was mathematically formalized and proved by Kennard [39], and then generalized to multiple observables (as in Thm.2.1) by Robertson [40].
The first Fourier-analytic proof of the uncertainty relation is sometimes attributed to Wiener (cf. [41]), who apparently discussed it during a seminar in Göttingen in 1924. A simple but elegant generalization thereof is due to Wigderson[2][42]. They point out specifically that if $H \in \mathbb{C}^{d\times d}$ is a complex Hadamard matrix (i.e., $|H_{ij}| = 1$ and $H^*H = n\mathbb{1}$ as fulfilled e.g. for a rescaled Fourier transform matrix), then $\forall v \in \mathbb{C}^n \setminus \{0\}$:

$$\Delta(v)\Delta(Hv) \geq n, \qquad \text{where} \quad \Delta(v) := \frac{\|v\|_1}{\|v\|_\infty}. \tag{2.16}$$

Exploiting further that the number $|\mathrm{supp}(v)|$ of non-zero entries of v is an upper bound on $\Delta(v) \leq |\mathrm{supp}(v)|$, Eq.(2.16) becomes the signal processing uncertainty relation of Donoho and Stark [43].
An overview on the vast literature on quantum speed limits and time-energy uncertainty relations can be found in the review article [44]. The Mandelstam-Tamm inequality goes

back to [45]. The Margolus-Levitin bound can be found in [46] and its generalization in the form of Thm.2.10 together with Lemma 2.9 in [47].

2.4 Energy–entropy

Gibbs states

Definition 2.14 (Gibbs hypothesis/state)**.** *A self-adjoint operator H on $\mathcal{H}$ is said to satisfy the* Gibbs-hypothesis *if for all $\beta \in (0,\infty)$ the operator $\exp[-\beta H]$ is trace-class and the state*

$$\rho_\beta := \frac{e^{-\beta H}}{Z}, \quad \textit{with} \quad Z := \operatorname{tr}\left[e^{-\beta H}\right] \tag{2.17}$$

has finite entropy. ρ_β is then called the Gibbs state *and Z the* canonical partition function*.*

Remarks: 1. The Gibbs-hypothesis requires that H is bounded from below, has pure point spectrum (i.e., every element of the spectrum is an eigenvalue) and finite degeneracies. Moreover, if $\dim(\mathcal{H}) = \infty$, H needs to be unbounded.

2. The parameter β is called *inverse temperature*. More precisely, if k is Boltzmann's constant (which is often made equal to one by a suitable choice of units) and T the temperature, then $\beta = 1/(kT)$. If $\dim(\mathcal{H}) = \infty$, the range of admissible β in Eq.(2.17) is $(0,\infty]$, where in the limiting case $\beta \to \infty$, the Gibbs state becomes the *ground state*, i.e., the normalized projector onto the eigenspace of the smallest eigenvalue of H. If $\dim(\mathcal{H}) < \infty$, we can allow $\beta \in [-\infty,\infty]$, where negative β means negative temperature for which higher energy levels in the Gibbs state are more populated than lower ones.

3. The energy $\operatorname{tr}[H\rho_\beta] = -\partial_\beta \ln Z_\beta$ is a decreasing function of β. In fact, $\partial_\beta \operatorname{tr}[H\rho_\beta] = -\operatorname{var}(H)$, where the variance is evaluated w.r.t. ρ_β. By varying β in the admissible range, $\operatorname{tr}[H\rho_\beta]$ takes on every value in the convex hull of the spectrum of H. That is, for every $E \in \operatorname{conv}\big(\operatorname{spec}(H)\big)$ there is a $\beta \in (0,\infty]$ (resp. $\beta \in [-\infty,\infty]$ if $\dim(H) < \infty$), s.t. $\operatorname{tr}[H\rho_\beta] = E$.

Example 2.3 (Harmonic oscillator)**.** The paradigm of a Hamiltonian that satisfies the Gibbs hypothesis in infinite dimensions is the *harmonic oscillator* Hamiltonian. Under the convention that the ground state energy is equal to 0, this can formally be written as $H := \omega \sum_{n=0}^{\infty} n|n\rangle\langle n|$, where $\omega > 0$ is the frequency of the oscillator. Taking the trace w.r.t. the eigenbasis $\{|n\rangle\}$, we can compute the canonical partition function via a geometric series and obtain

$$Z = \frac{1}{1-e^{-\beta\omega}}, \quad \text{and} \quad E = \frac{\omega}{e^{\beta\omega}-1}. \tag{2.18}$$

The entropy of a quantum harmonic oscillator Gibbs state is most easily expressed in terms of its energy $E = \operatorname{tr}[H\rho_\beta]$ as $S(\rho_\beta) = (E+1)\ln(E+1) - E\ln E$.

Theorem 2.15 (Free energy theorem)**.** *Let H be a self-adjoint operator on $\mathcal{H}$ that satisfies the Gibbs hypothesis and ρ_β a corresponding Gibbs state at inverse*

temperature $\beta \neq 0$. *The* free energy $F(\rho) := \mathrm{tr}\,[H\rho] - \beta^{-1}S(\rho)$ *of any density operator* $\rho \in \mathcal{B}(\mathcal{H})$ *of finite entropy then satisfies*

$$F(\rho) = F(\rho_\beta) + \tfrac{1}{\beta}S(\rho\|\rho_\beta) \geq F(\rho_\beta). \tag{2.19}$$

Moreover, $F(\rho_\beta) = -\frac{1}{\beta}\ln Z$.

Proof. Inserting the definition of the Gibbs state into the relative entropy we obtain

$$\beta^{-1}S(\rho\|\rho_\beta) = \underbrace{\mathrm{tr}\,[H\rho] + \beta^{-1}\mathrm{tr}\,[\rho\ln\rho]}_{=F(\rho)} + \beta^{-1}\ln Z. \tag{2.20}$$

Since this vanishes iff $\rho = \rho_\beta$, we get that $\beta^{-1}\ln Z = -F(\rho_\beta)$. Finally, the inequality follows from the positivity of the relative entropy. □

Corollary 2.16 (Gibbs' variational principle)**.** *For every density operator* ρ *of finite entropy:*

$$S(\rho) = \min\left\{\mathrm{tr}\,[\rho H] + \ln\mathrm{tr}\left[e^{-H}\right] \;\middle|\; H \geq 0\right\} \tag{2.21}$$

Proof. Setting $\beta = 1$, we can restrict the set of Hamiltonians over which we minimize to those for which the Gibbs state ρ_β w.r.t. H exists. The inequality

$$S(\rho) \leq \mathrm{tr}\,[\rho H] + \ln\mathrm{tr}\left[e^{-H}\right] \tag{2.22}$$

is then nothing but $-F(\rho) \leq \ln\mathrm{tr}\left[e^{-H}\right] = -F(\rho_\beta)$, which holds according to Eq.(2.19) with equality iff $\rho = \rho_\beta$. The latter is satisfied for $H = -\ln\rho$. □

Corollary 2.17 (Maximum entropy/minimum energy)**.** *Let* H *be a self-adjoint operator on* $\mathcal{H}$ *that satisfies the Gibbs hypothesis and* ρ_β *a corresponding Gibbs state at inverse temperature* $\beta > 0$.

1. Maximum entropy property*: If* $\rho \in \mathcal{B}(\mathcal{H})$ *is any density operator with the same or less energy, i.e.,* $\mathrm{tr}\,[H\rho] \leq \mathrm{tr}\,[H\rho_\beta]$*, then*

$$S(\rho) \leq S(\rho_\beta). \tag{2.23}$$

2. Minimum energy property*: If* $\rho \in \mathcal{B}(\mathcal{H})$ *is any density operator with equal or larger entropy, i.e.,* $S(\rho) \geq S(\rho_\beta)$*, then*

$$\mathrm{tr}\,[H\rho] \geq \mathrm{tr}\,[H\rho_\beta]\,. \tag{2.24}$$

Under the stated assumptions, equality holds in Eqs.(2.23,2.24) iff $\rho = \rho_\beta$.[2]

[2]This also implies that energy and entropy of a Gibbs state have to be strictly monotonic functions of each other. In fact, if we denote by $S(E)$ the entropy of the Gibbs state with energy E, then $\partial_E S(E) = \beta(E) > 0$ (cf. [48], Appendix A).

Proof. This follows directly from the free energy theorem since we can rewrite Eq.(2.19) as

$$\big(\operatorname{tr}[H\rho] - \operatorname{tr}[H\rho_\beta]\big) + \frac{1}{\beta}\big(S(\rho_\beta) - S(\rho)\big) = \frac{1}{\beta} S(\rho\|\rho_\beta).$$

The fact that the relative entropy is positive, and zero iff its arguments coincide, then completes the proof. □

A consequence of this discussion (specifically of the identity in Eq.(2.20)) is that every density operator that has finite energy (w.r.t. a Hamiltonian that satisfies the Gibbs hypothesis) also has finite entropy. Moreover, the assumption of finite energy restores the continuity of the von Neumann entropy. A qualitative argument for that would be to note that the relative entropy is lower semicontinuous. By Eq.(2.20) this implies that, under an energy constraint, the von Neumann entropy is upper semicontinuous. Since it is also lower semicontinuous in general, it becomes continuous. The following uses a different argument towards a quantitative bound:

Theorem 2.18 (Continuity of von Neumann entropy)**.** *Let $H \geq 0$ satisfy the Gibbs hypothesis and ρ_1, ρ_2 be two density operators with energies $\operatorname{tr}[H\rho_1]$, $\operatorname{tr}[H\rho_2]$ not exceeding $E < \infty$. If ρ_β is the Gibbs state of H with energy E/ϵ, where $\epsilon := \frac{1}{2}\|\rho_1 - \rho_2\|_1$, then*

$$\big|S(\rho_1) - S(\rho_2)\big| \leq 2\epsilon S(\rho_\beta) + h(\epsilon), \tag{2.25}$$

where $h(\epsilon) := -\epsilon \ln \epsilon - (1-\epsilon)\ln(1-\epsilon)$.

Remarks: If $\dim(\mathcal{H}) =: d < \infty$, we can simply replace $S(\rho_\beta)$ by $\ln d$. In general, since the derivative of the entropy of a Gibbs state w.r.t. its energy E is equal to the inverse temperature $\beta = \beta(E)$, which in turn goes to zero for $E \to \infty$, we have that the r.h.s. of Eq.(2.25) goes to zero for $\epsilon \to 0$.

Proof. We use the quantum coupling of ρ_1 and ρ_2 from Lemma 1.44 followed by the Araki-Lieb triangle inequality (see Exercise 1.30 (c)) and the concavity upper bound for the von Neumann entropy stated in Prop.1.29 to obtain:

$$\begin{aligned} \big|S(\rho_1) - S(\rho_2)\big| &\leq S(\rho) \\ &\leq \epsilon S(\sigma_1 \otimes \sigma_2) + h(\epsilon). \end{aligned} \tag{2.26}$$

If $\dim(\mathcal{H}) =: d < \infty$, we can bound $S(\sigma_1 \otimes \sigma_2) \leq 2\ln d$. In the infinite-dimensional case, consider the Hamiltonian $H' := H \otimes \mathbb{1} + \mathbb{1} \otimes H$. Since $\operatorname{tr}[H'\rho] = \operatorname{tr}[H\rho_1] + \operatorname{tr}[H\rho_2] \leq 2E$, we have that $\operatorname{tr}[H'(\sigma_1 \otimes \sigma_2)] \leq 2E/\epsilon$. Hence, by the maximum entropy property, the entropy $S(\sigma_1 \otimes \sigma_2)$ is at most that of the Gibbs state of H' with energy $2E/\epsilon$. However, this Gibbs state is $\rho_\beta \otimes \rho_\beta$. □

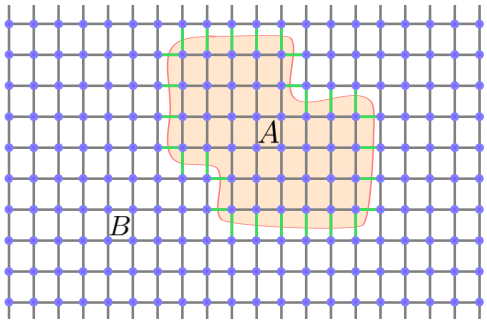

Figure 2.1: *Area law for mututal information.* For any non-zero temperature, consider a Gibbs state of a system whose interaction pattern is given by a graph (e.g. a square lattice). The mutual information between any region A and its complement B is bounded from above by a term linear in the number $|\partial A|$ of edges (which are depicted in green) that connect A and B.

Area law for correlations Another simple but insightful consequence of the free energy theorem is an *area law* for the mutual information. Consider a many-body quantum system whose interaction geometry is described by a graph (V, E). Specifically, we assume a finite-dimensional Hilbert space $\mathcal{H} = \otimes_{i\in V}\mathcal{H}_i$ and a Hamiltonian $H = \sum_{(i,j)\in E} h_{ij}$, where h_{ij} is supposed to act non-trivially only on $\mathcal{H}_i \otimes \mathcal{H}_j$ and as the identity operator on all other tensor factors. That is, the particles that are associated with the vertices of the graph interact pairwise along the edges. Let $A \subseteq V$ with complement $B := V \setminus A$ be any subsystem and $\partial A := \{(a, b) \in E | a \in A, b \in B\}$ the set of edges crossing its boundary. Within this framework, the following holds for the correlations of the system in equilibrium:

Theorem 2.19 (Area law)**.** *For any $\beta > 0$, the Gibbs state ρ_β w.r.t. H has mutual information bounded by*

$$I(A : B) \leq c\beta|\partial A|, \quad \text{with } c := 2 \max_{(i,j)\in\partial A} \|h_{ij}\|_\infty \ . \tag{2.27}$$

Proof. Let ρ_A, ρ_B be the reduced density operators of ρ_β w.r.t. subsystems A and B, respectively. The free energy, according to Thm.2.15, satisfies $F(\rho_\beta) \leq$

$F(\rho_A \otimes \rho_B)$. This can be rewritten as

$$\begin{aligned} \underbrace{S(\rho_A \otimes \rho_B) - S(\rho_\beta)}_{=I(A:B)} &\leq \beta \operatorname{tr}\left[H(\rho_A \otimes \rho_B - \rho_\beta)\right] \\ &= \beta \sum_{(i,j)\in E} \operatorname{tr}\left[h_{ij}(\rho_A \otimes \rho_B - \rho_\beta)\right] \\ &= \beta \sum_{(i,j)\in \partial A} \operatorname{tr}\left[h_{ij}(\rho_A \otimes \rho_B - \rho_\beta)\right] \qquad (2.28) \\ &\leq \beta \sum_{(i,j)\in \partial A} 2\|h_{ij}\|_\infty \leq c\beta|\partial A|. \qquad (2.29) \end{aligned}$$

In Eq.(2.28), we have used that for all $(i,j) \notin \partial A$ the reduced density operators of $\rho_A \otimes \rho_B$ and ρ_β are identical. The last inequality in Eq.(2.29) is an application of Hölder's inequality. □

In other words, the mutual information, i.e., the amount of information about B contained in A (and vice versa), scales at most like the surface area of the region A and not like its volume, which could be the case a priori. For this, it is necessary to assume the Gibbs state at a *non-zero* temperature (see Notes and literature).

Stability and passivity In the following, we are going to de- and reconstruct Gibbs states by examining increasingly strong properties together with the states that fulfill them: states that (i) commute with the Hamiltonian H, (ii) are a function of H, (iii) are a decreasing function of H, (iv) are Gibbs states.

First, we consider states that are a function of the Hamiltonian. Clearly, not every function will give rise to a density operator, and, especially in the infinite-dimensional case, we need to restrict the set of functions accordingly. Under such a condition, the following proposition shows that states that are functions of a Hamiltonian are *dynamically stable*. That is, if we let them evolve according to a slightly perturbed Hamiltonian, they will not move far even on arbitrarily long time scales.

Proposition 2.20 (Dynamical stability)**.** *Let H be a Hamiltonian satisfying the Gibbs hypothesis and $f : \mathbb{R} \to \mathbb{R}$ a positive function that is bounded by a decreasing exponential, i.e., $\exists a, b > 0\ \forall x : 0 \leq f(x) \leq ae^{-bx}$. For any self-adjoint, H-bounded[3] operator V, and sufficiently small $\lambda \geq 0$, define density operators $\rho_\lambda := f(H + \lambda V)/\operatorname{tr}\left[f(H + \lambda V)\right]$ as well as $\rho := \rho_0$ and its perturbed evolution $\rho(t, \lambda) := e^{-i(H+\lambda V)t} \rho\, e^{i(H+\lambda V)t}$. Then*

$$\sup_{t\geq 0} \|\rho - \rho(t,\lambda)\|_1 \leq 2\|\rho - \rho_\lambda\|_1 \overset{\lambda\to 0}{\longrightarrow} 0. \qquad (2.30)$$

[3] V is H-bounded if there are constants c, d s.t. $|\langle\psi, V\psi\rangle| \leq c\langle\psi, H\psi\rangle + d\|\psi\|^2$ holds for all ψ in the domain of H.

Proof. (sketch, see [49], Appendix A for details) With $U := e^{i(H+\lambda V)t}$ we can write $\rho(t,\lambda) = U^*\rho U$. Using that $U^*\rho_\lambda U = \rho_\lambda$ together with the triangle inequality for the norm, we obtain

$$\begin{aligned}\|\rho - U^*\rho U\|_1 &\leq \|\rho - \rho_\lambda\|_1 + \|\rho_\lambda - U^*\rho U\|_1 \\ &\leq \|\rho - \rho_\lambda\|_1 + \|U^*(\rho_\lambda - \rho)U\|_1 = 2\|\rho - \rho_\lambda\|_1.\end{aligned}$$

Under the assumptions, it can be shown that $\rho_\lambda \to \rho$ weakly and thus, by Thm.1.7, in trace-norm. □

Definition 2.21 (Passivity)**.** *A state ρ is called* passive *w.r.t. a Hamiltonian H if for all unitaries U:*

$$\operatorname{tr}[\rho H] \leq \operatorname{tr}[U^*\rho U H]\,. \tag{2.31}$$

Proposition 2.22 (Characterization of passivity)**.** *Let ρ be a density operator and H a Hamiltonian, both acting on a finite-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$.*

(i) The functional $E : U(d) \to \mathbb{R}$, $E(U) := \operatorname{tr}[U^*\rho U H]$ has a stationary point at U iff $[U^*\rho U, H] = 0$.

(ii) ρ is passive w.r.t. H iff there is an orthonormal basis $\{|k\rangle\}_{k=1}^d$ s.t. $\rho = \sum_k p_k |k\rangle\langle k|$, $H = \sum_k E_k |k\rangle\langle k|$ and for all $k, l \in \{1, \ldots, d\}$:

$$E_k < E_l \;\Rightarrow\; p_k \geq p_l. \tag{2.32}$$

Proof. (i): We parameterize a path through U by $t \mapsto U(t) := Ue^{iAt}$ for any Hermitian A. In this way, we cover the entire tangent space, so that U is a stationary point of E if for every such A:

$$\begin{aligned}0 \stackrel{!}{=} \frac{d}{dt}\Big|_{t=0} E\big(U(t)\big) &= i\operatorname{tr}[U^*\rho U A H] - i\operatorname{tr}[A U^*\rho U H] \\ &= i\operatorname{tr}[[H, U^*\rho U]A]\,.\end{aligned} \tag{2.33}$$

A vanishing commutator is clearly sufficient for stationarity. However, it is also necessary since we can choose $A = i[H, U^*\rho U]$ in which case Eq.(2.33) becomes $\|[H, U^*\rho U]\|_2^2$, which is zero only if the commutator vanishes.

(ii): As passivity requires stationarity of E, we can exploit *(i)* according to which there is a common basis of eigenvectors. So we can assume that $\rho = \sum_k p_k |k\rangle\langle k|$, and $H = \sum_k E_k |k\rangle\langle k|$. Then

$$E(U) = \sum_{k,l} E_l M_{kl} p_k, \quad \text{where} \quad M_{kl} := |\langle k|U|l\rangle|^2 \tag{2.34}$$

is a doubly stochastic matrix (see example 1.13). By Birkhoff's theorem, every doubly stochastic matrix is a convex combination of permutations. This implies that

$$\min_{U \in U(d)} E(U) = \min_{\pi \in S_d} \sum_{l=1}^{d} E_l \; p_{\pi(l)}, \tag{2.35}$$

where we have equality (rather than just '$\geq$') since every permutation matrix is in particular a unitary. Hence, characterizing passivity boils down to characterizing the permutations that yield the minimum in Eq.(2.35). If the implication of Eq.(2.32) is violated, the energy can be decreased by interchanging p_k and p_l. So Eq.(2.32) is clearly necessary for passivity. However, when applied to all k, l it is also sufficient since it determines a minimizing π for Eq.(2.35). This is unique unless there are degeneracies of the form $E_k = E_l$. In this case, however, interchanging $p_k \leftrightarrow p_l$ does not change the energy. Therefore, every permutation that is compatible with Eq.(2.32) leads to a passive state. □

If $H \in \mathcal{B}(\mathbb{C}^d)$ is non-degenerate, then ρ is passive w.r.t. H iff there is a non-increasing function $f : \mathbb{R} \to \mathbb{R}$ s.t.

$$\rho = \frac{f(H)}{\mathrm{tr}\,[f(H)]}. \tag{2.36}$$

Clearly, $f(x) = e^{-\beta x}$ for some $\beta \geq 0$ is one such function. In the degenerate case, there are passive states that are not of this form since $E_k = E_l$ does not require $p_k = p_l$ for a passive state. However, passive states that are not of the form in Eq.(2.36) are not *structurally stable* in the sense that a tiny perturbation of the Hamiltonian can lead to a state that is far from any passive state.

In order to get a better understanding of the meaning of passivity and also to ultimately close the circle and return the focus to the set of Gibbs states, we will introduce a graphical representation. To this end, we only consider states that have the same eigenbasis as a given Hamiltonian and assign to every common eigenvector a point $(E_k, \ln(1/p_k))$ in the extended plane $R^2 := \mathbb{R}^2 \cup \mathbb{R} \times \{\infty\}$. That is, the Hamiltonian-state pair is then assigned a set of points

$$V(H, \rho) := \{(E_k, \ln(1/p_k))\}_{k=1}^d \subset R^2. \tag{2.37}$$

The reason for using $\ln(1/p_k)$ instead of p_k in the second argument is that the former will be additive under taking multiple copies.

Let us equip R^2 with the following partial order: for $v_k := (x_k, y_k)$ and $v_l := (x_l, y_l)$ we will write

$$v_k \leq v_l \quad :\Longleftrightarrow \quad x_k \leq x_l \ \wedge \ y_k \leq y_l. \tag{2.38}$$

The relevance of this partial order in the present context stems from the fact that the condition for passivity in Eq.(2.32), when imposed for all k, l, is equivalent to $\forall v_k, v_l \in V(H, \rho) : v_k \leq v_l \vee v_l \leq v_k$. In other words, ρ is passive w.r.t. H iff the partial order becomes a total order on $V(H, \rho)$. For a graphical depiction see Fig.2.2.

Definition 2.23 (Complete passivity)**.** *A density operator ρ is said to be* completely passive *w.r.t. a Hamiltonian H if for every $n \in \mathbb{N} : \rho^{\otimes n}$ is passive w.r.t. the Hamiltonian $H^{(n)} := \sum_{i=1}^n H_i$, where each H_i denotes a tensor product operator that acts as H on the i'th tensor factor and as $\mathbb{1}$ on all the others.*

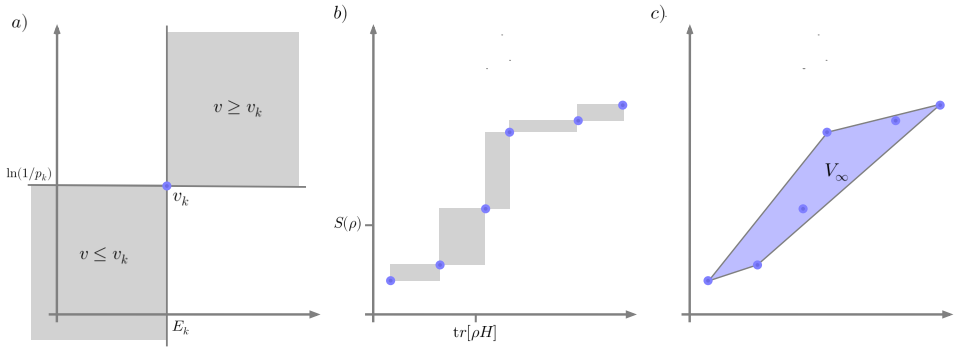

Figure 2.2: *Partial order and passivity.* a) The considered partial order defines every point in the northeast quadrant of v_k as larger, and everything southwest as smaller. b) A passive state is represented by a totally ordered set of points. In this case, the rectangular order-intervals form a non-decreasing staircase. The average of all points, averaged w.r.t. the probabilities p_k has coordinates $(\mathrm{tr}\,[\rho H]\,, S(\rho))$. c) The convex hull of the points shown in b) is the closure of V_n in the limit $n \to \infty$.

Theorem 2.24 (Complete passivity = Gibbs state)**.** *A finite-dimensional density operator ρ is completely passive w.r.t. a Hamiltonian H iff ρ is either a Gibbs state of H for some inverse temperature $\beta \geq 0$ or it is a ground state (i.e., $\mathrm{tr}\,[\rho H]$ is minimal).*

Proof. Let d be the Hilbert space dimension. In order to use the just discussed geometric characterization of passivity, we define

$$\begin{aligned} V_n &:= \frac{1}{n} V\left(H^{(n)}, \rho^{\otimes n}\right) \\ &= \left\{ \left(\frac{1}{n}\sum_{i=1}^{n} E_{k_i}, \frac{1}{n}\sum_{i=1}^{n} \ln(1/p_{k_i}) \right) \right\}_{k \in \{1,\ldots,d\}^n} \\ &= \left\{ \sum_{l=1}^{d} \lambda_l v_l \;\middle|\; v_l \in V_1,\ \sum_{l=1}^{d} \lambda_l = 1,\ \lambda_l \in \tfrac{1}{n}\mathbb{N}_0 \right\}. \end{aligned} \tag{2.39}$$

Then ρ is completely passive iff V_n is totally ordered for all $n \in \mathbb{N}$. Eq.(2.39) shows that in the limit $n \to \infty$ the set V_n becomes dense in the convex hull of V_1. The convex hull, in turn, is totally ordered iff it is contained in a line with positive (possibly infinite) slope β. Suppose first that this slope is finite. Then there is an offset $\ln Z$ so that for every $(E_k, \ln(1/p_k)) \in V_1$ we have $\ln(1/p_k) = \beta E_k + \ln Z$, which means

$$p_k = \frac{e^{-\beta E_k}}{Z}.$$

If the slope is infinite and the ground state energy level is degenerate, then the

corresponding weights p_k are not determined and any ground state is as good as any other. □

Landauer's principle c

Exercise 2.7 (Gibbs states with additional constraints). Let $H_1, \dots H_n$ be Hermitian operators on a finite-dimensional Hilbert space $\mathcal{H}$ and $\rho_\beta := \exp[-\sum_{i=1}^n \beta_i H_i]/Z$ for some $\beta \in \mathbb{R}^n$ and normalization constant Z s.t. $\operatorname{tr}[\rho_\beta] = 1$. Let ρ be any density operator on $\mathcal{H}$ s.t. $\operatorname{tr}[\rho H_i] = \operatorname{tr}[\rho_\beta H_i]$ holds for all i. Prove that $S(\rho) \leq S(\rho_\beta)$.

Exercise 2.8 (Passive states). Let $H = H^*$ be a Hamiltonian on a finite-dimensional Hilbert space $\mathcal{H}$.

(a) Suppose $\mathcal{H} \simeq \mathbb{C}^2$ and that H is not proportional to the identity. Show that w.r.t. H every passive state is a Gibbs state for some $\beta \geq 0$.

(b) Show that for every $r \in \{1, \dots, \dim(\mathcal{H})\}$ there is a state that is passive w.r.t. H and proportional to a rank r projection.

Exercise 2.9 (Active states). Given a Hamiltonian H on a finite-dimensional Hilbert space, let us call a density operator an *active state* w.r.t. H if no state in the unitary orbit $U\rho U^*$ has larger energy than ρ w.r.t. H.

(a) Provide a characterization of active states analogous to the one of passive states.

(b) Provide a characterization of all states that are *completely active* in the sense that for all $n \in \mathbb{N}$: $\rho^{\otimes n}$ is active w.r.t. $H^{(n)}$ (i.e., the n-fold independent action of H).

Exercise 2.10 (Conditional probabilities and area law for classical systems).

(a) Let $(p_{x,y})$ be a joint probability distribution of two classical, finite discrete random variables X and Y, and $S(X,Y)$ its entropy. Similarly, let $S(X)$ be the entropy of the corresponding marginal distribution $p_x := \sum_y p_{x,y}$. Show that $S(X,Y) \geq S(X)$.

(b) Within the setup of the area law for the mutual information, assume in addition that there is a product basis within which each of the interaction Hamiltonians h_{ij} is diagonal. Under this assumption, use (a) to prove an area law for the mutual information that does not depend on the inverse temperature β.

Exercise 2.11 (Landauer's bound).

(a) Estimate the energy (in units J=Ws) that has to be dissipated at room temperature at least according to Landauer's bound when erasing one bit of information by reinitializing it to '0'.

(b) Suppose a computer uses 100W of power and runs at a clock speed of 3GHz. How many bits could be reinitialized per clock cycle before violating Landauer's bound?

Notes and literature The inequality in Eq.(2.22) is also known as *Peierls-Bogolubov inequality*. Continuity of the von Neumann entropy for finite-dimensional systems was proven by Fannes [50] and improved by Audenaert [51]. The continuity-bound under an energy constraint of Thm.2.18 is due to Winter [31]. For the case of the reference Hamiltonian being the harmonic oscillator (discussed in example 2.3), the bound of Eq.(2.25) has been improved in [52] to

$$|S(\rho_1) - S(\rho_2)| \leq h(\epsilon) + Eh(\epsilon/E), \quad \forall \epsilon \in [0, E/(E+1)]. \tag{2.40}$$

Chapter 3

Statistical inference

Suppose an experiment produces an unknown quantum state ρ that is known to be described within a given set of density operators $\mathcal{Q} \subset \mathcal{B}(\mathcal{H})$, which is potentially equipped with an a priori probability distribution. How can we identify ρ within $\mathcal{Q}$? Depending on the size of $\mathcal{Q}$ this task runs under different names:

- *Hypothesis testing* if $\mathcal{Q}$ is a finite set.
- *Parameter estimation* if $\mathcal{Q}$ depends on one (or a few) real parameter(s).
- *Tomography* if the dimension of $\mathcal{Q}$ grows with the dimension of $\mathcal{H}$.

This distinction is obviously neither a sharp nor an exhaustive one, but it captures well the essential differences not only in terms of application contexts, but also in terms of the mathematical tools used.

3.1 Hypothesis testing

Optimality conditions Let $\rho_1, \ldots, \rho_m \in \mathcal{B}_1(\mathcal{H})$ be a finite number of density operators with associated probabilities $p_1, \ldots, p_m \geq 0$ satisfying $\sum_{x=1}^m p_x = 1$. We regard each ρ_x as one of m different *hypotheses* with corresponding a priori probability p_x. Suppose our goal is to discriminate among those hypotheses by means of a measurement that is described by a POVM $M = (M_x \in \mathcal{B}(\mathcal{H}))_{x=1}^m$ and to be chosen. Specifically, we aim at maximizing the average probability of identifying the hypothesis correctly. That is, we are interested in maximizing

$$\mathcal{P}(M) := \sum_{x=1}^m \mathrm{tr}\Big[M_x \underbrace{p_x \rho_x}_{=:\tilde{\rho}_x}\Big] \tag{3.1}$$

over all POVMs M. For later use it is helpful to introduce $L := \sum_x M_x \tilde{\rho}_x$ and to note that $\mathcal{P}(M) = \mathrm{tr}\,[L]$. A measurement that maximizes Eq.(3.1) is sometimes called a *maximum likelihood measurement* (see Fig.3.1 for this naming). The following Lemma shows that such an optimal measurement always exists.

Lemma 3.1 (Existence of optimal measurements)**.** *The supremum of Eq.(3.1) taken over all POVMs $(M_x \in \mathcal{B}(\mathcal{H}))_{x=1}^m$ is attained. That is, there exists a measurement that achieves the maximum.*

Proof. Let $\mathcal{M}_m \subseteq \mathcal{B}(\mathcal{H})^m$ be the space of all m-outcome POVMs on $\mathcal{H}$ equipped with the m-fold product of the weak*-topology on $\mathcal{B}(\mathcal{H})$. As the unit ball, which contains each M_x is (sequentially) compact, and this property is preserved under products, $\mathcal{M}_m$ is compact as well. Since the functional $\mathcal{M}_m \ni M \mapsto \sum_x \mathrm{tr}\,[M_x \tilde{\rho}_x]$ is continuous w.r.t. the product weak*-topology, its supremum over the compact space $\mathcal{M}_m$ is a maximum. □

Theorem 3.2 (Optimality conditions)**.** *Let $\left(\tilde{\rho}_x \in \mathcal{B}_1(\mathcal{H})\right)_{x=1}^m$ be a finite sequence of Hermitian operators and $\mathcal{M}_m$ the set of all m-outcome POVMs on $\mathcal{H}$. Then for $M \in \mathcal{M}_m$ with $L := \sum_x M_x \tilde{\rho}_x$ the following are equivalent.*

(i) M is an optimal measurement in the sense that for $\mathcal{P}(M) := \mathrm{tr}\,[L]$:

$$\mathcal{P}(M) = \sup_{M' \in \mathcal{M}_m} \mathcal{P}(M'),$$

(ii) $\forall x: \ \frac{1}{2}(L + L^*) \geq \tilde{\rho}_x$, *(iii)* $\forall x: \ L \geq \tilde{\rho}_x$,

(iv) There is an operator $K \in \mathcal{B}(\mathcal{H})$ such that $\forall x$: $K \geq \tilde{\rho}_x$ and $(K - \tilde{\rho}_x)M_x = 0$,

(v) $\mathcal{P}(M) = \min\{\mathrm{tr}\,[A] \,|A \in \mathcal{A}\}$ with $\mathcal{A} := \{A \in \mathcal{B}_1(\mathcal{H})|\ \forall x : A \geq \tilde{\rho}_x\}$.

Moreover, if these conditions are satisfied, then K from (iv) is equal to L and equal to the unique minimizer in (v).

Remarks: 1. Note that the inequalities in (iii) and (iv) imply in particular that L and K are Hermitian operators. 2. While it is common and instructive to think of $\tilde{\rho}_x$ as states weighted with their a priori probabilities, in which case $\mathcal{P}(M)$ is the expression in Eq.(3.1), one may also use a more general perspective where $\tilde{\rho}_x = \sum_y \delta(x,y) p_y \rho_y$. Here, $\delta : [m] \times [m] \to \mathbb{R}$ is any function that one wishes to use in order to quantify the 'reward' of measuring x when the true hypothesis was y. Using the terminology of Bayesian inference or statistical learning theory, this amounts to choosing a different *loss function*. 3. Uniqueness of K does not necessarily imply uniqueness of the optimal measurement.

Proof. $(i) \Rightarrow (ii)$ is proven by contraposition: suppose one of the inequalities of (ii) is violated, say for $x = 1$. Then $L + L^* - 2\tilde{\rho}_1$ is not positive, so we know that the orthogonal projector onto its negative part is non-zero. Let us denote this projector by P and define

$$M'_x := (\mathbb{1} - \epsilon P) M_x (\mathbb{1} - \epsilon P) + \epsilon(2 - \epsilon) P \delta_{1,x}.$$

This defines a valid POVM for all $\epsilon \in [0, 2]$, which leads to

$$\mathcal{P}(M') = \mathcal{P}(M) + \epsilon \underbrace{\mathrm{tr}\,[P(2\tilde{\rho}_1 - L - L^*)]}_{>0} \underbrace{- \epsilon^2 \mathrm{tr}\,[\tilde{\rho}_1 P] + \epsilon^2 \sum_x \mathrm{tr}\,[P M_x P \tilde{\rho}_x]}_{=\mathcal{O}(\epsilon^2)}.$$

Hence, for sufficiently small $\epsilon > 0$ we get $\mathcal{P}(M') > \mathcal{P}(M)$.

$(ii) \Rightarrow (i)$ and $(iii) \Rightarrow (i)$ are proven in the same way: using that $\mathcal{P}(M) = \operatorname{tr}[L]$ and that $\sum_x M'_x = \mathbb{1}$ we obtain

$$\mathcal{P}(M) - \mathcal{P}(M') = \sum_x \operatorname{tr}\left[LM'_x - \tilde{\rho}_x M'_x\right] = \sum_x \operatorname{tr}\left[(L - \tilde{\rho}_x)M'_x\right]. \tag{3.2}$$

Under the assumption of (iii) $(L - \tilde{\rho}_x)$ is positive, which implies positivity of Eq.(3.2) and therefore (i). Since $\operatorname{tr}[L] = \operatorname{tr}[L^*]$, we can replace L by its Hermitian part $\frac{1}{2}(L + L^*)$ in Eq.(3.2) and in this way prove that also $(ii) \Rightarrow (i)$.

$(ii) \Leftrightarrow (iii)$ only requires to show that (ii) implies that L is Hermitian. In order to see this, we use again that $\operatorname{tr}[L] = \operatorname{tr}[L^*]$ so that

$$\sum_x \operatorname{tr}\Big[\underbrace{\Big(\frac{1}{2}(L + L^*) - \tilde{\rho}_x\Big)}_{\geq 0} M_x\Big] = \operatorname{tr}\left[\frac{1}{2}(L + L^*) - L\right] = 0. \tag{3.3}$$

Due to the positivity assumptions of (ii), the l.h.s. of Eq.(3.3) is a sum of positive terms. So this vanishes only if each term vanishes individually, which in turn implies that $\big(\frac{1}{2}(L + L^*) - \tilde{\rho}_x\big)M_x = 0$. Summing over x this becomes $\frac{1}{2}(L - L^*) = 0$ so that L is indeed Hermitian.

$(iii) \Rightarrow (iv)$ follows when choosing $K = L$ and revisiting the argument after Eq.(3.3) in order to see that $(L - \tilde{\rho}_x)M_x = 0$.

$(iv) \Rightarrow (i)$: from $KM_x = \tilde{\rho}_x M_x$ we obtain $\operatorname{tr}[K] = \operatorname{tr}[L]$ by summing over x and taking the trace. Hence, $\operatorname{tr}[K] = \mathcal{P}(M)$ so that

$$\begin{aligned} \mathcal{P}(M) - \mathcal{P}(M') &= \sum_x \operatorname{tr}\left[KM'_x\right] - \operatorname{tr}\left[\tilde{\rho}_x M'_x\right] \\ &= \sum_x \operatorname{tr}\Big[\underbrace{(K - \tilde{\rho}_x)}_{\geq 0} M'_x\Big] \geq 0. \end{aligned}$$

$(v) \Leftrightarrow (i)$: From the proof of the equivalence of (i) and (iv) we know that $\sup_{M'} \mathcal{P}(M') = \operatorname{tr}[K]$ with K as specified in (iv). Clearly, $K \in \mathcal{A}$ and using that $KM_x = \tilde{\rho}_x M_x$ we obtain for any $A \in \mathcal{A}$:

$$\operatorname{tr}[K] = \sum_x \operatorname{tr}[KM_x] = \sum_x \operatorname{tr}[\tilde{\rho}_x M_x] \leq \sum_x \operatorname{tr}[AM_x] = \operatorname{tr}[A].$$

Consequently, K is an element of minimal trace in $\mathcal{A}$, which allows us to rewrite (v) as $\mathcal{P}(M) = \operatorname{tr}[K] = \sup_{M'} \mathcal{P}(M')$ and we finally recover (i).

The proof of the uniqueness of the minimizer is outsourced to Exercise 1.9 □

Example 3.1 (Commuting states)**.** If $\tilde{\rho}_1, \ldots, \tilde{\rho}_m$ mutually commute, i.e., they are simultaneously diagonal in a basis, say $\{|i\rangle\}_{i=1}^{\dim(\mathcal{H})}$, then

$$\max_M \mathcal{P}(M) = \sum_i \max_x \langle i|\tilde{\rho}_x|i\rangle. \tag{3.4}$$

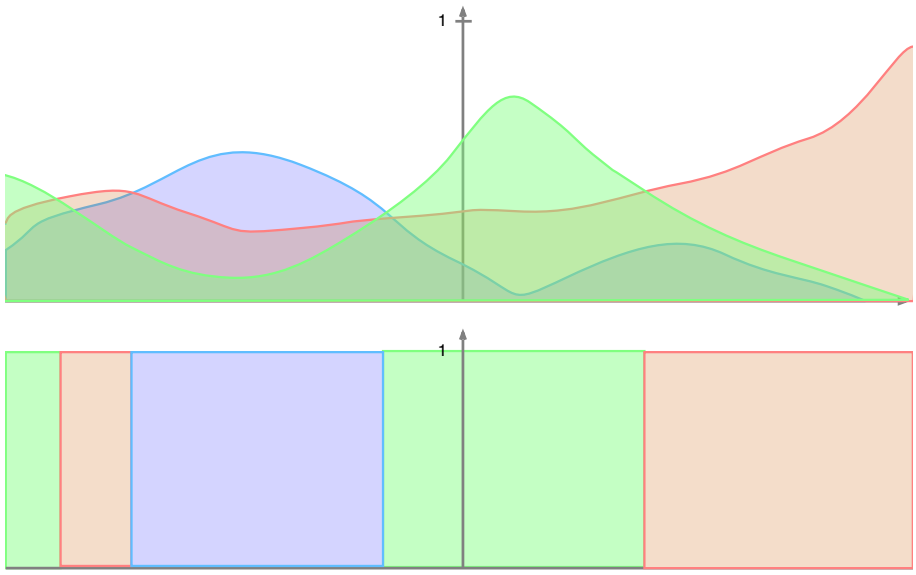

Figure 3.1: *Classical hypothesis testing.* Top: Depiction of (unnormalized) probability distributions of three hypotheses. Within the quantum formalism these are described by the diagonal elements of three diagonal unnormalized states $\tilde{\rho}_1, \tilde{\rho}_2$ and $\tilde{\rho}_3$ as discussed in Example 3.1. In this case, the maximum average success probability of correctly distinguishing the hypotheses is equal to the integral of the pointwise maximum of the distributions (see Eq.(3.4)). Bottom: the corresponding optimal measurement 'guesses' x wherever $\tilde{\rho}_x$ is maximal (see Exercise 3.4). Thus the name *maximum likelihood measurement*.

This follows from Thm.3.2 (v) since the diagonal operator K with diagonal entries $\langle i|K|i\rangle = \max_x \langle i|\tilde{\rho}_x|i\rangle$ by construction has minimal trace among all operators that are larger than each $\tilde{\rho}_x$. For the structure of maximum likelihood measurements in this case, see Exercise 3.4.

Example 3.2 (Binary hypothesis testing)**.** Consider two hypotheses ρ_1, ρ_2 with assigned a priori probabilities $p_1, p_2 = (1 - p_1)$ and define by P the orthogonal projector onto the positive part $(\tilde{\rho}_1 - \tilde{\rho}_2)_+$. We claim that the POVM $(M_1, M_2) := (P, \mathbb{1} - P)$ optimally distinguishes these hypotheses. Using (iii) in Thm.3.2 this is easily verified since $L = (\tilde{\rho}_1 - \tilde{\rho}_2)_+ + \tilde{\rho}_2$ is clearly larger than $\tilde{\rho}_2$ and also larger than $\tilde{\rho}_1$ since $L - \tilde{\rho}_1 = (\tilde{\rho}_1 - \tilde{\rho}_2)_+ - (\tilde{\rho}_1 - \tilde{\rho}_2) \geq 0$. The corresponding maximal average success probability is then

$$\begin{aligned} \mathcal{P}(M) &= \mathrm{tr}\,[L] = \mathrm{tr}\,[(\tilde{\rho}_1 - \tilde{\rho}_2)_+ + \tilde{\rho}_2] \\ &= \frac{1}{2}\Big(\|\tilde{\rho}_1 - \tilde{\rho}_2\|_1 + \underbrace{\mathrm{tr}\,[\tilde{\rho}_1 + \tilde{\rho}_2]}_{=1}\Big), \end{aligned} \tag{3.5}$$

where we have used that for any Hermitian A we have $\mathrm{tr}\,[A_+] = \frac{1}{2}(\|A\|_1 + \mathrm{tr}\,[A])$. Note that for $p_1 = p_2 = 1/2$ this recovers the achievability of the Lipschitz bound of Eq.(1.24). Eq.(3.5) can be regarded as a quantum version of the classical *Neyman-Pearson-Lemma*.

Example 3.3 (Uniformly distributed pure states)**.** Suppose $\rho_1, \ldots, \rho_m$ is a set of pure state density operators on $\mathbb{C}^d$ with associated a priori probabilities $p_x = 1/m$ and such that $\sum_x p_x \rho_x = \mathbb{1}/d$ (which, by comparing the ranks of both sides, implies in particular that $m \geq d$). The operators $M_x := \frac{d}{m}\rho_x$ then form a POVM describing a maximum likelihood measurement. In order to see that this is an optimal measurement, we use that $\rho_x^2 = \rho_x$ to conclude that $L = d\sum_y p_y^2 \rho_y = \frac{1}{m}\mathbb{1} \geq \frac{1}{m}\rho_x$ for all x. Hence, condition (iii) of Thm.3.2 is fulfilled. The optimal average success probability of distinguishing these states correctly is then $\mathcal{P}(M) = \mathrm{tr}\,[L] = d/m$, which is the general maximum derived in Prop.1.12.

Pretty good bounds While there is in general no closed-form expression for optimal measurements, there are some explicit common choices of measurements that perform 'pretty well' — in a sense that we will have to specify. For finite-dimensional Hilbert spaces, we will discuss two of these measurements that correspond to $\alpha \in \{1, 2\}$ in the family

$$M_x^{(\alpha)} := R^{-\frac{1}{2}}\, \tilde{\rho}_x^{\alpha}\, R^{-\frac{1}{2}} + \frac{1}{m}\mathbb{1}\big|_{\ker(R)}, \qquad R := \sum_x \tilde{\rho}_x^{\alpha}. \tag{3.6}$$

Here, $\mathbb{1}\big|_{\ker(R)}$ means the orthogonal projector onto the kernel of R and, in case R has a non-trivial kernel, $R^{-\frac{1}{2}}$ means the pseudo-inverse[1] of $R^{\frac{1}{2}}$. For $\alpha = 1$, $M^{(\alpha)}$ is known as *pretty good measurement* and for $\alpha = 2$ it is sometimes called *square measurement*.

We will now compare the average success probability as defined in Eq.(3.1) of these two types of measurements with the optimum

$$\mathcal{P}^{opt} := \max_{M \in \mathcal{M}_m(\mathcal{H})} \mathcal{P}(M), \tag{3.7}$$

that is attained by a, generally unknown, maximum likelihood measurement.

Proposition 3.3 (Square-measurement bounds)**.** *Let $\tilde{\rho}_1, \ldots, \tilde{\rho}_m$ be positive operators on $\mathbb{C}^d$ with $\sum_x \mathrm{tr}\,[\tilde{\rho}_x] = 1$ and $R := \sum_x \tilde{\rho}_x^2$. Then*

$$\left(\mathrm{tr}\left[R^{\frac{1}{2}}\right]\right)^2 \leq \mathcal{P}\big(M^{(2)}\big) \leq \mathcal{P}^{opt} \leq \mathrm{tr}\left[R^{\frac{1}{2}}\right]. \tag{3.8}$$

Proof. For the lower bound, we exploit the fact that the function $z \mapsto z^2$ is operator convex. This enables us to apply the operator Jensen's inequality [53],

[1] That is, we invert the operator on its range and keep the kernel.

which leads to

$$\begin{aligned}\Big(\mathrm{tr}\Big[R^{\frac{1}{2}}\Big]\Big)^2 &= \left(\sum_x \mathrm{tr}\Big[\tilde\rho_x\Big(\tilde\rho_x^{\frac12}R^{-\frac12}\tilde\rho_x^{\frac12}\Big)\Big]\right)^2 \le \sum_x \mathrm{tr}\Big[\tilde\rho_x\Big(\tilde\rho_x^{\frac12}R^{-\frac12}\tilde\rho_x^{\frac12}\Big)^2\Big] \\ &= \sum_x \mathrm{tr}\Big[\tilde\rho_x^2 R^{-\frac12}\tilde\rho_x R^{-\frac12}\Big] = \mathcal{P}\big(M^{(2)}\big)\,.\end{aligned}$$

For the upper bound we utilize that the function $z \mapsto z^{\frac12}$ is operator monotone. So $\tilde\rho_x^2 \le R$ implies that $\tilde\rho_x = \big(\tilde\rho_x^2\big)^{\frac12} \le R^{\frac12}$. An arbitrary POVM M (including the optimal one) then satisfies

$$\sum_x \mathrm{tr}\left[M_x\tilde\rho_x\right] \le \sum_x \mathrm{tr}\Big[M_x R^{\frac12}\Big] = \mathrm{tr}\Big[R^{\frac12}\Big]\,.$$

□

Proposition 3.4 (Pretty good measurement bound)**.** *Let $\tilde\rho_1,\ldots,\tilde\rho_m$ be positive operators on $\mathbb{C}^d$ and $R := \sum_x \tilde\rho_x$ with $\mathrm{tr}\,[R] = 1$. Then*

$$\big(\mathcal{P}^{opt}\big)^2 \le \mathcal{P}\big(M^{(1)}\big) \le \mathcal{P}^{opt}. \tag{3.9}$$

Proof. We begin the proof of the lower bound by applying the Cauchy-Schwarz inequality twice – once in the space $\mathcal{B}_2(\mathbb{C}^d)$ of Hilbert-Schmidt class operators and once in $\mathbb{R}^m$. In this way, we obtain for every POVM M:

$$\begin{aligned}\left(\sum_x \mathrm{tr}\left[M_x\tilde\rho_x\right]\right)^2 &= \left(\sum_x \mathrm{tr}\Big[\big(R^{\frac14}M_xR^{\frac14}\big)\big(R^{-\frac14}\tilde\rho_xR^{-\frac14}\big)\Big]\right)^2 \\ &\le \left(\sum_x \big\|\big(R^{\frac14}M_xR^{\frac14}\big\|_2\big\|R^{-\frac14}\tilde\rho_xR^{-\frac14}\big\|_2\right)^2 \\ &\le \left(\sum_x \big\|\big(R^{\frac14}M_xR^{\frac14}\big\|_2^2\right)\left(\sum_x \big\|\big(R^{-\frac14}\tilde\rho_xR^{-\frac14}\big\|_2^2\right) \qquad (3.10)\end{aligned}$$

We now claim that the first term in Eq.(3.10) can be dropped since it is smaller than one, and that the second term equals $\mathcal{P}\big(M^{(1)}\big)$. Both can be verified by simply writing out the Hilbert-Schmidt norm:

$$\begin{aligned}\sum_x \big\|\big(R^{\frac14}M_xR^{\frac14}\big\|_2^2 &= \sum_x \mathrm{tr}\Big[R^{\frac12}M_xR^{\frac12}\underbrace{M_x}_{\le\mathbb{1}}\Big] \le \mathrm{tr}\,[R] = 1, \\ \sum_x \big\|\big(R^{-\frac14}\tilde\rho_xR^{-\frac14}\big\|_2^2 &= \sum_x \mathrm{tr}\Big[\underbrace{R^{-\frac12}\tilde\rho_xR^{-\frac12}}_{=M_x^{(1)}}\tilde\rho_x\Big] = \mathcal{P}\big(M^{(1)}\big).\end{aligned}$$

□

To summarize these bounds in one statement, we define the error probability

$$\mathcal{P}_{err}(M) := 1 - \mathcal{P}(M) \quad \text{and} \quad \mathcal{P}_{err}^{opt} := 1 - \mathcal{P}^{opt}. \tag{3.11}$$

Corollary 3.5 (Pretty good bounds)**.** *For $\alpha \in \{1,2\}$ the measurements defined in Eq.(3.6) satisfy*

$$\big(\mathcal{P}^{opt}\big)^2 \leq \mathcal{P}\big(M^{(\alpha)}\big) \leq \mathcal{P}^{opt} \quad \text{and} \quad \mathcal{P}^{opt}_{err} \leq \mathcal{P}_{err}\big(M^{(\alpha)}\big) \leq 2\mathcal{P}^{opt}_{err}. \tag{3.12}$$

Proof. The chain of inequalities on the left is just a restatement of Prop.3.4 and a simple consequence of Prop.3.3. The statement on the right follows from there since

$$\mathcal{P}_{err}\big(M^{(\alpha)}\big) \leq 1 - \big(\mathcal{P}^{opt}\big)^2 = \big(1-\mathcal{P}^{opt}\big)\underbrace{\big(1+\mathcal{P}^{opt}\big)}_{\leq 2}.$$

□

Hence, the error of both the pretty good measurement and of the square-measurement is always within a factor of two of the optimal error probability. This becomes particularly meaningful when the error is small, which happens when more copies of the states are investigated jointly. In the limit of many copies, this implies that the error probabilities of both types of measurements have the same asymptotic scaling as a sequence of optimal measurements.

Proposition 3.6 (Fidelity bounds for optimal measurements)**.**
Let $\tilde{\rho}_1, \ldots, \tilde{\rho}_m$ be positive operators on $\mathbb{C}^d$ with $\sum_x \operatorname{tr}[\tilde{\rho}_x] = 1$. The error probability of the pretty good measurement $M^{(1)}$, defined in Eq.(3.6), satisfies[2]

$$\sum_{x<y} \big\|\tilde{\rho}_x^{\frac{1}{2}}\tilde{\rho}_y^{\frac{1}{2}}\big\|_1^2 \;\leq\; \mathcal{P}^{opt}_{err} \;\leq\; \mathcal{P}_{err}\big(M^{(1)}\big) \;\leq\; \sum_{x<y} \big\|\tilde{\rho}_x^{\frac{1}{2}}\tilde{\rho}_y^{\frac{1}{2}}\big\|_1. \tag{3.13}$$

Proof. Both, upper and lower bound rely on block-matrix inequalities for the proofs of which we refer to the original literature. We will, however, show how to obtain the result from those inequalities. For the general lower bound, we introduce an auxiliary Hilbert space $\mathbb{C}^m$ with ONB $\{|x\rangle\}_{x=1}^m$ and define $A := \sum_{x,y} \sqrt{M_x}\sqrt{\tilde{\rho}_y} \otimes |x\rangle\langle y|$, where $(M_x)_{x=1}^m$ is a POVM that we assume corresponds to an optimal measurement. We can regard A as a block matrix whose (x,y)-block $A_{xy} = \sqrt{M_x}\sqrt{\tilde{\rho}_y}$ satisfies $\|A_{xy}\|_2^2 = \operatorname{tr}[M_x\tilde{\rho}_y]$. Moreover, the normalization condition of the POVM implies that the corresponding block in A^*A is $\big(A^*A\big)_{xy} = \sqrt{\tilde{\rho}_x}\sqrt{\tilde{\rho}_y}$. Under the valid condition $\sum_{x,y}\|A_{xy}\|_2^2 = 1$ we can now apply the following block matrix inequality, which was proven in [54] and yields the lower bound upon inserting the expressions for the norms:

$$\sum_{x<y} \big\|(A^*A)_{xy}\big\|_1^2 \leq \sum_{x,y:x\neq y} \|A_{xy}\|_2^2. \tag{3.14}$$

For the upper bound, we use the construction $B := \sum_x \sqrt{\tilde{\rho}_x}R^{-1/4} \otimes |x\rangle$ with $R := \sum_x \tilde{\rho}_x$. This implies that $B^*B = \sqrt{R}$ and that BB^* and $(BB^*)^2$ are block matrices whose blocks take on the form $(BB^*)_{xy} = \sqrt{\tilde{\rho}_x}R^{-1/2}\sqrt{\tilde{\rho}_y}$ and $\big((BB^*)^2\big)_{xy} = \sqrt{\tilde{\rho}_x}\sqrt{\tilde{\rho}_y}$. In particular, $\|(BB^*)_{xy}\|_2^2 = \operatorname{tr}\big[\tilde{\rho}_x M_y^{(1)}\big]$. The claimed

[2] Here, the sums run over two indices, i.e., $\sum_{x<y} = \sum_x \sum_{y:y>x}$.

upper bound of the proposition is then implied by the following block matrix inequality that has been proven in [55]:

$$\sum_{x,y:x\neq y} \left\|(BB^*)_{xy}\right\|_2^2 \leq \sum_{x<y} \left\|\left((BB^*)^2\right)_{xy}\right\|_1. \tag{3.15}$$

The left-hand side is seen to equal the error probability after exploiting the identity $\sum_{x,y} \|(BB^*)_{xy}\|_2^2 = 1$. □

Discriminating on finitely many copies In practice, hypotheses are often discriminated not on the basis of a single experiment but rather on grounds of a finite repetition of an experiment followed by a statistical analysis. Assuming an n-fold repetition of the preparation of a quantum system (in an independent and identical manner), the hypotheses take on the form $\rho_x^{\otimes n}$ with a priori probability p_x that is independent of n. In this context, the mathematical analysis is often greatly simplified when using functionals for comparing the (dis)similarity of states that behave nicely under tensor products. One such functional is the *fidelity*.

Definition 3.7 (Fidelity)**.** *For two positive operators $\rho, \sigma \in \mathcal{B}_1(\mathcal{H})$ the* fidelity *is defined as*

$$F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1^2. \tag{3.16}$$

If ρ, σ are density operators, then $F(\rho, \sigma) \in [0, 1]$ and it is easy to see that the extremal values 0 and 1 are attained iff the states are orthogonal and identical, respectively. In particular, two hypotheses ρ, σ can be distinguished perfectly iff $F(\rho, \sigma) = 0$. This statement can be made more quantitative by the *Fuchs-van de Graaf inequalities* that state that any pair of density operators satisfies

$$1 - \sqrt{F(\rho, \sigma)} \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)}. \tag{3.17}$$

The following propositions will exploit the simple but useful multiplicativity property $F\big(\rho^{\otimes n}, \sigma^{\otimes n}\big) = F(\rho, \sigma)^n$.

Proposition 3.8. *Let $\epsilon, \delta > 0$ and $\rho_1, \ldots, \rho_m$ be density operators on $\mathcal{H} = \mathbb{C}^d$ s.t. for all $x \neq y : F(\rho_x, \rho_y) \leq 1 - \delta$. The following hold under the assumption*

$$n \geq \frac{2}{\delta} \ln\left(\frac{m-1}{2\epsilon}\right). \tag{3.18}$$

(1) For any choice of the prior $(p_x)_{x=1}^m$, the pretty good measurement ($M^{(1)}$ of Eq.(3.6)) applied to n copies (i.e., $\tilde{\rho}_x = p_x \rho_x^{\otimes n}$) satisfies $\mathcal{P}_{err}\big(M^{(1)}\big) \leq \epsilon$.

(2) There exists a POVM with $M_1, \ldots, M_x \in \mathcal{B}\big(\mathcal{H}^{\otimes n}\big)$ s.t. for all x :

$$\mathrm{tr}\left[\rho_x^{\otimes n}(\mathbb{1} - M_x)\right] \leq \epsilon. \tag{3.19}$$

Proof. (1) is an immediate consequence of the upper bound in Prop.3.6 and the multiplicativity of the fidelity:

$$\begin{aligned} \mathcal{P}_{err}\big(M^{(1)}\big) &\leq \sum_{x<y} \sqrt{F\big(\tilde{\rho}_x, \tilde{\rho}_y\big)} = \sum_{x<y} \sqrt{p_x p_y}\, F(\rho_x, \rho_y)^{\frac{n}{2}} \\ &\leq \tfrac{1}{2}(1-\delta)^{\frac{n}{2}} \left(\Big(\sum_x \sqrt{p_x}\Big)^2 - 1\right) \leq \tfrac{1}{2}(1-\delta)^{\frac{n}{2}}(m-1). \end{aligned} \tag{3.20}$$

Continuing with $(1 - \delta) \leq e^{-\delta}$ and inserting Eq.(3.18) then proves *(1)*.

(2) follows from *(1)* in the following way: consider the average success probability as a function of the POVM M *and* the prior distribution p, i.e., $\mathcal{P}(M) = \mathcal{P}(M, p)$. Then

$$\max_M \min_p \mathcal{P}(M, p) = \min_p \max_M \mathcal{P}(M, p) \geq 1 - \epsilon. \tag{3.21}$$

Here, the inequality on the right follows from *(1)* and the equality on the left is an application of Sion's minimax theorem, which applies since $\mathcal{P}$ is affine in both arguments and the corresponding optimization sets are convex. The POVM that achieves the maximum on the left then satisfies $\max_p \mathcal{P}_{err}(M, p) \leq \epsilon$. So this holds especially for every 'point mass distribution' where $p_y := \delta_{x,y}$. □

To summarize Prop.3.8, not only the average error probability but even the worst case error probability can be brought down to ϵ when measuring on $\mathcal{O}\big(\delta^{-1} \ln(m/\epsilon)$ copies.

Quantum Chernoff bound The discussion of the previous paragraph already suggests that the optimal error probability for discriminating hypotheses by measuring n copies of the state decreases exponentially with n. In fact, it turns out that

$$\mathcal{P}_{err}^{opt} = \exp\big[-\xi n + o(n)\big], \tag{3.22}$$

and the optimal error rate ξ is given by the *Quantum Chernoff distance*. For the case of $m = 2$ hypotheses with a priori probabilities $p_1, p_2 = 1 - p_1$, we do in principle know the optimal error probability, namely

$$\mathcal{P}_{err}^{opt}(n) := \tfrac{1}{2}\left(1 - \left\|p_1\rho_1^{\otimes n} - p_2\rho_2^{\otimes n}\right\|_1\right). \tag{3.23}$$

We also know from the previous paragraphs (Prop.3.6 and Prop.3.8) that the fidelity yields upper and lower bounds on the optimal error rate. More specifically, if both a priori probabilities are non-zero, then Prop.3.6 leads for $n \to \infty$ to

$$\begin{aligned} -\ln\sqrt{F(\rho_1, \rho_2)} \leq \quad & \liminf_n -\tfrac{1}{n}\ln \mathcal{P}_{err}^{opt}(n) \quad , \\ & \limsup_n -\tfrac{1}{n}\ln \mathcal{P}_{err}^{opt}(n) \quad \leq -\ln F(\rho_1, \rho_2)\,. \end{aligned} \tag{3.24}$$

Assuming that a limit, which we then denote by ξ, exists, we know that it has to lie within these bounds. In order to obtain the optimal error rate from Eq.(3.23), we need the following:

Lemma 3.9. *Let $\mathcal{H}$ be finite dimensional and $A, B \in \mathcal{B}(\mathcal{H})$ positive. Then for every $s \in [0, 1]$:*

$$\operatorname{tr}\left[(A^s - B^s)A^{1-s}\right] \;\leq\; \operatorname{tr}\left[(A - B)_+\right] \;\leq\; \operatorname{tr}\left[A^{1+s}B^{-s}\right]. \tag{3.25}$$

Proof. Left inequality: We apply operator monotonicity of the function $z \mapsto z^s, s \in [0, 1]$ to the inequalities $B \leq B + (A - B)_+$ and $A \leq A + (A - B)_- =$

$B + (A-B)_+$. In this way, we obtain

$$\begin{aligned}
\mathrm{tr}\left[(A^s - B^s)A^{1-s}\right] &\leq \mathrm{tr}\left[\left(\left(B+(A-B)_+\right)^s - B^s\right)A^{1-s}\right] \\
&\leq \mathrm{tr}\left[\left(\left(B+(A-B)_+\right)^s - B^s\right)\left(B+(A-B)_+\right)^{1-s}\right] \\
&= \mathrm{tr}\left[B\right] + \mathrm{tr}\left[(A-B)_+\right] - \mathrm{tr}\left[B^s\left(B+(A-B)_+\right)^{1-s}\right] \\
&\leq \mathrm{tr}\left[(A-B)_+\right].
\end{aligned}$$

Right inequality: We show first that it is sufficient to consider diagonal A and B. To this end we exploit that $\mathrm{tr}\left[A^{1+s}B^{-s}\right]$ is non-increasing under completely positive, trace-preserving maps. We apply this to the map $T(\cdot) := \sum_i P_i \cdot P_i$ which is constructed from the spectral decomposition $(A-B) =: \sum_i \lambda_i P_i$ where the P_i's are one-dimensional projections. Then since $\mathrm{tr}\left[(A-B)_+\right] = \mathrm{tr}\left[\left(T(A)-T(B)\right)_+\right]$ and now $[T(A), T(B)] = 0$ it is indeed sufficient to consider diagonal A and B. For those the assertion follows from the simple inequality $a - b \leq a(a/b)^s$ which holds for all positive real numbers a, b. □

Theorem 3.10 (Quantum Chernoff theorem)**.** *Assuming that* $\dim(\mathcal{H}) < \infty$ *and that both a priori probabilities are non-zero, the optimal error rate for discriminating two hypotheses* $\rho_1, \rho_2 \in \mathcal{B}(\mathcal{H})$ *satisfies*

$$\lim_{n\to\infty}\left(-\tfrac{1}{n}\ln \mathcal{P}_{err}^{opt}(n)\right) = \xi := -\ln\left(\inf_{s\in[0,1]}\mathrm{tr}\left[\rho_1^{1-s}\rho_2^s\right]\right). \tag{3.26}$$

Remarks: The remaining optimization problem turns out to be a benign one: since $s \mapsto \mathrm{tr}\left[\rho_1^{1-s}\rho_2^s\right]$ is convex, every local minimum is global. Moreover, if $\mathrm{supp}(\rho_1) = \mathrm{supp}(\rho_2)$, then the infimum is attained and thus a minimum. If the supports are different, the function can become discontinuous at the endpoints $s \in \{0, 1\}$. Regarded as a function of the states, $\xi = \xi(\rho_1, \rho_2)$ is called *Quantum Chernoff distance.*

Proof. We will prove that the rate ξ is achievable and sketch the main idea for showing that it cannot be exceeded. Achievability follows from the left inequality of Lemma 3.9, which can be rewritten as

$$\tfrac{1}{2}\left(\mathrm{tr}\left[A+B\right] - \|A-B\|_1\right) \leq \mathrm{tr}\left[B^s A^{1-s}\right],$$

when using that every Hermitian X satisfies $X_+ = \frac{1}{2}(|X| + X)$. Inserting $A = p_1\rho_1^{\otimes n}$ and $B = p_2\rho_2^{\otimes n}$ then leads to

$$\mathcal{P}_{err}^{opt}(n) \;\leq\; \mathrm{tr}\left[\rho_1^{1-s}\rho_2^s\right]^n p_1^{1-s}p_2^s\,. \tag{3.27}$$

Considering the limit (inferior) as in Eq.(3.26) and optimizing over $s \in [0, 1]$ then shows that ξ is indeed a lower bound on the optimal error rate.

That this rate cannot be surpassed is shown by regression to the classical case. In fact, the classical Chernoff bound for discrete probability distributions

states optimality of the bound in Eq.(3.26) – when only applied to commuting density operators. The core of the remaining part of the argument is thus to map the pair ρ_1, ρ_2 to a pair of commuting density operators $\hat{\rho}_1, \hat{\rho}_1$ in a way that guarantees in particular that

$$\operatorname{tr}\left[\rho_1^{1-s}\rho_2^s\right] = \operatorname{tr}\left[\hat{\rho}_1^{1-s}\hat{\rho}_2^s\right]. \tag{3.28}$$

This is achieved on the basis of the spectral decomposition of the density matrices via

$$\rho_x = \sum_i \lambda_i^{(x)} \left|\psi_i^{(x)}\right\rangle\!\left\langle\psi_i^{(x)}\right| \qquad \mapsto \qquad \hat{\rho}_1 := \sum_{i,j} \lambda_i^{(1)} |\langle\psi_i^{(1)}, \psi_j^{(2)}\rangle|^2 \; |ij\rangle\langle ij|$$
$$\hat{\rho}_2 := \sum_{i,j} \lambda_j^{(2)} |\langle\psi_i^{(1)}, \psi_j^{(2)}\rangle|^2 \; |ij\rangle\langle ij|$$

where $\{|ij\rangle\}$ is assumed to be a product ONB in $\mathcal{H} \otimes \mathcal{H}$. Using that $[\hat{\rho}_1, \hat{\rho}_2] = 0$ and Eq.(3.28) it can be shown that a quantum rate larger than ξ would contradict the classical Chernoff theorem. □

In general, the achievability of both, the optimal error probability of Eq.(3.23) and the optimal asymptotic error rate of the Quantum Chernoff theorem, assumes *global measurements*. That is, measurements that act collectively on all copies of the state at once. In practice, such measurements are significantly harder to implement than local measurements, which measure each copy individually.[3] If we restrict ourselves to local measurements that are in addition identical on all copies of the state, we can view the resulting restricted quantum hypothesis testing problem as a two-stage process: first, the measurement maps each of the quantum states to a corresponding 'classical' probability distribution and second, the measurement statistics is analyzed with the tools of 'classical' hypothesis testing. Taking into account the possibility of grouping measurement outcomes, the structure of the POVMs that appear under this restriction is the following:

Definition 3.11. *We say that a POVM* $(M_1, \ldots, M_m)$ *on* $\mathcal{H}^{\otimes n}$ *is implementable by* local, identical measurements *if there is a POVM* $(N_1, \ldots, N_k)$ *on* $\mathcal{H}$ *and a decomposition* $\{1, \ldots, k\}^n = \bigcup_{x=1}^m J_x$ *into disjoint subsets* J_x *such that*

$$M_x = \sum_{j \in J_x} N_{j_1} \otimes \cdots \otimes N_{j_n}. \tag{3.29}$$

In order to compare the power of global and local measurements, we need a little Lemma:

Lemma 3.12. *For* $m \in \mathbb{N}$, *density operators* $\rho, \sigma \in \mathcal{B}(\mathcal{H})$ *and a POVM* $(M_x)_{x=1}^m$ *on* $\mathcal{H}$, *define the outcome probabilities* $p_x := \operatorname{tr}[M_x\rho]$ *and* $q_x := \operatorname{tr}[M_x\sigma]$. *Then*

$$\sqrt{F(\rho,\sigma)} \leq \sum_x \sqrt{p_x q_x}, \tag{3.30}$$

[3] On the positive side, however, it can be shown that the effort increases only polynomially with n [56].

and if $\dim(\mathcal{H}) \leq m$*, there exists a PVM achieving equality in Eq.(3.30).*

Remark: In the classical statistics literature, the expression on the right-hand side of Eq.(3.30) is known as *Bhattacharyya coefficient*.

Proof. The inequality can be shown using triangle and Hölder's inequality:

$$\begin{aligned} \|\sqrt{\rho}\sqrt{\sigma}\|_1 &\leq \sum_x \|\sqrt{\rho}\sqrt{M_x}\sqrt{M_x}\sqrt{\sigma}\|_1 \\ &\leq \sum_x \|\sqrt{\rho}\sqrt{M_x}\|_2 \|\sqrt{M_x}\sqrt{\sigma}\|_2 = \sum_x \sqrt{p_x q_x}. \end{aligned}$$

In order to show that equality can be attained, let us first assume that ρ has full rank and introduce

$$R := \rho^{-\frac{1}{2}} \left(\sqrt{\rho}\sigma\sqrt{\rho}\right)^{\frac{1}{2}} \rho^{-\frac{1}{2}} =: \sum_x \lambda_x M_x,$$

where the right-hand side means a spectral decomposition of R with eigenvalues λ_x and corresponding eigenprojectors M_x. Using that $\sigma = R\rho R$ and $RM_xR = \lambda_x^2 M_x$ we obtain

$$\begin{aligned} \sum_x \sqrt{p_x q_x} &= \sum_x \sqrt{\mathrm{tr}\left[R\rho R M_x\right] \mathrm{tr}\left[\rho M_x\right]} \\ &= \sum_x \lambda_x \mathrm{tr}\left[\rho M_x\right] = \mathrm{tr}\left[R\rho\right] = \|\sqrt{\rho}\sqrt{\sigma}\|_1. \end{aligned} \tag{3.31}$$

If ρ does not have full rank, the same argument applies when restricted to the range of ρ. □

Proposition 3.13 (Local vs. global hypothesis testing)**.** *For the discrimination of two density operators* $\rho, \sigma \in \mathcal{B}(\mathbb{C}^d)$ *with non-zero a priori probabilities, let* ξ *be the quantum Chernoff distance, i.e., the optimal error rate of Thm.3.10 and* ξ_{loc} *the optimal error rate achievable by local, identical measurements. Then:*

(1) $-\frac{1}{2}\ln F(\rho,\sigma) \leq \xi_{loc} \leq \xi \leq -\ln F(\rho,\sigma) \leq 2\xi_{loc}$.

(2) If at least one of the states is pure, then $\xi_{loc} = \xi = -\ln F(\rho,\sigma)$.

Proof. (1) Suppose we perform a local measurement on each copy that produces an outcome x with probabilities $p_x := \mathrm{tr}\left[M_x\rho\right]$ or $q_x := \mathrm{tr}\left[M_x\sigma\right]$, depending on the actual hypothesis. Enabled by Lem.3.12, we choose the measurement such that $\sum_x \sqrt{p_x q_x} = \sqrt{F(\rho,\sigma)}$. Using the classical Chernoff theorem for the discrimination between p and q we get

$$\xi_{loc} \geq -\inf_{s\in[0,1]} \ln \sum_x p_x^s q_x^{1-s} \geq -\ln \sum_x \sqrt{p_x q_x} = -\tfrac{1}{2}\ln F(\rho,\sigma). \tag{3.32}$$

Clearly, $\xi_{loc} \leq \xi$ and from Eq.(3.24) we know that $\xi \leq -\ln F(\rho,\sigma)$. Combining this with Eq.(3.32) then proves the chain of inequalities.

(2) Suppose ρ is pure and consider a local PVM of the form $(\rho, \mathbb{1} - \rho)$. Applying this to n copies and 'guessing' ρ only if all n outcomes were correct, amounts to using an effective PVM of the form $(\rho^{\otimes n}, \mathbb{1} - \rho^{\otimes n})$, which is implementable by local identical measurements. This leads to an error probability that is up to the a priori probability of σ equal to $\operatorname{tr}[\sigma^{\otimes n}\rho^{\otimes n}] = F(\rho,\sigma)^n$. Hence, $\xi_{loc} \geq -\ln F(\rho,\sigma)$. However, since $-\ln F(\rho,\sigma) \geq \xi \geq \xi_{locc}$ all three quantities have to be equal. □

It is important to note that $\xi = \xi_{loc}$ does not necessarily mean that global measurements can not improve the error probability: focusing on the asymptotic error rate neglects any finite-n behaviour as well as any constant factor between the error probabilities. In fact, the measurement that was used in the proof of *(2)* is in general never optimal—not even for $n = 1$.

Finally, let us briefly comment on the general case of discriminating $m \geq 2$ quantum states $\rho_1, \ldots, \rho_m$ (under the assumption that all of them have non-zero a priori probability). Unsurprisingly, discriminating $m > 2$ states turns out to be at least as hard as discriminating any of the pairs of states in the set. Consequently, the optimal asymptotic error rate is bounded from above by $\min_{i\neq j} \xi(\rho_i, \rho_j)$ where $\xi(\rho_i, \rho_j)$ is the Quantum Chernoff distance from Thm.3.10. That this bound is, in fact, achievable was shown by Ke Li in [57]. What complicates the proof of this result is that, unlike in the case $m = 2$ where the quantum analog of the Neyman-Pearson-Lemma is available, there is no general explicit expression for the optimal error probability.

Asymmetric hypothesis testing Let us return to the case of $m = 2$ hypotheses, described by density operators ρ and σ. In the discussion of the asymptotic error rate of the quantum Chernoff theorem, we have treated the errors symmetrically and made no conceptual difference between misclassifying ρ and misclassifying σ. More precisely, if $(P_n, \mathbb{1} - P_n)$ is the POVM that is performed on n-copies of the state and

$$\alpha_n(P_n) := \operatorname{tr}\left[\rho^{\otimes n}(\mathbb{1} - P_n)\right] \quad \text{and} \quad \beta_n(P_n) := \operatorname{tr}\left[\sigma^{\otimes n} P_n\right] \tag{3.33}$$

are the *error of the first and second kind* (corresponding to 'false positive' and 'false negative'), respectively, then the quantum Chernoff distance is the largest ξ such that

$$\liminf_{n\to\infty} -\tfrac{1}{n}\ln \alpha_n(P_n) \geq \xi \quad \text{and} \quad \liminf_{n\to\infty} -\tfrac{1}{n}\ln \beta_n(P_n) \geq \xi, \tag{3.34}$$

can both be achieved simultaneously.

However, in some contexts (e.g., in medical statistical analysis) it is natural to treat these two errors in an asymmetric way and to try to achieve a faster decay rate for one of them at the cost of a slower rate for the other. In the following, we will consider the extreme scenario, where we only require one of them to be bounded by a constant. Specifically, we are interested in maximizing the decay rate of the error probability $\beta_n(P_n)$ under the constraint that $\alpha_n(P_n) \leq \epsilon$.

For a sequence of optimal measurements the error β_n will again decrease exponentially with n and the rate will turn out to be the *relative entropy* $S(\rho\|\sigma) = \operatorname{tr}[\rho(\ln\rho - \ln\sigma)]$. The analogous classical result is called *Stein's Lemma* where the *Kullback-Leibler divergence*, to which the relative entropy reduces for commuting density operators, appears as the optimal rate function. Before we come to the quantum version of this result we state a preparatory Lemma and a crucial ingredient for the proof of the Quantum Stein's Lemma:

Lemma 3.14. *Let $\rho \in \mathcal{B}(\mathcal{H})$ be a density operator and $\big(P_i \in \mathcal{B}(\mathcal{H})\big)_{i=1}^k$ a PVM. If $T(X) := \sum_{i=1}^k P_i X P_i$, then*

$$\rho \leq kT(\rho). \tag{3.35}$$

Proof. Due to linearity it is sufficient to consider pure states $\rho = |\psi\rangle\langle\psi|$. Then for every $|\phi\rangle \in \mathcal{H}$:

$$\langle\phi|kT(\rho) - \rho|\phi\rangle = \Big(k\sum_{i=1}^k |\langle\phi|P_i|\psi\rangle|^2\Big) - \big|\langle\phi|\sum_{i=1}^k P_i|\psi\rangle\big|^2 \geq 0,$$

where the inequality follows from Chauchy-Schwarz in $\mathbb{C}^k$ when applied to one vector with components $\langle\phi|P_i|\psi\rangle$ and one with all k components equal to 1. □

Theorem 3.15 (Hiai-Petz)**.** *Let $\rho, \sigma \in \mathcal{B}(\mathcal{H})$ be two density operators on $\mathcal{H} = \mathbb{C}^d$ and $n \in \mathbb{N}$. Define a map $T : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathcal{H}^{\otimes n})$ via $T(X) := \sum_{i=1}^k P_i X P_i$ from the spectral decomposition $\sigma^{\otimes n} = \sum_{i=1}^k \lambda_i P_i$ where the sum runs over k distinct eigenvalues and the P_i's project onto the corresponding eigenspaces. Then*

$$S\big(T(\rho^{\otimes n})\big\|\sigma^{\otimes n}\big) \;\leq\; nS(\rho\|\sigma) \;\leq\; S\big(T(\rho^{\otimes n})\big\|\sigma^{\otimes n}\big) + d\log(n+1), \tag{3.36}$$

where $S(\rho\|\sigma) = \operatorname{tr}[\rho(\log\rho - \log\sigma)]$ is the relative entropy.

Proof. The left inequality follows from monotonicity of the relative entropy under trace-preserving completely positive maps together with additivity $nS(\rho\|\sigma) = S\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big)$ and the invariance $T(\sigma^{\otimes n}) = \sigma^{\otimes n}$.

For the right inequality we use again additivity and proceed as follows:

$$\begin{aligned} n\,S(\rho\|\sigma) &\leq \operatorname{tr}\big[\rho^{\otimes n}\big(\log T(\rho^{\otimes n}) - \log\sigma^{\otimes n}\big)\big] + \log k && (3.37)\\ &\leq S\Big(T(\rho^{\otimes n})\Big\|\sigma^{\otimes n}\Big) + d\log(n+1). && (3.38)\end{aligned}$$

Here the first inequality uses $\log\rho^{\otimes n} \leq \log T(\rho^{\otimes n}) + \log k$ which follows from Lemma 3.14 together with the operator monotonicity of the logarithm. The inequality in Eq.(3.38) is derived as follows: for the first term we use that $\log T(\rho^{\otimes n}) = T\big(\log T(\rho^{\otimes n})\big)$, since the logarithm preserves the block structure, and that T satisfies $\operatorname{tr}[AT(B)] = \operatorname{tr}[T(A)T(B)]$ for any A, B. The second term in Eq.(3.38) follows from the simple combinatorial bound $k \leq (n+1)^d$ on the number of different eigenvalues of $\sigma^{\otimes n}$. This in turn follows from the fact that those eigenvalues are all of the form $\prod_{i=1}^d \nu_i^{a_i}$ where $(\nu_i)_{i=1}^d$ are the eigenvalues of σ and $0 \leq a_i \leq n$. □

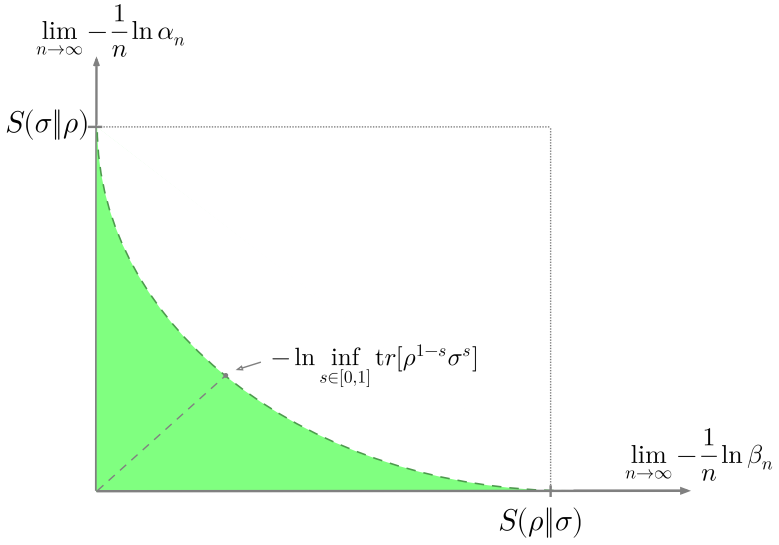

Figure 3.2: *Asymptotic binary hypothesis testing of quantum states.* The green shaded region depicts all pairs of achievable error rates (of the first and second kind). While the Quantum Chernoff distance characterizes the optimum point of symmetric error rates, the relative entropies correspond to the asymmetric limit points. The trade-off curve that delineates the achievable region is given by the Quantum Hoeffding bound in Eq.(3.44).

Note that Eq.(3.36) implies in particular that

$$S(\rho\|\sigma) = \lim_{n\to\infty} \frac{1}{n} S\big(T(\rho^{\otimes n})\big\|\sigma^{\otimes n}\big). \tag{3.39}$$

Moreover since $\sigma^{\otimes n}$ commutes with $T(\rho^{\otimes n})$ the relative entropy on the r.h.s. equals the one of the classical probability distributions which are obtained from measuring the two states in the basis in which they are simultaneously diagonal. Since the relative entropy (or Kullback-Leibler divergence) is the optimal rate function appearing in the classical Stein's lemma, the above Thm.3.15 implies that $S(\rho\|\sigma)$ is an achievable rate in the quantum context as well. The following shows that it is indeed the optimal rate:

Theorem 3.16 (Quantum Stein's Lemma)**.** *Consider the task of distinguishing two quantum states $\rho, \sigma \in \mathcal{M}_d(\mathbb{C})$. Let β_n be the error probability as defined in Eq.(3.33), minimized over all measurements. For every $\epsilon \in (0,1)$ we have that*

$$\lim_{n\to\infty} -\frac{1}{n}\ln\beta_n = S(\rho\|\sigma). \tag{3.40}$$

Proof. As discussed above achievability of the rate $S(\rho\|\sigma)$ follows from Thm.3.15 together with the classical Stein's lemma. What remains to prove is thus an

upper bound on the optimal rate, i.e., a lower bound on the error probability β_n. To this end, we assume that σ can be inverted on a space that contains the range of ρ—otherwise $S(\rho\|\sigma) = \infty$, which can be treated separately (see Example 3.4). Under this assumption, we can apply the right inequality of Lemma 3.9 to $A = \rho^{\otimes n}$ and $B = e^{\lambda n}\sigma^{\otimes n}$ for some $\lambda \in \mathbb{R}$ to be chosen later. For $s \in [0,1]$ this gives

$$\begin{aligned} e^{-s\lambda n}\mathrm{tr}\left[\rho^{1+s}\sigma^{-s}\right]^n &\geq \mathrm{tr}\left[\left(\rho^{\otimes n} - e^{\lambda n}\sigma^{\otimes n}\right)P_n\right] && (3.41)\\ &\geq (1-\epsilon) - e^{n\lambda}\beta_n(P_n), && (3.42)\end{aligned}$$

where the second inequality follows from Eq.(3.33). Rewriting these inequalities we obtain

$$\beta_n \geq e^{-n\lambda}\Big[(1-\epsilon) - e^{-n\left(\lambda s - f(s)\right)}\Big], \tag{3.43}$$

where we set $f(s) := \ln \mathrm{tr}\left[\rho^{1+s}\sigma^{-s}\right]$. Since $f(0) = 0$ and $f'(0) = S(\rho\|\sigma)$ the choice $\lambda = S(\rho\|\sigma) + \delta$ will for any $\delta > 0$ guarantee that there is an $s \in (0,1]$ such that $\lambda s > f(s)$. Thus, $\limsup_{n\to\infty} -\frac{1}{n}\ln\beta_n \leq S(\rho\|\sigma) + \delta$ and since this holds for arbitrary $\delta > 0$ the relative entropy is indeed the optimal asymptotic rate. □

Example 3.4 (Infinite rates)**.** If $S(\rho\|\sigma) = \infty$, then Quantum Stein's Lemma admits a decay of β_n with unbounded exponential rate. This can be understood as follows: $S(\rho\|\sigma) = \infty$ means that $\ker(\sigma) \not\subseteq \ker(\rho)$. That is, there is a unit vector ψ with $\sigma\psi = 0$ and $\rho\psi \neq 0$. Choosing $P_n := \mathbb{1} - (\mathbb{1} - |\psi\rangle\langle\psi|)^{\otimes n}$ then gives $\beta_n(P_n) = \mathrm{tr}\left[\sigma^{\otimes n}P_n\right] = 0$ while $\alpha_n(P_n) = \mathrm{tr}\left[\rho^{\otimes n}(\mathbb{1} - P_n)\right] = \mathrm{tr}\left[\rho(\mathbb{1} - |\psi\rangle\langle\psi|)\right]^n$ decays exponentially.

While the Quantum Chernoff theorem deals with *symmetric* hypothesis testing, the Quantum Stein's Lemma treats the two kinds of errors in a maximally asymmetric way. In between there is a trade-off that is described by the *Quantum Hoeffding theorem*, which we only mention without going into details. It states that the region of asymptotically achievable rates is bounded by an error function $e : [0,\infty) \to [0,\infty]$

$$e(r) := \sup_{s\in[0,1]} \frac{-rs - \ln \mathrm{tr}\left[\sigma^s\rho^{1-s}\right]}{1-s}. \tag{3.44}$$

More precisely, the pair of asymptotic error rates $(e, e(r))$ is achievable but (assuming the limits exist) if the rate for the error of the second kind exceeds r, then the rate for the error of the first kind is necessarily at most $e(r)$.

Exercise 3.1. Show that two hypotheses described by density operators ρ_1, ρ_2 with corresponding a priori probabilities $p_1, p_2 \in (0,1)$ can be discriminated perfectly iff they are orthogonal, i.e., $\rho_1\rho_2 = 0$.

Exercise 3.2. Let $\rho_1, \ldots, \rho_m$ be density operators on $\mathbb{C}^d$ with associated a priori probabilities $p_x = 1/m$, and $M = (M_1, \ldots, M_m)$ a corresponding maximum likelihood measurement. Show that for every $\lambda \in [0,1]$ M is still optimal if we replace every ρ_x by $\lambda\rho_x + (1-\lambda)\mathbb{1}/d$ with equal a priori probabilities.

Exercise 3.3. For a given $\mu \in [-1, 1]$ consider the qubit density operators

$$\rho_x := \frac{1}{2}\begin{pmatrix} 1 & \mu e^{2\pi i x/m} \\ \mu e^{-2\pi i x/m} & 1 \end{pmatrix}, \quad x = 1, \ldots, m, \tag{3.45}$$

each occurring with a priori probability $1/m$. Construct a maximum likelihood measurement for optimally distinguishing these states and compute the corresponding maximum average success probability.

Exercise 3.4 (Optimal classical strategy). Consider the scenario of commuting states of Example 3.1. Let $M = (M_1, \ldots, M_m)$ be a POVM whose elements are diagonal in the same basis as the considered states and define $m(i) := \max_x \langle i|\tilde{\rho}_x|i\rangle$. Show that M is a maximum likelihood measurement iff for all i and x:

$$\langle i|\tilde{\rho}_x|i\rangle < m(i) \quad \Rightarrow \quad \langle i|M_x|i\rangle = 0. \tag{3.46}$$

Exercise 3.5 (Pure states). Let $\psi_1, \ldots, \psi_d \in \mathbb{C}^d$ be linearly independent unit vectors. Consider the task of distinguishing the states $\rho_x := |\psi_x\rangle\langle\psi_k|$ with assigned a priori probabilities $p_x > 0$. Show that an optimal POVM $M = (M_1, \ldots, M_d)$

(a) satisfies $\operatorname{tr}[\rho_x M_x] > 0 \ \forall x$,

(b) corresponds to an orthonormal basis in the sense that there is an ONB $|\phi_x\rangle$ such that $M_x = |\phi_x\rangle\langle\phi_x|$,

(c) is unique.

Exercise 3.6 (Trace-norm vs. fidelity).

(a) For two density operators $\rho, \sigma \in \mathcal{B}(\mathcal{H})$ show that

$$1 - \tfrac{1}{2}\|\rho - \sigma\|_1 \leq F(\rho, \sigma)^{1/2}. \tag{3.47}$$

Hint: You may use the relation between the fidelity and the Bhattacharyya coefficient (Lemma 3.12) together with the inequality $(\sqrt{a} - \sqrt{b})^2 \leq |a - b|$.

(b) Let $\phi, \psi \in \mathcal{H}$ be two unit vectors and $a, b \in \mathbb{R}_+$. Show that

$$\big\|a|\phi\rangle\langle\phi| - b|\psi\rangle\langle\psi|\big\|_1 = \sqrt{(a+b)^2 - 4ab|\langle\phi|\psi\rangle|^2}. \tag{3.48}$$

Hint: Represent the operator in the trace-norm as a 2×2-matrix in the space spanned by ϕ, ψ and compute the eigenvalues.

Exercise 3.7. Let $\rho_1, \rho_2 \in \mathcal{B}(\mathcal{H})$ be two density matrices.

(a) Show that if the two states can be discriminated perfectly using $n \in \mathbb{N}$ copies (i.e. $\rho_1^{\otimes n}$ vs. $\rho_1^{\otimes n}$), then they can be discriminated perfectly using only a single copy. (*Hint:* You may want to consider the Fuchs van de Graaf inequalities)

(b) Prove that the use of an ancilla does not improve the optimal success probability when discriminating the two states.

Exercise 3.8 (Chernoff distance vs. fidelity). Let $\rho, \sigma \in \mathcal{B}(\mathbb{C}^d)$ be two density operators and $Q(s) := \operatorname{tr}\left[\rho^s \sigma^{1-s}\right]$ for $s \in [0, 1]$.

(a) What is the relation between $Q(s)$ and the fidelity $F(\rho, \sigma)$ if both states are pure?

(b) Show that $\inf_{s\in[0,1]} Q(s) \leq F(\rho, \sigma)^{1/2}$.

(c) Show that for every $s \in [0,1]$: $F(\rho,\sigma) \leq Q(s)$. Hint: you may write

$$\rho^{\frac{1}{2}}\sigma^{\frac{1}{2}} = \rho^{\frac{1-s}{2}}\left(\rho^{\frac{s}{2}}\sigma^{\frac{1-s}{2}}\right)\sigma^{\frac{s}{2}} ,$$

and use the Hölder inequality $\|AB\|_r \leq \|A\|_p\|B\|_q$ for $q,p,r \geq 1$ with $\frac{1}{q}+\frac{1}{p}=\frac{1}{r}$.

Exercise 3.9 (Necessary number of copies)**.** Assume m hypotheses of the form $\rho_x^{\otimes n}$ with equal a priori probabilities and such that for all x,y : $F(\rho_x,\rho_y) \geq f > 0$. Show that in order to discriminate the hypotheses with an average error probability of at most ϵ the number of copies has to be at least

$$n \geq \frac{\log 4\epsilon}{\log f} .$$

Exercise 3.10. Use the Quantum Stein's Lemma to argue that the relative entropy has to be non-increasing under completely positive, trace-preserving maps.

Notes and literature

Binary hypothesis testing for quantum states was first considered by Helstrom [15]. The multi-hypotheses optimality conditions of Thm.3.2 are due to Holevo [16, 58] and Yuen, Kennedy and Lax [19]. Their proofs were based on duality in convex analysis. The presented proof circumvents this approach by using an idea presented in [59]. Pretty good measurements appeared in the works of Belavkin [17, 18] and Hausladen and Wootters [60] first for pure states. The corresponding bound in Prop.3.4 goes back to Barnum and Knill [61], the presentation of the proof, however, follows Watrous ([62], p.138). The upper bound in Prop.3.3 appeared first in [63], where it has been formulated for all operator monotone functions $z \mapsto z^\alpha, \alpha \in (0,1]$. The measurement $M^{(2)}$ appeared as part of a numerical strategy in [64] and was analyzed in more detail by Tyson [65, 66] and Audenaert and Mosonyi [55]. The latter also prove the upper bound in Prop.3.6. If all states are pure, this bound can be quadratically improved [67, 55]. The lower bound, as well as its application in Exercise 3.9 is due to Montanaro [54]. The bound of Prop.3.8 on the number of sufficient copies is due to Harrow and Winter [68]. Montanaro [69] showed that the worst-case bound that was obtained by a non-constructive minimax argument in [68] can in fact be obtained by pretty good measurements.

The optimal asymptotic rate for discriminating two classical probability distributions was derived by Chernoff in [70]. Possible quantum extensions were discussed by Ogawa and Hayashi [71]. The proof of the optimality part of the Quantum Chernoff theorem appeared in [72] while achievability was proven in [73]. The combination and a good overview on asymptotic quantum hypothesis testing can be found in [74]. Extensions of these results to the (infinite-dimensional) von Neumann algebra setting appeared in [75]. [75] also contains the presented proof (attributed to N. Ozawa) of the left inequality of Lemma 3.9, which originally appeared in [73]. The comparison of local and global measurements in Prop.3.13 goes back to Kargin [76] and was investigated further e.g. in [77].

Hiai and Petz proved the eponymous theorem and, as a consequence, the achievability part of the Quantum Stein's Lemma in [78]. The optimality was shown by Ogawa and Nagaoka [79]. The quantum Hoeffding theorem is due to Hayashi [80] and Nagaoka [81] (see also [74]).

All the hitherto mentioned asymptotic hypothesis testing results are derived within a *fixed sample size* framework, where n is fixed in advance in the sense that only strategies that use all n copies are considered. Alternatively, one can consider *sequential* strategies that are allowed to stop at any n, once they have acquired sufficient confidence about the nature of the hypothesis. In the quantum context, *sequential hypothesis testing* was studied in [82] and [83], where it was shown that such methods admit error rates beyond the Hoeffding bound. In [84] it was shown that the optimal rate pair for adaptive sequential hypothesis testing of binary quantum states is given by $\big(S(\rho\|\sigma), S(\sigma\|\rho)\big)\big)$.

Hypothesis testing of quantum channels We will now turn to quantum hypothesis testing in scenarios where the hypotheses are given by quantum channels. This subject turns out to be significantly richer than the one that considers quantum states as hypotheses, so we will only introduce the basic concepts and visit some interesting examples.

Suppose there are hypotheses given by quantum channels $T_i : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$. In the extremal case where $\dim(\mathcal{H}_1) = 1$ we are back to hypothesis testing for quantum states. In fact, we can always at least partly reduce the general case to the discrimination of quantum states: after all, in order to discriminate channels in practice we will have to send states through them, which produces new states that can then be discriminated using known methods. There are, however, two additional questions to be addressed: (i) does it help to use a composite input on which the channel acts only on one subsystem? and (ii) how to choose the input state? The two resulting opportunities for optimization, namely the size of the ancilla space and the nature of the input state, motivate considering a new norm that can be considered a generalization of the trace norm, with which it coincides for the case $\dim(\mathcal{H}) = 1$.

Definition 3.17 (Induced and completely bounded trace norm)**.** *For a linear map $\Phi : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ we define*

$$\|\Phi\|_1 := \sup\left\{\|\Phi(X)\|_1 \,\middle|\, \|X\|_1 \leq 1\right\}, \tag{3.49}$$

$$\|\Phi\|_\diamond := \sup_{n\in\mathbb{N}} \left\|\Phi \otimes \mathrm{id}_n\right\|_1, \tag{3.50}$$

where $\mathrm{id}_n : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ denotes the identity map. $\|\cdot\|_\diamond$ is called the 'diamond norm' *or* 'completely bounded trace norm' *and the map Φ is said to be* 'completely bounded' *if $\|\Phi\|_\diamond < \infty$.*

It follows immediately from the definition that $\|\Phi\|_1 \leq \|\Phi\|_\diamond$ and that both are indeed norms on the space of completely bounded maps. Other properties that we will make use of are summarized in the following Lemma. (For more, see the exercises, where we will for instance see that $\|\Phi\|_\diamond \leq \|\Phi\|_1 \dim(\mathcal{H}_1)$.)

Lemma 3.18. *Let $\Phi : \mathcal{B}_1(\mathcal{H}_1) \to \mathcal{B}_1(\mathcal{H}_2)$ be a completely bounded linear map.*

1. *$\|\Phi\|_1 = \sup_{\varphi,\psi} \|\Phi(|\varphi\rangle\langle\psi|)\|_1$ where the suprema run over all unit vectors.*
2. *If Φ is* Hermiticity preserving*, i.e., $\forall X \in \mathcal{B}_1(\mathcal{H}_1) : \Phi(X^*) = \Phi(X)^*$, then*

$$\|\Phi\|_\diamond = \sup_n \sup_{\|\varphi\|=1} \left\|(\Phi \otimes \mathrm{id}_n)(|\varphi\rangle\langle\varphi|)\right\|_1, \tag{3.51}$$

   where the first supremum runs over all natural numbers $n \leq \dim(\mathcal{H}_1)$.

3. *If Φ is completely positive, then $\|\Phi\|_\diamond = \|\Phi\|_1 = \sup_\varphi \mathrm{tr}\left[\Phi(|\varphi\rangle\langle\varphi|)\right]$ with supremum over all unit vectors.*

Note: Any real linear combination of positive maps (e.g. quantum channels) is Hermiticity preserving.

Proof. 1. follows from convexity of the norm together with the fact that rank-one operators of the form $|\varphi\rangle\langle\psi|$ are the extreme points of the $\mathcal{B}_1$-unit ball (see Example 1.10).

2. For $\mathcal{H} := \mathcal{H}_1 \otimes \mathbb{C}^n$ and any $X \in \mathcal{B}_1(\mathcal{H})$ with $\|X\|_1 \leq 1$ define a Hermitian operator $H := \frac{1}{2}(X \otimes |0\rangle\langle 1| + X^* \otimes |1\rangle\langle 0|) \in \mathcal{B}_1(\mathcal{H} \otimes \mathbb{C}^2)$. Then $\|H\|_1 = \|X\|_1$ and using that Φ and thus $\Phi \otimes \mathrm{id}_n$ is Hermiticity preserving, we get

$$\begin{aligned} \big\|(\Phi \otimes \mathrm{id}_{2n})(H)\big\|_1 &= \tfrac{1}{2}\big\|(\Phi \otimes \mathrm{id}_n)(X) \otimes |0\rangle\langle 1| + (\Phi \otimes \mathrm{id}_n)(X)^* \otimes |1\rangle\langle 0|\big\|_1 \\ &= \big\|(\Phi \otimes \mathrm{id}_n)(X)\big\|_1. \end{aligned} \tag{3.52}$$

For computing $\|\Phi\|_\diamond$ we can thus restrict ourselves to Hermitian operators. Using convexity as in the proof of *1.* then leads to Eq.(3.51). The fact, that $n \leq \dim(\mathcal{H}_1)$ suffices then follows from the observation that for $n > \dim(\mathcal{H}_1)$ the Schmidt decomposition of φ only makes use of a subspace of $\mathbb{C}^n$ of dimension at most $\dim(\mathcal{H}_1)$.

3. uses Eq.(3.51) together with the fact that the trace-norm of a positive operator equals its trace. Therefore:

$$\|\Phi\|_\diamond = \sup_{n,\varphi} \underbrace{\mathrm{tr}\left[(\Phi \otimes \mathrm{id}_n)(|\varphi\rangle\langle\varphi|)\right]}_{=\mathrm{tr}[\Phi(\rho)]} = \sup_{\varphi} \mathrm{tr}\left[\Phi(|\varphi\rangle\langle\varphi|)\right], \tag{3.53}$$

where ρ is the reduced state of $|\varphi\rangle\langle\varphi|$ and the last equality exploits once again convexity, which allows us to replace the mixed state again by a pure one. □

Together with the already known interpretation of the trace norm, Eq.(3.51) provides the diamond norm with a clear operational meaning: suppose we want to discriminate a black-box quantum channel that is either given by T_1 with a priori probability p_1 or by T_2 with $p_2 = 1 - p_1$. When using an ancilla system whose dimension is at least that of the input of the T_i's and optimizing over all composite input states as well as over all global measurements, the infimum over all error probabilities is

$$\mathcal{P}_{err}^{opt} = \frac{1}{2}\left(1 - \big\|p_1 T_1 - p_2 T_2\big\|_\diamond\right). \tag{3.54}$$

If instead no ancilla system is used, then the optimal error probability is given by Eq.(3.54) with $\|\cdot\|_\diamond$ replaced by $\|\cdot\|_1$. That the use of an ancilla can make a difference is nicely demonstrated by the following example:

Example 3.5 (Discrimination of Werner-Holevo channels)**.** Define two quantum channels $T_+, T_- : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ via

$$T_\pm(\rho) := \frac{1}{d \pm 1}\big(\mathrm{tr}\,[\rho]\,\mathbb{1} \pm \rho^T\big),$$

and set the a priori probability of T_+ to $p := (d+1)/(2d)$. Then

$$\begin{aligned} \big\|pT_+ - (1-p)T_-\big\|_\diamond &= 1, \quad \text{whereas} \\ \big\|pT_+ - (1-p)T_-\big\|_1 &= \tfrac{1}{d}. \end{aligned}$$

That is, while the use of an ancilla enables perfect discrimination, the optimal error probability without the use of an ancilla does, for large d, barely improve over flipping a coin. In order to verify these claims, observe that the parameters are chosen such that $\big(pT_+ - (1-p)T_-\big) = \frac{1}{d}T$, where $T(\rho) := \rho^T$ is the transposition. This satisfies $\|T\|_1 = 1$ and $\|T\|_\diamond = d$.

Discrimination of unitaries In order to analyze the discrimination of unitaries (or isometries), we equip our toolbox with the following concepts:

Definition 3.19 (Numerical range and spectral arc-length)**.** *Let $U \in \mathcal{B}(U)$.*

- *The* numerical range *of U is defined as*

$$\mathcal{N}(U) := \{\langle\psi, U\psi\rangle \;|\psi \in \mathcal{H}, \|\psi\| = 1\} \subset \mathbb{C}. \tag{3.55}$$

- *If U is unitary, we define the* spectral arc-length $\Theta(U) \in [0, 2\pi]$ *as the length of the shortest compact interval $I \subset \mathbb{R}$ such that the corresponding segment* $\exp(iI)$ *of the unit circle contains the spectrum of U.*

The central property of the numerical range that will allow us to relate it to the spectral arc-length is the follow:

Lemma 3.20 (Toeplitz-Hausdorff theorem for normal operators)**.** *The convex hull of the spectrum of any normal operator $U \in \mathcal{B}(\mathcal{H})$ is equal to the closure of its numerical range, i.e.,*

$$\mathrm{conv}\big(\mathrm{spec}(U)\big) = \overline{\mathcal{N}(U)}. \tag{3.56}$$

Remarks: 1. For finite-dimensional Hilbert spaces the numerical range is closed. 2. By the general *Toeplitz-Hausdorff theorem* the numerical range is convex for any, even unbounded, operator (see [85] for a surprisingly simple proof). However, the convex hull of the spectrum is in general only a proper subset of the closed numerical range.

Proof. Since U is normal there is a sequence of diagonal operators U_n with $\mathrm{spec}(U_n) = \mathrm{spec}(U)$ that converges to U in operator norm (cf. Example 1.6). We can therefore consider a spectral decomposition of the form $U_n = \sum_k \nu_k |k\rangle\langle k|$ where the closure of the set of eigenvalues $\{\nu_k\}$ is equal to the spectrum of U. Hence,

$$\overline{\mathrm{conv}\{\nu_k\}} = \mathrm{conv}\big(\mathrm{spec}(U)\big). \tag{3.57}$$

On the other hand, we have $\langle\psi, U_n\psi\rangle = \sum_k \nu_k \lambda_k$ with $\lambda_k := |\langle\psi|k\rangle|^2$. As varying over all unit vectors ψ is therefore equivalent to varying over all discrete probability distributions λ, we have

$$\overline{\mathrm{conv}\{\nu_k\}} = \overline{\mathcal{N}(U_n)}, \tag{3.58}$$

so that the fact $U_n \to U$ completes the proof. □

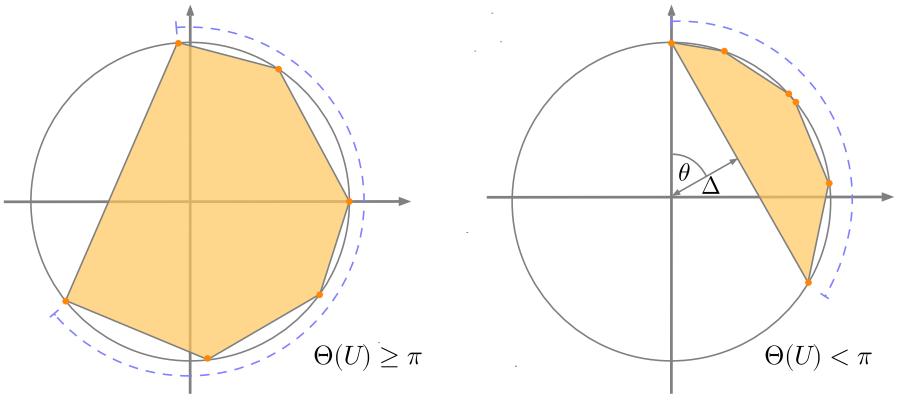

Figure 3.3: *Spectral arc-length vs. numerical range of a unitary* U. The points on the unit circle represent the elements of spec(U), the shaded region is the numerical range. The dashed line indicates the corresponding spectral arc-length. *Left:* If $\Theta(U) \geq \pi$, then 0 is contained in the convex hull of the spectrum and thus (by Lem.3.20) in the closure of the numerical range of U. *Right:* If $\Theta(U) < \pi$, then the distance Δ between the numerical range and the origin is given by $\Delta = \cos\theta$ where $\theta = \Theta(U)/2$.

The spectral arc-length $\Theta(U)$ of a unitary U can be characterized in many equivalent ways. Via spectral decomposition of the unitary one obtains for instance

$$\Theta(U) = 2\inf\left\{\|H\|_\infty \middle|\ H = H^* \wedge e^{iH}U^* \propto \mathbb{1}\right\}. \tag{3.59}$$

Moreover, we will see in the Exercise that $\Theta(U)$ can be interpreted as the length of the shortest path from $\mathbb{1}$ to U in the unitary group.

By definition, the spectral arc-lengths of $U, U^*, \mathbb{1} \otimes U$ and $e^{i\alpha}U$ with $\alpha \in \mathbb{R}$ are all the same. Less obvious properties, including the relation to the numerical range, are summarized in the following proposition.

Proposition 3.21. *Let* $U, V \in \mathcal{B}(\mathcal{H})$ *be unitaries.*

1. *With* $\theta := \frac{1}{2}\min\{\Theta(U), \pi\}$ *we have*

$$\inf_{\|\psi\|=1} \left|\langle\psi, U\psi\rangle\right| = \cos(\theta). \tag{3.60}$$

   In particular, $0 \in \overline{\mathcal{N}(U)} \ \Leftrightarrow\ \Theta(U) \geq \pi$.

2. $\Theta(U \otimes V) \leq \Theta(U) + \Theta(V)$ *with equality if the r.h.s. is not larger than* π.

3. $\Theta(U) + \Theta(V) \geq \pi \ \Rightarrow\ \Theta(U \otimes V) \geq \pi$.

4. *If* $\dim(\mathcal{H}) < \infty$, *then* $\Theta(UV) \leq \Theta(U) + \Theta(V)$.

5. $\Theta(UV) = \Theta(VU)$.

Proof. 1. Let us denote the infimum in Eq.(3.60) by Δ. By definition, Δ is equal to the minimum absolute value of any element in the closed numerical range $\overline{\mathcal{N}(U)}$. According to Lemma 3.20 the latter equals the convex hull of the spectrum of U. If $\Theta(U) \geq \pi$, then clearly $0 \in \mathrm{conv}\big(\mathrm{spec}(U)\big)$ and thus $\Delta = 0$. If $\Theta(U) < \pi$, then a simple geometric consideration (see Fig.3.3) shows that $\Delta = \cos\left(\frac{1}{2}\Theta(U)\right)$.

2. Let us denote the shortest intervals that cover the spectra of U and V by I and J, respectively and assume w.l.o.g. (by multiplication with a phase factor) that $J = [0, \Theta(V)]$ and $I = [a, b]$. The spectrum of a tensor product of bounded operators is equal to the product of their spectra [86]. On the level of the unit circle, this amounts to taking the Minkowski sum of the two sets. Hence, the interval $I + J$ whose length is bounded by $|I| + |J| = \Theta(U) + \Theta(V)$ covers the spectrum of UV, which proves the inequality. Moreover, if this sum is at most π, then the interval $[a, b + \Theta(V)]$ is the shortest path between the points $\exp(ia)$ and $\exp(i(b + \Theta(V)))$ on the unit circle. Therefore any other interval that covers these points must be at least as long.

3. As the spectra of both U and V are contained in the spectrum of $U \otimes V$ (up to a global phase factor) we have that $\Theta(U \otimes V) \geq \max\{\Theta(U), \Theta(V)\}$. Therefore the implication remains to be proven only in the case where the individual arc-lengths are both less than π. In this case, we observe that the premise $\Theta(U) + \Theta(V) \geq \pi$ excludes the existence of a segment of length larger than π in the complement of the spectrum of $U \otimes V$. Consequently, $\Theta(U \otimes V) \geq \pi$.

4. Assume that H_1 and H_2 are the Hermitian operators that achieve the minimum norm in Eq.(3.59) for U and V, respectively. Then *Thompson's theorem* [87] guarantees the existence of unitaries $W_1, W_2 \in \mathcal{B}(\mathcal{H})$ such that

$$UV \propto e^{iH_1} e^{iH_2} = e^{i(W_1 H_1 W_1^* + W_2 H_2 W_2^*)}, \tag{3.61}$$

where the neglected proportionality factor is of the form $e^{i\alpha}$, $\alpha \in \mathbb{R}$ and does thus not affect the spectral arc-lengths. In combination with Eq.(3.59) and the triangle-inequality for the norm we obtain

$$\Theta(UV) \leq 2\big\|W_1 H_1 W_1^* + W_2 H_2 W_2^*\big\|_\infty \leq 2\big(\|H_1\|_\infty + \|H_2\|_\infty\big) = \Theta(U) + \Theta(V).$$

5. follows from the general fact that the non-zero spectrum of a product of two operators does not change if we change the order of the product. Since zero is excluded in our case, the spectra are the same and so are $\Theta(UV)$ and $\Theta(VU)$. □

Example 3.6 (Shrinking spectral arc-length)**.** The restriction to $\Theta(U)+\Theta(V) \leq \pi$ in *2., 4.* of the preceding Proposition might seem odd at first glance. However, it is easy to see that this is necessary. Take for instance any unitary U with $\mathrm{spec}(U) = \{-1, 1\}$ like the Pauli matrices. Then $\Theta(U) = \pi$ but $\Theta(UU) = \Theta(\mathbb{1}) = 0$ and $\Theta(U \otimes U) = \pi$.

Proposition 3.22 (Distance between isometries)**.** *Let $V_1, V_2 : \mathcal{H}_1 \to \mathcal{H}_2$ be isometries between Hilbert spaces, $T_i(\cdot) := V_i \cdot V_i^*$ the corresponding quantum*

channels and $p \in [0,1]$. If $\nu \in [0,1]$ denotes the distance between the numerical range $\mathcal{N}(V_1^* V_2)$ and the origin in the complex plane, then:

$$\big\|pT_1 - (1-p)T_2\big\|_\diamond = \big\|pT_1 - (1-p)T_2\big\|_1 = \sqrt{1 - 4p(1-p)\nu^2}. \tag{3.62}$$

Moreover, if $V_1^* V_2$ is unitary and $\theta := \frac{1}{2}\min\{\Theta(V_1^* V_2), \pi\}$, then $\nu = \cos(\theta)$.

Remark: Note that this means in particular that in the unitary case,

$$\tfrac{1}{2}\|T_1 - T_2\|_\diamond = \sin\theta. \tag{3.63}$$

Relating this to the graphical depiction of Fig.(3.3, right) means that $\|T_1 - T_2\|_\diamond$ is precisely the length of the numerical range's largest edge (i.e., the one facing the origin).

Proof. We exploit that, according to Eq.(3.51), we can restrict the optimization for the diamond norm to pure states. Together with the fact that the trace-norm distance of two weighted pure states can be expressed in terms of their fidelity, as in Eq.(3.48), we obtain

$$\big\|\big(pT_1 \otimes \mathrm{id}_n - (1-p)T_2 \otimes \mathrm{id}_n\big)(|\varphi\rangle\langle\varphi|)\big\|_1 = \sqrt{1 - 4p(1-p)\big|\langle\varphi, (V_1^* V_2 \otimes \mathbb{1}_n)\varphi\rangle\big|^2}.$$

It remains to take the suprema over n and over all unit vectors φ. The latter, however, can be simplified since

$$\inf_{\|\varphi\|=1} \big|\langle\varphi, (V_1^* V_2 \otimes \mathbb{1}_n)\varphi\rangle\big| = \inf_\rho \big|\mathrm{tr}\,[V_1^* V_2 \rho]\,\big| = \inf_{\|\varphi\|=1} \big|\langle\varphi, (V_1^* V_2)\varphi\rangle\big| =: \nu,$$

where the second infimum runs over all density operators (as those arise as reduced states of φ) and the last step uses convexity of the numerical range (cf. Lemma 3.20) in order to restrict to pure states again—now, however, on a smaller Hilbert space. Eq.(3.1) shows that we can w.l.o.g. restrict to $n = 1$, which also proves the first equality in Eq.(3.62). Finally, if $V_1^* V_2$ is unitary, Eq.(3.60) completes the proof. □

We obtain an immediate corollary when looking at the cases of maximal diamond norm, meaning vanishing optimal error probability according to Eq.(3.54):

Corollary 3.23 (Perfect single-shot discrimination of unitaries)**.** *Two unitaries $V_1, V_2 : \mathcal{H}_1 \to \mathcal{H}_2$ can be distinguished with arbitrarily small error probability in a single-shot experiment (with or without ancilla) iff $\Theta(V_1^* V_2) \geq \pi$.*

If we recall the relation of Eq.(3.60) between the spectral arc-length and the numerical range (and assume for the moment that the latter is closed), the criterion of Cor.3.23 becomes easy to understand: it is equivalent to saying that there exists a unit vector $\psi \in \mathcal{H}$ for which $V_2\psi$ is orthogonal to $V_1\psi$ (or arbitrarily close to orthogonal, in general).

Combining this result with the properties of the spectral arc-length, which we collected in Prop.3.21, we can extend this result to the case where multiple (say n) uses of the unknown unitary time-evolution are allowed. Let us first consider the case where the time-evolutions are used in parallel:

Proposition 3.24 (Perfect parallel discrimination of unitaries)**.** *Let $n \in \mathbb{N}$ and $V_1, V_2 : \mathcal{H}_1 \to \mathcal{H}_2$ be unitaries. Then $V_1^{\otimes n}$ can be distinguished from $V_2^{\otimes n}$ with arbitrarily small error probability iff $n\Theta(V_1^* V_2) \geq \pi$.*

Proof. According to Cor.3.23, discrimination with arbitrarily small error probability is possible iff $\pi \leq \Theta\big((V_1^*)^{\otimes n} V_2^{\otimes n}\big) = \Theta(U^{\otimes n})$ with $U := V_1^* V_2$. Due to *2., 3.* in Prop.3.21 this is equivalent to $\pi \leq n\Theta(U)$. □

This means in particular that any pair of different unitaries can be distinguished with arbitrarily small (in the finite-dimensional case *zero*) error probability by using only a finite number of copies. In practice, however, a parallel scheme like the one underlying Prop.3.24 has the drawback that global measurements and entangled input states might be necessary. It turns out, however, that a sequential scheme, which avoids these obstacles, is sufficient. As *3.* of Prop.3.21 only holds for tensor products but in general not for composition, there is slightly more work required for the result. This is accomplished by the following Lemma.

Lemma 3.25. *Let $U \in \mathcal{B}(\mathcal{H})$ be a unitary and* [4] *$n := \lceil \frac{\pi}{\Theta(U)} \rceil$. Then there exists a unitary $W \in \mathcal{B}(\mathcal{H})$ such that $0 \in \overline{\mathcal{N}(W^* U W U^{n-1})}$.*

Proof. For $n = 1$ the result follows (with $W = \mathbb{1}$) from Prop. 3.21. So $n \geq 2$. Moreover, we can w.l.o.g. assume U to be a diagonal operator (or otherwise approximate it by one with equal spectrum). Consider the two eigenvalues that determine the boundary of the 'spectral arc', whose length is measured by $\Theta(U)$. Restricted to the two-dimensional subspace spanned by the corresponding eigenvectors, we can, up to a global phase factor, view U as a 2×2-matrix u of the form $u = \mathrm{diag}(1, e^{i\beta})$ where $n\beta \geq \pi > (n-1)\beta$ since $\beta = \Theta(U)$.

Within this two-dimensional subspace, define a unitary $W_\alpha := \exp(i\alpha\sigma_2)$ that represents a rotation by an angle α. Note that for $\alpha = \pi/2$ we have $W_\alpha^* \mathrm{diag}(1, e^{i\beta}) W_\alpha = \mathrm{diag}(e^{i\beta}, 1)$. Now consider a function $r : \alpha \mapsto r(\alpha) \in \mathbb{R}$ defined via

$$e^{ir(\alpha)} := \frac{\lambda_2\big(W_\alpha^* u W_\alpha u^{n-1}\big)}{\lambda_1\big(W_\alpha^* u W_\alpha u^{n-1}\big)}, \tag{3.64}$$

where $\lambda_1(\cdot), \lambda_2(\cdot)$ denote the two eigenvalues of the unitary in their argument. Then $r(0) = n\beta \geq \pi$ whereas $r(\pi/2) = (n-2)\beta < \pi$. So by continuity of r, the intermediate value theorem guarantees the existence of an angle α for which $r(\alpha) = \pi$. For this angle, the two eigenvalues are therefore antipodal on the unit circle so that $0 \in \mathcal{N}(W_\alpha^* u W_\alpha u^{n-1})$. Choosing $W := W_\alpha \oplus \mathbb{1}$ and considering the closure of the numerical range (in case $\dim(\mathcal{H}) = \infty$) then completes the proof. □

Theorem 3.26 (Perfect sequential discrimination of unitaries)**.** *Let $V_1, V_2 \in \mathcal{B}(\mathcal{H})$ be different unitaries, and $n := \lceil \frac{\pi}{\Theta(V_1^* V_2)} \rceil$.*

[4] $\lceil x \rceil$ denotes the smallest integer not less than x.

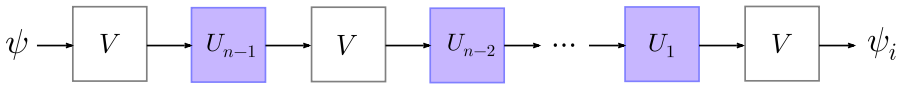

Figure 3.4: *Sequential scheme* for the discrimination of unitary hypotheses $V \in \{V_1, V_2\}$. If and only if $n\Theta(V_1^* V_2) \geq \pi$ there exist an input state ψ and unitary evolutions $U_1, \ldots, U_{n-1}$ so that the two outputs ψ_i, $i \in \{1, 2\}$ that correspond to the two different hypotheses are orthogonal.

1. *There exist unitaries* $U_1, \ldots, U_{n-1} \in \mathcal{B}(\mathcal{H})$ *s.t. for any* $\epsilon > 0$ *(and* $\epsilon = 0$ *if* $\dim(\mathcal{H}) < \infty$*) there is a unit vector* $\psi \in \mathcal{H}$ *for which* $|\langle \psi_1, \psi_2 \rangle| \leq \epsilon$ *where* $\psi_i := V_i U_1 V_i U_2 \cdots U_{n-1} V_i \psi$, $i \in \{1, 2\}$.

2. *If* $\dim(\mathcal{H}) < \infty$ *and* $m < n$*, then for all unitaries* $U_1, \ldots, U_{m-1} \in \mathcal{B}(\mathcal{H})$*:*

$$\inf_{\|\psi\|=1} |\langle V_1 U_1 V_1 U_2 \cdots U_{m-1} V_1 \psi, V_2 U_1 V_2 U_2 \cdots U_{m-1} V_2 \psi \rangle| \neq 0. \tag{3.65}$$

In other words, $n := \lceil \frac{\pi}{\Theta(V_1^* V_2)} \rceil$ sequential uses of the unitary $V \in \{V_1, V_2\}$ are necessary (*2.*) and sufficient (*1.*) to discriminate the two cases with arbitrarily small error probability in a sequential scheme as depicted in Fig.3.4.

Proof. 1. Applying Lemma 3.25 to $U = V_1^* V_2$ we know that there exists a unitary W and a unit vector ψ such that $W\psi$ and $V_1^* V_2 W (V_1^* V_2)^{n-1} \psi$ are orthogonal up to ϵ. This becomes the statement in the theorem when setting $U_1 := W V_1^*$ and $U_2 := \ldots = U_{n-1} := V_1^*$.

2. As the statement is for all tuples of unitaries, we can w.l.o.g. replace each U_k by $(U_k V_1^*)$. Defining $U := V_1^* V_2$ and using *1.* of Prop.3.21, we see that Eq.(3.65) becomes equivalent to

$$\Theta\big(U_{m-1}^* \cdots U_2^* U_1^* U U_1 U U_2 \cdots U U_{m-1} U\big) < \pi. \tag{3.66}$$

Making repeated use of *5.* and *4.* of Prop.3.21 together with the fact that $\Theta(U_k U U_k^*) = \Theta(U)$, we can bound the l.h.s. of Eq.(3.66) by $m\Theta(U)$ which is smaller than π by assumption. □

Perfect discrimination of arbitrary quantum channels In this paragraph, we will investigate perfect discrimination of pairs of quantum channels within two types of strategies that both use a finite number of copies of the unknown quantum channel: *parallel strategies* and more general *adaptive strategies* (see Fig.3.5). For both cases, we will find necessary and sufficient conditions, which will eventually allow us to show that adaptive strategies are strictly more powerful than parallel ones.

Theorem 3.27 (Perfect parallel discrimination of quantum channels)**.** *For* $\dim(\mathcal{H}_1) < \infty$*, let* $T_A, T_B : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ *be two quantum channels with sets of Kraus operators* $\{A_k\}, \{B_l\} \subset \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ *and define* $S := \operatorname{span}\{A_k^* B_l\} \subseteq \mathcal{B}(\mathcal{H}_1)$ *and*

$$\Delta_n := \|T_A^{\otimes n} - T_B^{\otimes n}\|_\diamond \quad \textit{for} \quad n \in \mathbb{N}. \tag{3.67}$$

1. $\Delta_n = 2$ *iff there is a density operator* $\rho \in \mathcal{B}(\mathcal{H}_1^{\otimes n})$ *in the orthogonal complement of* $S^{\otimes n}$ *(w.r.t. the Hilbert-Schmidt inner product).*

2. *If* S *contains a positive definite operator, then* $\forall n \in \mathbb{N} : \Delta_n < 2$.

Remark: That is, *1.* gives a necessary and sufficient condition for the possibility of perfectly discriminating the two cases with the possible use of an ancilla and n channels in parallel, whereas *2.* provides a condition under which no finite n suffices.

Proof. 1. Consider the case $n = 1$. $\Delta_1 = 2$ is equivalent to the existence of a pure input state whose outputs under the two channels are orthogonal. That is, a unit vector $\varphi \in \mathcal{H}_1 \otimes \mathcal{H}_1$ such that

$$\begin{aligned} 0 &= \operatorname{tr}\left[(T_A \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)(T_B \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)\right] \\ &= \sum_{k,l} \left|\langle \varphi, (A_k^* B_l \otimes \mathbb{1})\varphi\rangle\right|^2 = \sum_{k,l} \left|\operatorname{tr}\left[A_k^* B_l \rho\right]\right|^2, \end{aligned} \tag{3.68}$$

where in the last step ρ is the reduced state of $|\varphi\rangle\langle\varphi|$. As Eq.(3.68) is a sum of positive terms, it vanishes iff all of them vanish individually, which in turn means $\rho \perp S$. The case for general $n \in \mathbb{N}$ follows by realizing that the operator subspace corresponding to the n-fold tensor power of the channels is $S^{\otimes n}$.

2. follows from *1.* in the following way: suppose there is a positive definite $P \in S$. Then $P^{\otimes n} \in S^{\otimes n}$ is also positive definite so that any density operator $\rho \in \mathcal{B}(\mathcal{H}_1^{\otimes n})$ satisfies $\operatorname{tr}\left[\rho P^{\otimes n}\right] > 0$ and can thus not be orthogonal to $S^{\otimes n}$. □

In order to derive a similar result for more general (adaptive) strategies, we will need the following notion:

Definition 3.28 (Entanglement-assisted disjointness)**.** *Let* $\mathcal{H}_1, \mathcal{H}_2$ *be finite-dimensional. Two quantum channels* $T_A, T_B : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ *are called* entangle-ment-assisted disjoint *if there is a unit vector* $\varphi \in \mathcal{H}_1 \otimes \mathcal{H}_1$ *such that the ranges of* $(T_A \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)$ *and* $(T_B \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)$ *have trivial intersection* $\{0\}$.

Remark: The range of a positive semi-definite matrix is often also called its *support* and it is equal to the orthogonal complement of its kernel. An alternative characterization of 'entanglement assisted disjointness' is thus the existence of a unit vector φ s.t.

$$\ker\big[(T_A \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)\big] + \ker\big[(T_B \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)\big] = \mathcal{H}_2 \otimes \mathcal{H}_1. \tag{3.69}$$

Clearly, if the ranges of the two output states are orthogonal, the states and thus the channels can be perfectly discriminated. Def.3.28 defines a slightly weaker notion as it only requires that the ranges have a non-zero minimal angle between them. The relevance of this property in the present context stems from the following (see [88, 89] for a proof):

Proposition 3.29. *Consider finite dimensional Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$. Let $\rho_A, \rho_B \in \mathcal{B}(\mathcal{H}_1)$ be density operators with P_A and P_B the respective projectors onto their ranges. Given two unit vectors $\psi_A, \psi_B \in \mathcal{H}_2$ there exists a quantum channel $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ s.t. $T(\rho_i) = |\psi_i\rangle\langle\psi_i|$ for $i \in \{A, B\}$ iff*

$$\|P_A P_B\|_\infty \leq |\langle\psi_A, \psi_B\rangle|. \tag{3.70}$$

Remark: $\arccos \|P_A P_B\|_\infty$ is the mentioned angle between the subspaces given by the ranges. An alternative way of expressing the l.h.s. of Eq.(3.70), which may also clarify the interpretation of the angle, is

$$\|P_A P_B\|_\infty = \max\big\{|\langle\phi_A, \phi_B\rangle| \;\big|\; \phi_i \in P_i\mathcal{H}_1, \|\phi_i\| = 1\big\}. \tag{3.71}$$

Theorem 3.30 (Perfect adaptive discrimination of quantum channels)**.** *For $\dim(\mathcal{H}_1), \dim(\mathcal{H}_2) < \infty$, let $T_A, T_B : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ be two quantum channels with sets of Kraus operators $\{A_k\}, \{B_l\} \subset \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ and define $S := \mathrm{span}\{A_k^* B_l\} \subseteq \mathcal{B}(\mathcal{H}_1)$. Then T_A and T_B are perfectly distinguishable by a finite number of uses iff the following conditions are both satisfied:*

(i) T_A and T_B are entanglement assisted disjoint.

(ii) $\mathbb{1} \notin S$.

Proof. Suppose T_A, T_B can be distinguished perfectly with n uses and that n is the smallest such number. Let φ describe the input state before the first use of $T \in \{T_A, T_B\}$ and $\rho_A := (T_A \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)$ and $\rho_B := (T_A \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)$ the states after the first use. If the ranges of ρ_A and ρ_B had non-trivial intersection containing e.g. a unit vector ψ, then there is a $\lambda > 0$ such that $\rho_A, \rho_B \geq \lambda|\psi\rangle\langle\psi|$. Under this assumption, the success (i.e., vanishing error probability) of the overall protocol can thus be expressed by the equation

$$\mathrm{tr}\Big[(\mathbb{1} - M_i)\,\Phi^{(i)}\big(\lambda|\psi\rangle\langle\psi| + (1-\lambda)\rho^{(i)}\big)\Big] = 0, \quad \forall i \in \{A, B\}, \tag{3.72}$$

where M_i is the POVM element corresponding to a measurement outcome i, $\Phi^{(i)}$ is some quantum channel that contains $n-1$ uses of T_i and $\rho^{(i)}$ is some density operator. If Eq.(3.72) holds for some $\lambda > 1$ it also holds for $\lambda = 1$.

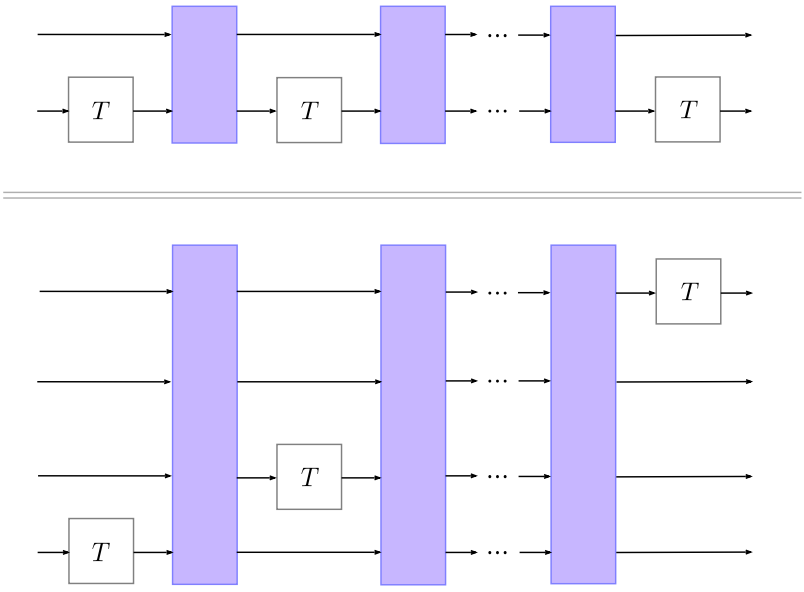

Figure 3.5: Two equivalent ways to visualize *general adaptive schemes*. The depiction on the bottom might make it easier to see that adaptive schemes include parallel ones.

Consequently, the subsequent $n-1$ uses of T would have to be capable of distinguishing the two cases upon input of $|\psi\rangle\langle\psi|$, contradicting the minimality of n. So ρ_A, ρ_B must have ranges with trivial intersection, proving the necessity of *(i)*.

In order to show necessity of *(ii)* consider the state of the overall system right before the last use of T and denote this state by ρ_A, ρ_B depending on whether T is T_A or T_B, respectively. Assuming the last use of T is not obsolete for perfect discrimination, we have that $\rho_A \not\perp \rho_B$ but $(T_A \otimes \mathrm{id})(\rho_A) \perp (T_B \otimes \mathrm{id})(\rho_B)$. Hence, there are unit vectors ψ_A and ψ_B in the range of ρ_A and ρ_B, respectively, such that $\langle \psi_A, \psi_B \rangle \neq 0$ while

$$\begin{aligned} 0 &= \mathrm{tr}\left[(T_A \otimes \mathrm{id})(|\psi_A\rangle\langle\psi_A|)(T_B \otimes \mathrm{id})(|\psi_B\rangle\langle\psi_B|)\right] \\ &= \sum_{k,l} \left|\langle \psi_A, (A_k^* B_l \otimes \mathbb{1})\psi_B \rangle\right|^2. \end{aligned} \tag{3.73}$$

Since all terms in Eq.(3.73) have to vanish individually, $\mathbb{1}$ cannot be contained in S since otherwise Eq.(3.73) would contradict $\langle \psi_A, \psi_B \rangle \neq 0$. This proves necessity of *(ii)*.

Sufficiency of *(i,ii)* will be proven by constructing an admissible procedure for discriminating the two cases. Since S is a subspace of $\mathcal{B}(\mathcal{H}_1)$, *(ii)* implies that $S^\perp \neq \{0\}$. Moreover, there must be an $M \in S^\perp$, with $\mathrm{tr}\,[M] \neq 0$. If this were not the case, i.e. if $S^\perp \subseteq \{\mathbb{1}\}^\perp$, then (by taking the orthogonal complement of

this inclusion) $S \supseteq \{\mathbb{1}\}$, which is excluded by *(ii)*. Furthermore, we can rescale M such that $\|M\|_2 = 1/\sqrt{d}$ with $d := \dim(\mathcal{H}_1)$. Defining a maximally entangled state $|\psi_A\rangle := \sum_k |kk\rangle/\sqrt{d} \in \mathcal{H}_1 \otimes \mathcal{H}_1$ and a unit vector $|\psi_B\rangle := (M \otimes \mathbb{1})|\psi_A\rangle$ this leads to

$$\mathrm{tr}\left[(T_A \otimes \mathrm{id})(|\psi_A\rangle\langle\psi_A|)(T_B \otimes \mathrm{id})(|\psi_B\rangle\langle\psi_B|)\right] = \tfrac{1}{d}\sum_{k,l} \left|\mathrm{tr}\left[A_k^* B_l M\right]\right|^2 = 0, \tag{3.74}$$

while $\langle\psi_A, \psi_B\rangle \neq 0$.

In order to exploit this, we use that *(i)* guarantees the existence of a unit vector $\varphi \in \mathcal{H}_1 \otimes \mathcal{H}_1$ such that the ranges of $\rho_A := (T_A \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)$ and $\rho_B := (T_B \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)$ have trivial intersection. Denoting the projections onto these ranges by P_A and P_B, respectively, this means that $\|P_A P_B\|_\infty < 1$. This enables us to pick an $m \in \mathbb{N}$ such that

$$\left\|P_A^{\otimes m} P_B^{\otimes m}\right\|_\infty = \|P_A P_B\|_\infty^m \leq |\langle\psi_A, \psi_B\rangle|. \tag{3.75}$$

As $P_A^{\otimes m}$ and $P_B^{\otimes m}$ are the projectors onto the ranges of $\rho_A^{\otimes m}$ and $\rho_B^{\otimes m}$, respectively, Prop.3.29 guarantees the existence of a quantum channel Φ that maps $\Phi : \rho_i^{\otimes m} \mapsto |\psi_i\rangle\langle\psi_i|$ for both $i \in \{A, B\}$. Finally, we can exploit Eq.(3.74), which implies that by using T one more time, i.e. $m+1$ times in total, we can perfectly discriminate T_A from T_B. □

Lemma 3.31. *For* $\dim(\mathcal{H}_1), \dim(\mathcal{H}_2) < \infty$ *two quantum channels* $T_A, T_B : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ *with corresponding sets of Kraus operators* $\{A_k\}_{k=1}^a, \{B_l\}_{l=1}^b \subset \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ *are entanglement assisted disjoint if* $\{A_k, B_l\}$ *is a set of* $a+b$ *linearly independent operators.*

Remark: Note that we can assume w.l.o.g. that $\{A_k\}$ is itself a set of linearly independent operators and that the same holds for $\{B_l\}$.

Proof. For any $\varphi \in \mathcal{H}_1 \otimes \mathcal{H}_1$, the range of $(T_A \otimes \mathrm{id})(|\varphi\rangle\langle\varphi|)$ is equal to $\mathcal{H}_A := \mathrm{span}\{(A_k \otimes \mathbb{1})\varphi\}_{k=1}^a$. If we choose φ such that it has maximal Schmidt rank, then $\dim(H_A) = a$ and $\dim(\mathcal{H}_B) = b$ where $\mathcal{H}_B := \mathrm{span}\{(B_l \otimes \mathbb{1})\varphi\}_{l=1}^b$. Moreover, if all $a+b$ Kraus operators are linearly independent, then $\dim(\mathcal{H}_A + \mathcal{H}_B) = a+b$ so that necessarily $\dim(\mathcal{H}_A \cap \mathcal{H}_B) = 0$. □

Example 3.7 (Adaptive vs. parallel discrimination)**.** Consider two quantum channels $T_A, T_B : \mathcal{B}(\mathbb{C}^3) \to \mathcal{B}(\mathbb{C}^3)$, where $T_B = \mathrm{id}$ and $T_A(\rho) := \sum_{i=1}^2 A_i \rho A_i^*$ with $A_1 := \mathrm{diag}(p_1, p_2, p_3)$, $A_2 := \sqrt{\mathbb{1} - A_1^2}$ and $0 < p_1 < p_2 < p_3 < 1$. The claim is that T_A, T_B can be discriminated perfectly using finitely many copies by an adaptive scheme but not via a parallel scheme. Let us first look into the general, adaptive case. Since the p_i's are distinct, the three operators $A_1, A_2, \mathbb{1}$ are linearly independent (as the vectors forming their diagonals are). This implies that both conditions of Thm.3.30 are satisfied: first, T_A, T_B are entanglement assisted disjoint by Lemma 3.31 and, second, $\mathbb{1} \notin S$. So perfect adaptive discrimination is feasible. However, as A_1 is clearly positive definite and contained in S, part *2.* of Thm.3.27 rules out the possibility of perfect parallel discrimination with a finite number of copies.

Exercise 3.11 (Diamond norm). Let $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ be a linear map between finite-dimensional spaces with Choi matrix C and $d := \dim(\mathcal{H}_1)$.

(a) Prove that

$$\|\Phi\|_\diamond = \sup_{\rho_1,\rho_2} \big\|(\mathbb{1} \otimes \sqrt{\rho_1})C(\mathbb{1} \otimes \sqrt{\rho_2})\big\|_1, \tag{3.76}$$

where the supremum is taken over all pairs of density operators ρ_1, ρ_2.

(b) Show that

$$\|\Phi\|_\diamond \leq \|C\|_1 \leq d\|\Phi\|_\diamond.$$

(c) Show that $\|\Phi\|_\diamond \leq d\|\Phi\|_1$.

(d) Let $\Phi(X) := X^T$ be the matrix transposition. Show that $\|\Phi\|_\diamond = d$.

(e) Let Ψ be another linear map between finite-dimensional matrix spaces. Show that

$$\|\Phi \otimes \Psi\|_\diamond = \|\Phi\|_\diamond \|\Psi\|_\diamond.$$

Is this true if the diamond norm is replaced by the induced trace-norm $\|\cdot\|_1$?

Exercise 3.12. Let $T_1, T_2 : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be quantum channels $\rho \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ a *separable* density operator (i.e. a convex combination of product states) and (M_1, M_2) a POVM on $\mathcal{H} \otimes \mathcal{H}$ that is built up from two measurements that act locally on the two subsystems and whose measurement outcomes are combined arbitrarily to two distinct sets that are eventually labeled 1 and 2. Show that

$$\mathrm{tr}\left[\big((T_1 - T_2) \otimes \mathrm{id}(\rho)\big) M_1\right] \leq \|T_1 - T_2\|_1. \tag{3.77}$$

How do you interpret this inequality?

Exercise 3.13 (Paths in the unitary group).

(a) Let $\gamma : [0,1] \to U(d)$ be a differentiable path in the unitary group. Show that for each $t \in (0,1)$ there is a Hermitian matrix $H(t)$ such that $\dot{\gamma}(t) = iH(t)\gamma(t)$.

(b) For $U, V \in U(d)$ unitary define the *geodesic distance* $g(U,V) := \inf_{\gamma,\alpha} \int_0^1 \|\dot{\gamma}(t)\|_\infty dt$ where the infimum is over all continuously differentiable paths $\gamma : [0,1] \to U(d)$ from $\gamma(0) = U$ to $\gamma(1) = Ve^{i\alpha}$ for some $\alpha \in \mathbb{R}$. Show that the geodesic distance equals the arc-length in the sense that $\Theta(U^*V) = 2g(U,V)$.

Hint: Argue that w.l.o.g. $U = \mathbb{1}$ and approximate the path by a finite number of pieces on which $H(t)$ (defined as in (a)) is independent of t.

Exercise 3.14 (Perfect discrimination of unitaries – again). Consider two quantum channels T_1, T_2 describing unitary time evolutions with unitaries $V_1, V_2 \in \mathcal{B}(\mathcal{H})$ to be discriminated.

(a) Use the general theorem for 'perfect parallel discrimination of quantum channels' in order to show that the two channels can be perfectly discriminated for a sufficiently large finite number of uses.

(b) Use the general theorem for 'perfect adaptive discrimination of quantum channels' in order to show that the two channels can be perfectly discriminated for a sufficiently large finite number of uses.

Exercise 3.15 (Perfect discrimination of measurement channels). Let $\{|k\rangle\}_{k=1}^{d}$ be an orthonormal basis for $\mathcal{H}$ and $U \in \mathcal{B}(\mathcal{H})$ a unitary that is not diagonal in this basis. Define two quantum channels $T_1, T_2 : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ as $T_1(\rho) := \sum_k |k\rangle\langle k|\rho|k\rangle\langle k|$ and $T_2(\rho) := T_1(U\rho U^*)$.

(a) Prove that T_1, T_2 can be perfectly discriminated using a finite number of copies and an adaptive scheme.

(b) Show that T_1, T_2 can be perfectly discriminated using a single copy iff there is a density matrix $\rho \in \mathcal{B}(\mathcal{H})$ such that all diagonal elements of $(U\rho)$ are zero (in the basis $\{|k\rangle\}$).

Exercise 3.16 (Perfect adaptive discrimination of quantum channels). Let $T_1, T_2 : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ be quantum channels with $\dim(\mathcal{H}_1) < \infty$.

(a) Provide a necessary criterion for the possibility of perfectly discriminating T_1 from T_2 with a finite number of uses in terms of the ranks of the Choi matrices of T_1 and T_2.

(b) Let $\sigma_0 := \mathbb{1} \in \mathbb{B}(\mathbb{C}^2)$ and $\{\sigma_i\}_{i=1}^{3}$ be the Pauli matrices. Define a quantum channel $T_1(\rho) := \sum_{i=0}^{3} p_i \sigma_i \rho \sigma_i$ from any probability vector p with non-zero entries. Can you construct a quantum channel $T_2 : \mathbb{B}(\mathbb{C}^2) \to \mathbb{B}(\mathbb{C}^2)$ that can be discriminated perfectly from T_1 with finitely many uses?

Notes and literature The diamond norm and its dual, the completely bounded norm, are discussed in detail in the books of Paulsen [90] and Watrous [62]. The latter is also the source of Example 3.5. The discrimination of unitaries goes back to Acin [91] who considered parallel schemes and was extended to sequential schemes by Duan, Feng and Ying in [92]. The proof of subadditivity (*4.* in Prop.3.21) of the spectral arc-length via Thompson's theorem [87] is inspired by [93], where geodesics have in this way been identified with one-parameter subgroups of the unitary group (see Exercise 3.13). For a partial extension of these results to infinite dimensions see [94].
Thm.3.27 on the perfect parallel discrimination of quantum channels was proven in [95]. The analogous theorem for more general adaptive discrimination (Thm.3.30) was proven in [89]. The discrimination of measurements with rank-one effects has been studied in [96, 97], where a close relation to the discrimination of unitaries was found, and in [98], where an explicit example (however, different from Exp.3.7) has been constructed that proves that adaptive strategies are more powerful than parallel ones.

3.2 Parameter estimation

Fisher information and Cramer-Rao bound Consider the task of estimating a parameter $\theta \in \mathbb{R}^n$ that influences an $\mathcal{X}$-valued random variable by observing a sample of the latter. We assume that we know an underlying parameter-dependent family of probability density functions $p_\theta : \mathcal{X} \to \mathbb{R}$ and that this is sufficiently regular. More specifically, we assume that the (co-) variance and expectation of the first and second derivatives of the *log-likelihood* function $\theta \mapsto \ln p_\theta(x)$ exist and that expectation and differentiations commute. On this basis, which we will tacitly assume throughout, we can define the following:

Definition 3.32 (Fisher information)**.** *The* Fisher information $I \in \mathbb{R}^{n\times n}$ *of a parameter-dependent probability density p_θ is defined at a particular parameter value $\theta = \theta_0$ in any of the following equivalent ways:*[5]

$$I_{kl} := \mathbb{E}\left[\Big(\frac{\partial}{\partial\theta_k}\ln p_\theta(x)\Big)\Big(\frac{\partial}{\partial\theta_l}\ln p_\theta(x)\Big)\right]\Big|_{\theta=\theta_0} \tag{3.78}$$

$$= \mathrm{Cov}\left[\frac{\partial}{\partial\theta_k}\ln p_\theta(x), \frac{\partial}{\partial\theta_l}\ln p_\theta(x)\right]\Big|_{\theta=\theta_0} \tag{3.79}$$

$$= -\mathbb{E}\left[\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\theta_l}\ln p_\theta(x)\right]\Big|_{\theta=\theta_0}, \tag{3.80}$$

where all expectations are w.r.t. $x \sim p_{\theta_0}$. Depending on how carefully we want to specify the dependencies, we will write $I = I(p_\theta) = I(p_\theta, \theta_0)$.

Proof. (that the three equations define the same.) Eq.(3.79) becomes Eq.(3.78) when noticing that the first-order terms vanish, i.e.:

$$\mathbb{E}\left[\frac{\partial}{\partial\theta_k}\ln p_\theta(x)\right] = \int_{\mathcal{X}} p_\theta(x)\frac{\partial}{\partial\theta_k}\ln p_\theta(x)\,dx \tag{3.81}$$

$$= \frac{\partial}{\partial\theta_k}\underbrace{\int p_\theta(x)\,dx}_{=1} = 0.$$

Differentiating the r.h.s. of Eq.(3.81) further w.r.t. θ_l, we obtain

$$0 = \frac{\partial}{\partial\theta_l}\int_{\mathcal{X}} p_\theta(x)\frac{\partial}{\partial\theta_k}\ln p_\theta(x)\,dx$$

$$= \int p_\theta(x)\Big(\frac{\partial}{\partial\theta_l}\ln p_\theta(x)\Big)\Big(\frac{\partial}{\partial\theta_k}\ln p_\theta(x)\Big)\,dx$$

$$+ \int p_\theta(x)\Big(\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\theta_l}\ln p_\theta(x)\Big)\,dx,$$

which shows equality between Eq.(3.78) and Eq.(3.80). □

[5] Recall that $\mathrm{Cov}[X, Y] := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

In order to form an intuition for the Fisher information consider the *maximum likelihood estimator* $\hat{\theta}(x) := \text{argmax}_\theta \, p_\theta(x)$ for a single parameter ($n = 1$). That is, we guess the value of the parameter to be the one for which the likelihood and therefore also the log-likelihood is maximal. The more peaked the log-likelihood function is around this maximum the more accurate can we expect this estimator to be. The Fisher information, as defined in Eq.(3.80), now quantifies this 'peakedness' via the negative second derivative.

Theorem 3.33 (Cramer-Rao bound)**.** *For $\hat{\theta} : \mathcal{X} \to \mathbb{R}^n$ and $\theta_0 \in \mathbb{R}^n$ define $C_{kl} := \frac{\partial}{\partial \theta_l} \mathbb{E}[\hat{\theta}_k]|_{\theta=\theta_0}$ and $\text{Cov}[\hat{\theta}]_{kl} := \text{Cov}[\hat{\theta}_k, \hat{\theta}_l]|_{\theta=\theta_0}$ where all expectations are w.r.t. $x \sim p_{\theta_0}$. Then for $I = I(p_\theta, \theta_0)$:*

$$\text{Cov}[\hat{\theta}] \geq C I^{-1} C^T, \tag{3.82}$$

where I^{-1} denotes the Moore-Penrose pseudo inverse if I is not invertible. In this case, in addition, $\ker[I] \subseteq \ker[I]$.

Quantum Fisher information

Quantum multi-parameter estimation

3.3 Tomography

In addition to the statistical questions familiar from hypothesis testing and parameter estimation, the task of tomography raises in addition the question about the size of the measurement setting. That is, how many types of measurements or how many different outcomes are required in order to identify a state? We begin with considering these types of questions that become non-trivial if a priori information is available that confines the state to a known subset of states $\mathcal{Q}$ (e.g., pure states, states with constrained rank or complexity, a given symmetry or known preparation history).

Injectivity, dimension, and topology Two often discussed ways of identifying a quantum state are: (i) in terms of the outcome statistics of a set of measurements or (ii) via the expectation values of a set of observables. For both cases, we define the notion of *informational completeness* with respect to an a priori given set of states:

Definition 3.34 (Informational completeness)**.** *Let $\mathcal{Q} \subset \mathcal{B}(\mathcal{H})$ be a set of density operators.*

- *A family $\{M^{(\lambda)}\}_{\lambda\in\Lambda}$ of POVMs on $\mathcal{B}(\mathcal{H})$ is called* informationally complete w.r.t *$\mathcal{Q}$ if for all $\rho_1, \rho_2 \in \mathcal{Q}$:*

$$\mathrm{tr}\left[\rho_1 M^{(\lambda)}(Y)\right] = \mathrm{tr}\left[\rho_2 M^{(\lambda)}(Y)\right] \ \forall Y \ \forall\lambda \quad \Rightarrow \quad \rho_1 = \rho_2.$$

- *A family $\{H_\lambda\}_{\lambda\in\Lambda} \subset \mathcal{B}(\mathcal{H})$ of Hermitian operators is called* informationally complete w.r.t *$\mathcal{Q}$ if for all $\rho_1, \rho_2 \in \mathcal{Q}$:*

$$\mathrm{tr}\left[\rho_1 H_\lambda\right] = \mathrm{tr}\left[\rho_2 H_\lambda\right] \ \forall\lambda \quad \Rightarrow \quad \rho_1 = \rho_2.$$

A simple but useful observation that allows us to relate different notions of informational completeness is the following:

Lemma 3.35. *Let $\mathcal{Q} \subset \mathcal{B}(\mathcal{H})$ be a set of density operators and $\mathcal{X} \subset \mathcal{B}(\mathcal{H})$ a set of Hermitian operators. Then the following are equivalent, if orthogonal complement and closure are understood within the Hilbert-Schmidt Hilbert space $\mathcal{B}_2(\mathcal{H})$:*

(i) *For $\rho_1, \rho_2 \in \mathcal{Q}$ we have:* $\mathrm{tr}\left[\rho_1 X\right] = \mathrm{tr}\left[\rho_2 X\right] \ \forall X \in \mathcal{X} \quad \Rightarrow \quad \rho_1 = \rho_2.$

(ii) $(\mathcal{Q} - \mathcal{Q}) \cap \left(\mathrm{span}\{\mathcal{X} \cup \{\mathbb{1}\}\}\right)^\perp = \{0\}.$

(iii) $(\mathcal{Q} - \mathcal{Q})^\perp + \overline{\mathrm{span}\{\mathcal{X} \cup \{\mathbb{1}\}\}} = \mathcal{B}_2(\mathcal{H}).$

Proof. $(ii) \Leftrightarrow (iii)$ follows by taking the orthogonal complement. For proving $\neg(i) \Leftrightarrow \neg(ii)$ let us assume there are distinct states $\rho_1, \rho_2 \in \mathcal{Q}$ for which

$\mathrm{tr}[\rho_1 X] = \mathrm{tr}[\rho_2 X]$ holds for all $X \in \mathcal{X}$. Since the ρ_i's are density operators this also holds for all $X \in \mathcal{X} \cup \{\mathbb{1}\}$, so that the statement becomes equivalent to

$$\rho_1 - \rho_2 \in \big(\mathcal{X} \cup \{\mathbb{1}\}\big)^\perp = \big(\mathrm{span}\big\{\mathcal{X} \cup \{\mathbb{1}\}\big\}\big)^\perp,$$

where the last step uses the fact that a set and its linear span have the same orthogonal complement. □

The point this Lemma is supposed to make is that informational completeness only depends on the space spanned by the operators of $\mathcal{X} \cup \{\mathbb{1}\}$. This is used in the following:

Proposition 3.36. *Let $\mathcal{Q} \subset \mathcal{B}(\mathcal{H})$ be a set of density operators and $n \in \mathbb{N}$. The following are equivalent:*

1. *There is an n-outcome POVM that is informationally complete w.r.t. $\mathcal{Q}$.*
2. *There is a set $\{H_1, \ldots, H_{n-1}\}$ of Hermitian operators that is informationally complete w.r.t. $\mathcal{Q}$.*
3. *There is a set of k POVMs $M^{(1)}, \ldots, M^{(k)}$ with $n_1, \ldots, n_k$ outcomes that is informationally complete w.r.t. $\mathcal{Q}$ where $n - 1 = \sum_{i=1}^k (n_i - 1)$.*

Proof. *1.⇒2.*: Let $(M_j)_{j=1}^n$ be the operators the POVM assigns to the n measurement outcomes. Define $H_j := M_j$ for $j \leq n-1$. Then $\{H_j\}_{j=1}^{n-1} \cup \{\mathbb{1}\}$ and $\{M_j\}_{j=1}^n \cup \{\mathbb{1}\}$ span the same space since $M_n = \mathbb{1} - \sum_{j=1}^{n-1} M_j$. So by Lemma 3.35 the set $\{H_j\}_{j=1}^{n-1}$ is informationally complete iff the initial POVM is.

2.⇒1.: Define $M_j := \frac{1}{2}(\mathbb{1} - \|H_j\|_\infty^{-1} H_j)/(n-1)$ for $j \leq n-1$ and $M_n := \mathbb{1} - \sum_{j=1}^{n-1} M_j$. By construction, $(M_j)_{j=1}^n$ forms a POVM and by Lemma 3.35 the information completeness of $\{H_j\}$ is inherited.

3.⇒1.: We construct an n-outcome POVM M by setting $\{M_j\}_{j=1}^{n-1} := \bigcup_{i=1}^k \big\{\frac{1}{k} M_l^{(i)}\big\}_{l=1}^{n_i-1}$ and $M_n := \mathbb{1} - \sum_{j=1}^{n-1} M_j$. As the collections of operators span the same space, Lemma 3.35 again guarantees information completeness of M.

1.⇒3.: We define the k POVMs such that $\bigcup_{i=1}^k \big\{M_l^{(i)}\big\}_{l=1}^{n_i-1} := \{M_j\}_{j=1}^{n-1}$ and $M_{n_i}^{(i)} := \mathbb{1} - \sum_{l=1}^{n_i-1} M_l^{(i)}$. □

Example 3.8 (Information completeness for all states)**.** Let $\mathcal{Q} \subset \mathcal{B}(\mathcal{H})$ be the set of all density operators on $\mathcal{H} = \mathbb{C}^d$. Since $(\mathcal{Q} - \mathcal{Q})^\perp$ is the one-dimensional space spanned by $\mathbb{1}$, and $\mathbb{C}^{d\times d}$ is d^2-dimensional, *(iii)* of Lemma 3.35 implies that a POVM $(M_i)_{i=1}^n$ is informationally complete w.r.t. all states iff $\mathrm{span}\{M_i\} = \mathbb{C}^{d\times d}$. Consequently, $n \geq d^2$ is necessary for informational completeness. Similarly, every informationally complete set of Hermitian operators $\{H_j\}$ must contain $d^2 - 1$ linearly independent operators (or d^2 linearly independent ones if their span contains $\mathbb{1}$).

Moreover, following *3.* in Prop. 3.36, any informationally complete set of binary measurements has to consist of at least $k = d^2 - 1$ measurements. For

$d = 2$ this means $k = 3$ with the measurements corresponding to the three Pauli matrices being an example.

Example 3.9 (Symmetric informationally complete (SIC)-POVMs)**.**
A POVM $(M_j)_{j=1}^{d^2}$ on $\mathbb{C}^d$ with $n = d^2$ outcomes is called *symmetric informationally complete POVM* (short SIC-POVM) if

(i) each M_j is of rank one,

(ii) $\exists a \in \mathbb{R}\ \forall j : \operatorname{tr}[M_j] = a$ and

(iii) $\exists b \in \mathbb{R}$: $\operatorname{tr}[M_i M_j] = b$ for all $i \neq j$.

The POVM requirements then determine the constants to be $a = 1/d$ and $b = a^2/(d+1)$. Moreover, the M_j's are then necessarily linearly independent as can be seen by observing that the Gram matrix $G_{ij} := \operatorname{tr}[M_i M_j]$ has full rank since all off-diagonal entries equal b whereas all diagonal entries equal a^2, which is larger than b. Consequently, the M_j's form a basis of $C^{d\times d}$ and are thus indeed informationally complete w.r.t. the set of all density matrices. SIC-POVMs are known in all dimensions up to the hundreds (at least numerically). Whether they exist for every $d \in \mathbb{N}$ is the content of *Zauner's conjecture*. For $d = 2$ the four corners of any regular tetrahedron inscribed in the Bloch sphere correspond (when properly normalized) to the four elements of a SIC-POVM.

Before looking at some explicit examples where $\mathcal{Q}$ is a strict subset of density operators, we will analyze the utility of randomly chosen measurements in terms of the 'dimension' of the set $\mathcal{Q}$. For this purpose, we introduce the following:

Definition 3.37 (Covering number & Minkowski dimension)**.** *For any bounded subset S of a finite-dimensional Banach space we define:*

- *The* covering number $N(\epsilon, S) := \min\left\{m \in \mathbb{N} \mid S \subseteq \bigcup_{i=1}^{m} B_\epsilon(x_i),\ x_i \in S\right\}$, *where $B_\epsilon(x_i)$ denotes the open ball of radius $\epsilon > 0$ around x_i.*
- *The* Minkowski dimension $\mathcal{D}(S) := \limsup_{\epsilon\to 0} \frac{\ln N(\epsilon,S)}{\ln(1/\epsilon)}$.

Remarks: Strictly speaking, this defines the *inner* covering number (as we require the centers of the balls to lie *in* S) and the *upper* Minkowski dimension (as we consider the lim sup). $\mathcal{D}(S)$ is also known as *upper box counting dimension*. Due to the equivalence of norms in finite dimensions, the choice of the norm does not influence $\mathcal{D}(S)$ and neither do other choices such as open vs. closed balls or inner vs. unconstrained covering number.

In order to understand the definition of the Minkowski dimension, it is helpful to observe that if S is a d-dimensional manifold, then $N(\epsilon, S)$ scales as $(1/\epsilon)^d$. The Minkowski dimension utilizes this to define the dimension for arbitrary sets while preserving the familiar notion of dimension where it is already defined otherwise. In fact, if S is a d-dimensional manifold, then $\mathcal{D}(S) = d$.

Theorem 3.38 (Generic informational completeness)**.** *Let $\mathcal{Q}$ be a closed set of density operators on a finite-dimensional Hilbert space $\mathcal{H}$. Then almost any*

(i.e., up to Lebesgue-measure zero) set of $m > 2\mathcal{D}(\mathcal{Q})$ Hermitian operators is informationally complete w.r.t. $\mathcal{Q}$.

Proof. For $n \in \mathbb{N}$ the set $S_n := \{\rho_1 - \rho_2 | \rho_1, \rho_2 \in \mathcal{Q} \wedge \|\rho_1 - \rho_2\|_1 > \frac{1}{n}\}$ has dimension $\mathcal{D}(S_n) \leq \mathcal{D}(\mathcal{Q} - \mathcal{Q}) \leq 2\mathcal{D}(\mathcal{Q}) < m$. Any set of Hermitian operators $H_1, \ldots, H_m$ can be represented by a linear map $h : x \mapsto h(x) := (\mathrm{tr}\,[H_1 x], \ldots, \mathrm{tr}\,[H_m x])$ from the Hermitian subspace of $\mathcal{B}(\mathcal{H})$ into the Euclidean space $\mathbb{R}^m$. Since informational completeness is unchanged under multiplication by a positive number, we can w.l.o.g. restrict ourselves to maps with operator norm $\|h\| \leq 1$. The set of operators represented by h is not informationally complete w.r.t. $\mathcal{Q}$ iff $h(x) = 0$ for some $x \in S_n$ and $n \in \mathbb{N}$. Since the union of countably many null sets is again a null set, it suffices to show that for every fixed $n \in \mathbb{N}$, the set of linear maps h whose kernel has a non-empty intersection with S_n is a null set.

In order to formalize a 'random' choice of h let μ be a probability measure on the operator-norm unit ball such that for any Hermitian y the distribution of $h(y)$ is uniform in the ball of radius $\|y\|$ in $\mathbb{R}^m$. In fact, any measure with respect to which the Lebesgue measure is absolutely continuous (i.e., has a density function) would work equally well. With respect to such a measure, we can now bound the probability that h is not informationally complete. To this end, we utilize an optimal ϵ-covering of S_n:

$$\begin{aligned}
\mathbb{P}_{h\sim\mu}\big[\exists x \in S_n : h(x) = 0\big] &\leq \sum_{i=1}^{N(\epsilon,S_n)} \mathbb{P}\big[\exists x \in B_\epsilon(x_i) : h(x) = 0\big] \\
&\leq \sum_{i=1}^{N(\epsilon,S_n)} \mathbb{P}\big[\exists x \in B_\epsilon(x_i) : \|h(x)\| < \epsilon\big] \\
&\leq \sum_{i=1}^{N(\epsilon,S_n)} \mathbb{P}\big[\|h(x_i)\| < 2\epsilon\big] \qquad (3.83) \\
&= \sum_{i=1}^{N(\epsilon,S_n)} \left(\frac{2\epsilon}{\|x_i\|}\right)^m \leq (2n)^m N(\epsilon, S_n)\epsilon^m. \qquad (3.84)
\end{aligned}$$

Here, the first inequality uses the union bound, Eq.(3.83) exploits that $\|h(x_i)\| \leq \|h(x_i - x)\| + \|h(x)\| \leq \|h\|\epsilon + \|h(x)\| < 2\epsilon$ and the first step in Eq.(3.84) follows from the fact that $h(x_i)$ is uniformly distributed within an m-dimensional Euclidean ball of radius $\|x_i\|$.

Finally, we can take the limit $\epsilon \to 0$ in Eq.(3.84) and complete the proof by inserting $\lim_{\epsilon\to 0} N(\epsilon, S_n)\epsilon^m = 0$, which follows from

$$\frac{\ln\big[N(\epsilon, S_n)\epsilon^m\big]}{\ln(1/\epsilon)} = \frac{\ln N(\epsilon, S_n)}{\ln(1/\epsilon)} - m \overset{\epsilon\to 0}{\longrightarrow} \mathcal{D}(S_n) - m < 0.$$

□

Proposition 3.39 (Informational completeness for pure states in $\mathbb{C}^3$)**.** *A POVM $(M_i)_{i=1}^n$ is informationally complete w.r.t. all pure states in $\mathbb{C}^3$ iff it falls into one of the following classes:*

(i) $\{M_i\}^\perp = \{0\}$ *(which is equivalent to informational completeness for all states and thus $n \geq 9$).*

(ii) $\{M_i\}^\perp = \mathrm{span}\{H\}$ *for some Hermitian operator H with* $\mathrm{tr}\,[H] = 0$ *and* $\det(H) \neq 0$*. In this case $n \geq 8$.*

Proof. Clearly, *(i)* is sufficient for informational completeness. So let us focus on the case where the M_i's do not span the whole space. We will first show that the existence of any nonzero, singular Hermitian $H \in \{M_i\}^\perp$ is in contradiction with informational completeness: as H must be traceless with one eigenvalue zero, it is of the form $H \propto |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \in (\mathcal{Q} - \mathcal{Q})$, which indeed contradicts informational completeness by *(ii)* of Lemma 3.35. Now suppose $\dim\{M_i\}^\perp \geq 2$. Then there are two linearly independent Hermitian operators $H_1, H_2 \in \{M_i\}^\perp$, which must be non-singular by the argument just made. By switching $H_2 \to -H_2$, if necessary, we can assume that $\det(H_1)$ and $\det(H_2)$ have opposite signs. However, by the intermediate value theorem, $H_t := tH_1 + (1-t)H_2 \in \{M_i\}^\perp$ has vanishing determinant for some $t \in [0,1]$, again contradicting informational completeness. Hence, *(ii)* is necessary (in case *(i)* is not already fulfilled).

Conversely, *(ii)* implies that every non-zero element of $\{M_i\}^\perp$ has rank three. Elements of $\mathcal{Q} - \mathcal{Q}$ in contrast have rank at most two. Consequently, $(\mathcal{Q} - \mathcal{Q}) \cap \{M_i\}^\perp = \{0\}$, which implies informational completeness by Lemma 3.35. □

The set of pure states on $\mathbb{C}^d$ is a smooth real manifold of dimension $2d-2$. In fact, it can be identified with the *complex projective space* $\mathbb{C}\mathbf{P}^{d-1}$. For the just discussed case $d = 3$, the dimension $\mathcal{D}(\mathcal{Q}) = 4$ thus differs from the number 8 of measurement outcomes that are minimally required for informational completeness by a factor of 2. In general, we know from Thm.3.38 that this gap can not be much larger than a factor of two. But where does it originate in the first place and how large is it for other choices of $\mathcal{Q}$? The answer to this lies in the topological properties of $\mathcal{Q}$.

Theorem 3.40 (Informational completeness and topological embeddings)**.** *Let $\mathcal{Q} \subset \mathcal{B}(\mathcal{H})$ be a closed subset of density operators. A POVM $(M_i)_{i=1}^n$ on $\mathcal{H}$ with $n \in \mathbb{N}$ outcomes is informationally complete w.r.t. $\mathcal{Q}$ iff the map $h : \mathcal{Q} \to \mathbb{R}^{n-1}$, $h(\rho) := (\mathrm{tr}\,[M_1\rho], \ldots, \mathrm{tr}\,[M_{n-1}\rho])$ is a topological embedding (i.e. a homeomorphism onto its image).*

Proof. A topological embedding is an injective continuous map with a continuous inverse on its image. Injectivity already implies informational completeness. Conversely, informational completeness implies injectivity. Moreover, since h is linear and bounded it is continuous so that $h : \mathcal{Q} \to h(\mathcal{Q})$ becomes a continuous bijection. As $\mathcal{Q}$ is compact, the inverse is continuous as well. So h is a topological embedding. □

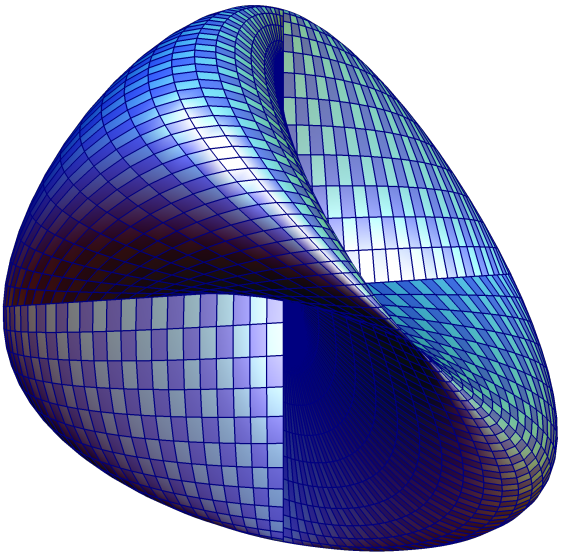

Figure 3.6: The *Roman surface* is obtained by displaying the first three dimensions of the embedding arising from Eq.(3.85) of the real projective plane $\mathbb{R}\mathbf{P}^2$. A continuous map from $\mathbb{R}\mathbf{P}^2$ into $\mathbb{R}^3$ that is injective (i.e., does not display self-intersections as the Roman surface) does not exist. This corresponds to the non-existence of an informationally complete POVM with four outcomes.

Corollary 3.41 (2-dimensional manifolds)**.** *Let $\mathcal{Q} \in \mathcal{B}(\mathcal{H})$ be a closed 2-dim. manifold without boundary in the set of density operators. Then every POVM that is informationally complete w.r.t. $\mathcal{Q}$ has at least 4 outcomes. If $\mathcal{Q}$ is in addition non-orientable, then 5 outcomes are necessary.*

Proof. The first statement follows from Thm.3.40 and the fact that $\mathcal{Q}$ as a compact two-dimensional manifold without boundary cannot be embedded into $\mathbb{R}^2$ where every compact two-dimensional manifold has a boundary. The second statement follows from Thm.3.40 together with the classification of two-dimensional manifolds by which non-orientability does not permit an embedding in $\mathbb{R}^3$. □

Example 3.10 (Pure states with real amplitudes in $\mathbb{C}^3$)**.** Let $\mathcal{Q}$ be the set of pure states in $\mathbb{C}^3$ with real amplitudes. That is, every element of $\mathcal{Q}$ can be represented by a unit vector $x \in \mathbb{R}^3$. Since x and $-x$, however, represent the same state, we have to identify antipodal points, and $\mathcal{Q}$ in this way becomes homeomorphic to the *real projective plane* $\mathbb{R}\mathbf{P}^2$. $\mathbb{R}\mathbf{P}^2$ is known to be a non-orientable, two-dimensional closed manifold without boundary. Hence, Cor.3.41 implies that at least 5 outcomes are necessary for any POVM to be informationally complete w.r.t. $\mathcal{Q}$. That 5 outcomes are also sufficient can be seen by using Prop.3.36 and realizing that $x \mapsto (x_1x_2, x_2x_3, x_3x_1, x_1^2 - x_2^2)$ is an injective map from $\mathbb{R}\mathbf{P}^2$

into $\mathbb{R}^4$ that can be obtained as the vector of expectation values of the four Hermitian matrices

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.85}$$

In some cases, there exists a deeper relation between informational completeness and the existence of *smooth* embeddings. Such a relation is the basis for the following result:

Theorem 3.42 (Pure-state information completeness)**.** *Let $\mathcal{Q} \in \mathcal{B}(\mathbb{C}^d)$ be the set of pure state density operators and $n(d)$ the minimum number of outcomes of any POVM that is informationally complete w.r.t. $\mathcal{Q}$. Then*

$$n(d) = 4d - 3 - c(d)\alpha(d),$$

where $c(d) \in [1, 2]$ and $\alpha(d)$ is the number of ones that appear in the binary expansion of $d - 1$.

Exercise 3.17. In a dusty corner on the engine deck, Chief Engineer Scotty finds three different Romulan quantum measurement devices, each capable of producing three different outcomes. Do you see a chance that Scotty can use these devices to perform quantum tomography that is informationally complete on the set of all pure states on $\mathbb{C}^3$?

Exercise 3.18. How many 3-outcome measurements (i.e. POVMs) are necessary for tomography that is informationally complete w.r.t. all states on a four-qubit system?

Exercise 3.19. A pure state $|\psi\rangle\langle\psi|$ on $\mathbb{C}^d \otimes \mathbb{C}^d$ is called *maximally entangled* if there is a unitary $U \in U(d)$ s.t. $|\psi\rangle \propto (U \otimes \mathbb{1}) \sum_{i=1}^d |i, i\rangle$. How many randomly chosen Hermitian operators on $\mathbb{C}^d \otimes \mathbb{C}^d$ are sufficient (with probability one) for tomography that is informationally complete w.r.t. all maximally entangled states?

Exercise 3.20. Prove that a POVM $(M_i)_{i=1}^n$ on $\mathbb{C}^d$ is informationally complete w.r.t. all pure states on $\mathbb{C}^d$ if and only if every non-zero Hermitian operator in $\left(\{M_i\}_{i=1}^n\right)^\perp$ has rank at least 3.

Exercise 3.21. Show that for $d = 2$ a POVM is informationally complete w.r.t. all pure states iff it is informationally complete w.r.t. all states.

Notes and literature The question about the minimal number of measurement settings or outcomes that are required to identify a quantum states with the help of prior information is addressed in numerous papers, e.g. [99, 100, 101, 102, 103, 104, 105, 106]. The results presented in this section are mainly from [104], where informational completeness has been studied in general and related to the existence of topological embeddings. This was then pursued further in [105] where the relation has been strengthened by incorporating notions of stability. Relations between dimension and generic injectivity similar to Thm. 3.38 (albeit outside quantum theory) have a long history dating back to Whitney's [107] (weak) embedding theorem for manifolds. The formulation in terms of the Minkowski dimension goes back to Mañé [108] (see also [109] for a textbook exposition). Hölder continuity of the corresponding inverse map has been shown in [110, 111]. While the results presented in this section merely deal with the existence of an injective map,

for the case of low-rank states, *compressed sensing* approaches [103] give slightly weaker bounds on the number of measurement settings/outcomes but lead to efficient inversion algorithms.

2-designs In this paragraph, we have a closer look at a special, highly symmetric type of POVMs that turn out to be close relatives of the SIC-POVMs discussed in Example 3.9. For that, we need the following classical theorem, which we will only use for the group $G = U(d)$ of unitaries but formulate in slightly more general terms:

Theorem 3.43 (Haar measure)**.** *For every compact Hausdorff topological group G there is a unique measure μ defined on the Borel σ-algebra, called* normalized Haar measure*, that satisfies*

(i) normalization*:* $\mu(G) = 1$

(ii) invariance*:* $\mu(gS) = \mu(S) = \mu(Sg)$ *for every* $g \in G$ *and Borel set* $S \subseteq G$

Together with the usual concept of Lebesgue integration, this defines the *normalized Haar integral* $\int f(g)d\mu(g)$ for every Borel measurable function $f : G \to \mathbb{C}$. In case of the unitary group $G = U(d)$ we will simply write $\int f(U)dU$. If f maps instead unit vectors from $\mathbb{C}^d$ into $\mathbb{C}$, we define

$$\int f(\psi)\, d\psi := \int f(U\varphi)\, dU, \tag{3.86}$$

where $\varphi \in \mathbb{C}^d$ is any unit vector. Invariance of the Haar integral then guarantees that the choice of φ does not influence the integral. As usual, integrals over a function f that has values in a finite-dimensional vector space are defined component-wise.

Definition 3.44 (t-design)**.** *For $t \in \mathbb{N}$, a set of unit-vectors $\psi_1, \ldots, \psi_m$ in $\mathbb{C}^d$ with associated probability distribution $(p_i)_{i=1}^m$ is called a* t-design *if*

$$\sum_{i=1}^{m} p_i \big(|\psi_i\rangle\langle\psi_i|\big)^{\otimes t} = \int \big(|\psi\rangle\langle\psi|\big)^{\otimes t} d\psi, \tag{3.87}$$

where the integral on the right is w.r.t. the normalized Haar measure. We call a t-design uniform *if $p_i = 1/m$ for all i.*

Remark: To distinguish the defined version from other *designs* (block designs, spherical designs, unitary designs, etc.), some authors sometimes attributed the above ones with the adjectives 'complex projective'. Non-uniform t-designs are also called 'generalized' or 'weighted' t-designs.

By taking the partial trace we see that every t-design is an s-design if $s \leq t$.

The operator in Eq.(3.87) is clearly supported on the *symmetric subspace* $\mathcal{H}_{\text{sym}}^{(t)} := \text{span}\{\psi^{\otimes t} |\ \psi \in \mathbb{C}^d\} \subseteq (\mathbb{C}^d)^{\otimes t}$. Denote by $P_{\text{sym}}^{(t)}$ the orthogonal projector onto $\mathcal{H}_{\text{sym}}^{(t)}$. Its trace is equal to the dimension of the space, which can be

shown to be

$$\mathrm{tr}\big[P_{\mathrm{sym}}^{(t)}\big] = \binom{t+d-1}{d-1}. \tag{3.88}$$

Due to the invariance of the Haar integral the r.h.s. of Eq.(3.87) is, in fact, equal to the normalized projector onto the symmetric subspace. That is:

Lemma 3.45.

$$\int |\psi\rangle\langle\psi|^{\otimes t} d\psi = \frac{P_{\mathrm{sym}}^{(t)}}{\mathrm{tr}\big[P_{\mathrm{sym}}^{(t)}\big]}. \tag{3.89}$$

This means in particular that

$$\int |\psi\rangle\langle\psi|\, d\psi = \frac{\mathbb{1}}{d}, \quad \text{and} \quad \int |\psi\rangle\langle\psi|^{\otimes 2} d\psi = \frac{\mathbb{1}+\mathbb{F}}{d(d+1)}. \tag{3.90}$$

When multiplying the left expression in Eq.(3.90) by d and replacing the Haar integral by the average over a t-design, we see that every t-design with n elements forms an n-outcome POVM $(M_i)_{i=1}^m$ via

$$M_i = d\, p_i\, |\psi_i\rangle\langle\psi_i|. \tag{3.91}$$

In order to arrive at a better understanding of the properties of POVMs that stem from 2-designs, we need the following Lemma:

Lemma 3.46. *If unit vectors $\psi_1, \ldots, \psi_m \in \mathbb{C}^d$ with associated probability distribution $(p_i)_{i=1}^m$ form a 2-design, then for every $A \in \mathbb{C}^{d\times d}$:*

$$\sum_{i=1}^m p_i |\psi_i\rangle\langle\psi_i| A |\psi_i\rangle\langle\psi_i| = \frac{1}{d(d+1)}\Big(\mathrm{tr}\,[A]\,\mathbb{1} + A\Big). \tag{3.92}$$

Proof. We will prove the identity by showing that both sides of the equation coincide under the trace with an arbitrary $X \in \mathbb{C}^{d\times d}$. Starting with the l.h.s. we obtain

$$\begin{aligned} \mathrm{tr}\left[(A\otimes X)\sum_i p_i |\psi_i\rangle\langle\psi_i|^{\otimes 2}\right] &= \frac{1}{d(d+1)}\mathrm{tr}\,[(A\otimes X)(\mathbb{1}+\mathbb{F})] \\ &= \frac{1}{d(d+1)}\Big(\mathrm{tr}\,[A]\,\mathrm{tr}\,[X] + \mathrm{tr}\,[AX]\Big), \end{aligned}$$

where the first equality stems from inserting the 2-design property from Eq.(3.90) and the second one uses that $\mathrm{tr}\,[(A\otimes X)\mathbb{F}] = \mathrm{tr}\,[AX]$. □

Theorem 3.47 (2-design vs. SIC-POVM)**.** *The following relations hold between 2-designs and POVMs on $\mathbb{C}^d$, with correspondence as in Eq.(3.91):*

1. *Every SIC-POVM corresponds to a uniform 2-design.*
2. *Every 2-design contains at least d^2 elements and the corresponding POVM is informationally complete w.r.t. all states.*

3. *Every 2-design with d^2 elements corresponds to a SIC-POVM. In particular, every such t-design is uniform (i.e., $p_i = 1/d^2$ for all i).*

Proof. *1.* From the SIC-POVM operators $(M_i)_{i=1}^m$ (where $m = d^2$) we define the rank-one projectors $|\psi_i\rangle\langle\psi_i| := \frac{1}{d}M_i$. We have to show that $\sum_i |\psi_i\rangle\langle\psi_i|^{\otimes 2} = d(\mathbb{1} + \mathbb{F})/(d+1)$. Since the SIC-POVM operators form a basis of $\mathbb{C}^{d\times d}$ it suffices to verify this identity under the trace with an operator of the form $|\psi_k\rangle\langle\psi_k| \otimes |\psi_l\rangle\langle\psi_l|$ for all k, l. Hence, this amounts to proving

$$\sum_i |\langle\psi_i, \psi_k\rangle|^2 |\langle\psi_i, \psi_l\rangle|^2 \stackrel{?}{=} d\big(1 + |\langle\psi_k, \psi_l\rangle|^2\big)/(d+1). \tag{3.93}$$

Equality in Eq.(3.93), however, is readily verified by inserting the central property of the SIC-POVM, namely that $|\langle\psi_i, \psi_k\rangle|^2 = 1/(d+1)$ for all $i \neq k$.

2. Consider the r.h.s. of Eq.(3.92). Varying over all $A \in \mathbb{C}^{d\times d}$ we obtain the d^2-dimensional vector space $\mathbb{C}^{d\times d}$. Consequently, the l.h.s., which is a linear combination of terms of the form $|\psi_i\rangle\langle\psi_i|$ must span the same space, i.e., $\mathrm{span}\{|\psi_i\rangle\langle\psi_i|\} = \mathbb{C}^{d\times d}$.

3. Consider the l.h.s. of Eq.(3.90). Replacing the Haar integral by the two-design and taking the trace of the square of the expression, we obtain

$$\frac{1}{d} = \sum_{i,j=1}^m p_i p_j |\langle\psi_i, \psi_j\rangle|^2 = \sum_{i\neq j} p_i p_j |\langle\psi_i, \psi_j\rangle|^2 + \sum_i p_i^2. \tag{3.94}$$

If we do the same with the r.h.s. of Eq.(3.90) we obtain the first line in

$$\begin{aligned} \frac{2}{d(d+1)} &= \sum_{i\neq j} p_i p_j |\langle\psi_i, \psi_j\rangle|^4 + \sum_i p_i^2 \\ &\geq \Big(\sum_{i\neq j} p_i p_j |\langle\psi_i, \psi_j\rangle|^2\Big)^2 \Big/ \underbrace{\Big(\sum_{i\neq j} p_i p_j\Big)}_{=1-q} + \underbrace{\sum_i p_i^2}_{=:q} \qquad (3.95) \\ &= \frac{\frac{1}{d^2} + q\big(1 - \frac{2}{d}\big)}{1-q} \;\geq\; \frac{2}{d(d+1)}. \qquad (3.96) \end{aligned}$$

Here, we have used the Cauchy-Schwarz inequality (w.r.t. a weighted inner product) to obtain Eq.(3.95). The first step in Eq.(3.96) follows from replacing $\sum_{i\neq j} p_i p_j |\langle\psi_i, \psi_j\rangle|^2$ via Eq.(3.94) and setting $q := \sum_i p_i^2$. The inequality in Eq.(3.96) stems from $q \geq 1/d^2$. Since both ends of the chain of inequalities are the same, equality has to hold in every step. Equality in the Chauchy-Schwarz inequality of Eq.(3.95) implies proportionality of the vectors, which means $|\langle\psi_i, \psi_j\rangle|^2 = \beta$ for all $i \neq j$ and some $\beta \in \mathbb{R}$. Equality in Eq.(3.96) implies that $p_i = 1/d^2$ for all i. Hence, by the definition of a SIC-POVM $\big(\frac{1}{d}|\psi_i\rangle\langle\psi_i|\big)_{i=1}^{d^2}$ is a SIC-POVM. □

For later use we show an implication of the uniform 2-design condition for measurements: the 2-norm distance of outcome distributions is, up to a multiplicative constant, equal to the operator 2-norm distance of the measured operators:

Lemma 3.48. *Let $\psi_1, \dots, \psi_m \in \mathbb{C}^d$ be unit-vectors forming a uniform 2-design, and $h : \mathbb{C}^{d\times d} \to \mathbb{C}^m$, $h(Z) := (\mathrm{tr}\,[ZM_1], \dots, \mathrm{tr}\,[ZM_m])$ with $M_i = \frac{d}{m}|\psi_i\rangle\langle\psi_i|$. Then for all $X, Y \in \mathbb{C}^{d\times d}$ of equal trace:*

$$\|h(X) - h(Y)\|_2 = \left[\frac{d}{m(d+1)}\right]^{1/2} \|X - Y\|_2 \tag{3.97}$$

Proof. We begin by exploiting linearity of h and define the traceless $Z := X - Y$:

$$\begin{aligned} \|h(X) - h(Y)\|_2^2 &= \left(\frac{d}{m}\right)^2 \sum_i \mathrm{tr}\left[(Z \otimes Z^*)\big(|\psi_i\rangle\langle\psi_i| \otimes |\psi_i\rangle\langle\psi_i|\big)\right] \\ &= \frac{d}{m(d+1)} \mathrm{tr}\left[(Z \otimes Z^*)(\mathbb{1} + \mathbb{F})\right] \qquad (3.98) \\ &= \frac{d}{m(d+1)}\Big(\big|\mathrm{tr}\,[Z]\,\big|^2 + \|Z\|_2^2\Big), \end{aligned}$$

where Eq.(3.98) uses the r.h.s. of Eq.(3.90) together with the 2-design property. Reinserting $Z = X - Y$ and using $\mathrm{tr}\,[Z] = 0$ then completes the proof. □

Notes and literature There is a vast literature on t-designs and SIC-POVMs and on the closely related topics of *frames* [112, 113] and *equiangular lines* [114, 115, 116]. Early investigations of (complex projective) t-designs can be found in the works of Hoggar [117] and, in the more general context of designs in compact metric spaces, in the works of Levenshtein [118]. The conjecture that 2-designs of size d^2 (a.k.a. SIC-POVMs) exist in every dimension, can be found in the 1999 PhD thesis of Zauner (translated in [119]). Zauner suggested constructing them as orbits of fiducial vectors under the action of the Weyl-Heisenberg group, which also makes the problem amenable to numerical investigations [120]. An analysis of 2-designs and SIC-POVMs from the perspective of quantum information theory can be found in [121, 122]. An in-depth discussion of the topic from the point of view of *finite tight frames* is provided in the text book by Waldron [123].

Least-squares estimators So far, our approach towards quantum tomography assumed exact knowledge of the probabilities of measurement outcomes—we have essentially tried to invert Born's rule. In practice, however, the probabilities are not measured exactly and can only be approximated by the corresponding observed frequencies of measurement outcomes. To be more precise, assume that n copies of a state $\rho \in \mathcal{B}(\mathcal{H})$ have been prepared and measured independently by a device that is described by a m-outcome POVM $(M_i)_{i=1}^m$. If the outcome $i \in \{1, \dots, m\}$ has thereby been obtained n_i times, we can use the frequency $f_i^{(n)} := n_i/n$ as an approximation to the probability $\mathrm{tr}\,[\rho M_i]$ to which it converges in the limit $n \to \infty$.

In this paragraph, we will discuss explicit ways of computing a *statistical estimator* $\hat{\rho}$ for the true state ρ from the frequencies $f := f^{(n)}$. A simple and useful choice is the *Hermitian least-squares estimator*

$$\hat{H} := \operatorname*{argmin}_{H=H^*} \hat{R}(H) \quad \text{where} \quad \hat{R}(H) := \sum_i \Big(f_i - \operatorname{tr}\left[HM_i\right] \Big)^2. \tag{3.99}$$

While $\hat{R}$ has a unique local minimum that is automatically its global minimum, the minimizer may in general not be unique. In this case, $\hat{H}$ denotes *a* minimizer and we may still specify a particular one via Eq.(3.101) of the following proposition:

Proposition 3.49 (Hermitian least-squares estimator)**.** *Every Hermitian least-squares estimator is a solution to the linear equation*

$$\sum_i \operatorname{tr}\left[HM_i\right] M_i = \sum_i f_i M_i. \tag{3.100}$$

With $C(X) := \big(\operatorname{tr}\left[XM_i\right]\big)_{i=1}^m$ *a solution can be expressed explicitly as*

$$\hat{H} = (C^*C)^{-1} C^* f, \tag{3.101}$$

where $(\cdot)^{-1}$ *means the pseudo-inverse (i.e., the inverse computed on the range). Moreover, if all* M_i *have the same trace, then every solution of Eq.(3.100) has trace one.*

Proof. Due to the quadratic nature of $\hat{R}$, $\hat{H}$ must be a solution of

$$0 \overset{!}{=} \nabla \hat{R}(H) = 2 \sum_i \big(f_i - \operatorname{tr}\left[HM_i\right] \big) M_i, \tag{3.102}$$

which proves Eq.(3.100). From here we can obtain Eq.(3.101) by using that $C^*(y) = \sum_i y_i M_i$ and realizing that Eq.(3.100) can be rewritten as $C^*C(H) = C^* f$. Finally, if all M_i have the same trace, we can show that every H that solves Eq.(3.100) has trace one, by taking the trace of Eq.(3.100) and exploiting $\sum_i f_i = 1$ and the POVM property $\sum_i M_i = \mathbb{1}$. □

In general, the computational cost of solving a linear equation scales cubic with the dimension. However, in special cases, the solution to the linear equation that yields a Hermitian least-squares estimator can be derived analytically. One such case is when the POVM corresponds to a uniform 2-design:

Proposition 3.50 (Hermitian least-squares estimators for uniform 2-designs)**.**

Let $\psi_1, \ldots, \psi_m \in \mathbb{C}^d$ *be unit vectors that form a uniform 2-design (i.e.* $\forall i$: $p_i = 1/m$*) and* $M_i := \frac{d}{m} |\psi_i\rangle\langle\psi_i|$*. Then the Hermitian least-squares estimator is unique and given by*

$$\hat{H} = \Big((d+1) \sum_i f_i \, |\psi_i\rangle\langle\psi_i| \Big) - \mathbb{1}. \tag{3.103}$$

Proof. Rewriting the l.h.s. of Eq.(3.100) with the help of Eq.(3.92), we obtain that every Hermitian least-squares estimator must solve

$$\frac{1}{d+1}\big(\mathbb{1}\mathrm{tr}\,[H] + H\big) = \sum_i f_i\, |\psi_i\rangle\langle\psi_i|.$$

Taking the trace of both sides of this equation shows that $\mathrm{tr}\,[H] = 1$ is necessary. Inserting this, we can solve for H and obtain Eq.(3.103) as a unique solution. □

A Hermitian least-squares estimator $\hat{H}$ is not necessarily positive. This motivates the definition and use of the so-called *projected least-squares (PLS) estimator* that is obtained as

$$\hat{\rho} := \operatorname*{argmin}_{\rho} \|\rho - \hat{H}\|_2, \tag{3.104}$$

where the minimum is taken over all density operators ρ. For any given $\hat{H}$, the PLS estimator is unique and can be regarded as the projection of $\hat{H}$ onto the closed convex set of density operators w.r.t. the Hilbert-Schmidt distance.

Instead of constructing a statistical estimator that is a density operator via this two-step procedure (first $\hat{H}$, then PLS) we can also imagine performing the minimization in Eq.(3.99) right away over the set of density operators. This leads to the so-called *positive least-squares* estimator. The latter is, in fact, equal to the PLS estimator for uniform 2-designs due to Lemma 3.48. In general, however, the two-step procedure can lead to a different estimator that can be obtained as follows:

Lemma 3.51. *Let $H \in \mathbb{C}^{d\times d}$ be Hermitian, diagonalized by a unitary U s.t. $H = U\mathrm{diag}(\lambda_1, \ldots, \lambda_d)U^*$. The density operator $\hat{\rho}$ that is closest to H in $\|\cdot\|_2$-norm is of the form*

$$\hat{\rho} = U\mathrm{diag}\big(\max\{0, \lambda_i - \alpha\}\big)_{i=1}^{d} U^*, \tag{3.105}$$

where $\alpha \in \mathbb{R}$ is s.t. $\mathrm{tr}\,[\hat{\rho}] = 1$.

Proof. (sketch): As the $\|\cdot\|_2$-norm is unitarily invariant, we can evaluate it in any basis, for instance in the basis in which $\hat{H}$ is diagonal. Then

$$\|\rho - \hat{H}\|_2^2 = \sum_{i,j} \big|\rho_{ij} - \delta_{i,j}\lambda_i\big|^2$$

shows that every non-zero off-diagonal element of ρ increases the norm unnecessarily so that we can assume $\rho_{ij} = \delta_{i,j}\mu_i$, i.e., ρ is diagonal in the same basis as $\hat{H}$. For the optimization of the eigenvalues of ρ we first parameterize $\mu_i = x_i^2$ with $x_i \in \mathbb{R}$ in order to incorporate the positivity constraint and then use a Lagrange multiplier Λ to minimize $f(x) := \sum_i (x_i^2 - \lambda_i)^2$ under the constraint $g(x) := \sum_i x_i^2 = 1$. Every minimizer then has to solve

$$0 \overset{!}{=} \nabla f(x) + \Lambda \nabla g(x), \tag{3.106}$$

which is equivalent to $2(x_i^2 - \lambda_i)x_i + \Lambda x_i = 0$ for all i. This has two types of solutions: either $x_i = 0$ (and thus $\mu_i = 0$) or

$$\mu_i = \lambda_i - \frac{\Lambda}{2}. \tag{3.107}$$

Hence, all non-zero eigenvalues are obtained by shifting the corresponding eigenvalue of $\hat{H}$ by the same constant $\alpha := \Lambda/2$. It remains to determine the set of eigenvalues that have to be set to zero. Unsurprisingly, these turn out to be the smallest ones (see [124] for details), which eventually results in Eq.(3.105). □

Notes and literature Projected least-squares estimators were studied in detail in [125] and compared to other statistical estimators in [126]. Lemma 3.51 is from [124].

Error bounds and confidence regions The error of a statistical estimator can be specified either in expectation or in terms of confidence regions. In the following we will have a look at both for the following simple quantum state tomography scenario:

Consider $m = d^2 - 1$ hermitian operators $H_1, \ldots, H_m \in \mathbb{C}^{d\times d}$ that we assume to be informationally complete w.r.t. the set of all density operators on $\mathbb{C}^d$, we define $h(\rho) := (\operatorname{tr}[H_1\rho], \ldots \operatorname{tr}[H_m\rho])$. Informational completeness demands that h is a linear bijection between the space of trace-less hermitian $d \times d$-matrices and R^m. Hence, an inverse map h^{-1} exists between these spaces.

Suppose that each observable H_i is measured n times, independently. That is, in total nm independent measurements are performed on nm independent and identically prepared systems, each described by the density operator ρ. Let f_i be the empirical estimate for $\operatorname{tr}[H_i\rho]$, obtained by averaging the outcomes of the n individual measurements. Abusing notation and denoting by f_i also the corresponding random variable, $\mathbb{E}[f_i] = \operatorname{tr}[H_i\rho]$ is the corresponding expectation value. Moreover, if H_i is the effect operator of a POVM, $\operatorname{tr}[H_i\rho]$ is the probability of obtaining the corresponding outcome, so that f_i follows a binomial distribution with variance $\operatorname{Var}[f_i] = \frac{1}{n}\operatorname{tr}[H_i\rho](1 - \operatorname{tr}[H_i\rho]) \leq \frac{1}{4n}$.

Consider an estimator $\hat{\rho}$ that is a minimizer of

$$\min_X \Big\{ \underbrace{\sum_{i=1}^{m} (\operatorname{tr}[H_i X] - f_i)^2}_{=\|h(X)-f\|_2^2} \;\Big|\; X = X^* \wedge \operatorname{tr}[X] = 1 \Big\}. \tag{3.108}$$

More explicitly, we choose $\hat{\rho} := \rho_0 + h^{-1}\big(f - h(\rho_0)\big)$, where ρ_0 is an arbitrary reference density matrix, e.g. $\rho_0 := \mathbb{1}/d$. This choice is hermitian, has unit trace and satisfies $\|h(\hat{\rho}) - f\|_2 = 0$ therefore minimizing Eq.(3.108). [6]

[6] The reason for not simply choosing $\hat{\rho} := h^{-1}(f)$ with h^{-1} mapping into the set of density matrices, is that we want h^{-1} to be a *linear map* in order to be able to talk about its operator norm (largest singular value) in the following proposition. As a linear map, however, it maps into the set of trace-zero matrices, so that an additional *offset* is necessary, and this is given by ρ_0.

Proposition 3.52. *In the above scenario, the distance between the true density matrix ρ and the estimator $\hat{\rho}$, which is computed from the outcomes of nm independent measurements, satisfies:*

$$\mathbb{E}\Big[\|\hat{\rho}-\rho\|_2^2\Big] \;\leq\; \|h^{-1}\|^2 \sum_{i=1}^{m} \mathrm{Var}\left[f_i\right] \;\leq\; \frac{\|h^{-1}\|^2 \Delta^2 m}{4n}, \tag{3.109}$$

where Δ is any number such that the spectrum of every H_i is contained in an interval of length at most Δ. $\|h^{-1}\|$ denotes the operator norm of h^{-1} when regarded as a linear map from $\mathbb{R}^m$ into the space of trace-less hermitian matrices, equipped with the Euclidean and Hilbert-Schmidt norm, respectively.

Remark: Note that $\mathrm{Var}\left[f_i\right]$ can be estimated from the measured data as the observed variance divided by n.

Proof. Inserting $\hat{\rho}$ and using linearity of the expectation value we obtain

$$\begin{aligned}
\mathbb{E}\Big[\|\hat{\rho}-\rho\|_2^2\Big] &= \mathbb{E}\Big[\|\rho_0 + h^{-1}\big(f-h(\rho_0)\big)-\rho\|_2^2\Big] \\
&= \mathbb{E}\Big[\|h^{-1}\big(f-h(\rho_0)\big)-h^{-1}h(\rho-\rho_0)\|_2^2\Big] \\
&\leq \|h^{-1}\|^2 \mathbb{E}\Big[\|f-h(\rho)\|_2^2\Big] \\
&= \|h^{-1}\|^2 \sum_{i=1}^{m} \underbrace{\mathbb{E}\Big[\big(f_i-\mathrm{tr}\left[H_i\rho\right]\big)^2\Big]}_{=\mathrm{Var}[f_i]}.
\end{aligned}$$

Since the variance does not change under translation, we can w.l.o.g. assume that the range of the outcomes of the i'th observable is $[-\frac{\Delta}{2},\frac{\Delta}{2}]$. The maximal variance under this constraint is $\frac{\Delta^2}{4}$, which is reduced by a factor $\frac{1}{n}$ when averaging n independent outcomes. □

Example 3.11. Let $(H_i)_{i=1}^m$ be Hilbert-Schmidt orthogonal hermitian operators with a common normalization, i.e. $\exists c\in\mathbb{R}\ \forall i,j : \mathrm{tr}\left[H_iH_j\right] = c\delta_{i,j}$. Examples would be the three Pauli matrices for $d=2$ or tensor products thereof for $d=2^k$. In this case, we have $hh^* = c\mathbb{1}$ so that $\|h^{-1}\|^2 = c^{-1}$. If in addition, as in the case of Pauli matrices, all observables satisfy $H_i^2 \propto \mathbb{1}$, we can choose $\Delta = 2\sqrt{c/d}$ so that

$$\mathbb{E}\Big[\|\hat{\rho}-\rho\|_2^2\Big] \;\leq\; \frac{m}{dn} \;\leq\; \frac{d}{n}\,. \tag{3.110}$$

In the above-specified context of quantum state tomography, a *confidence region* C with corresponding *confidence level* $\delta\in[0,1]$ is a subset of the $d\times d$ Hermitian matrices that is determined from the measured data and has the following property: if C is repeatedly determined in i.i.d. experiments that are all described by a density operator ρ, the probability that C contains ρ is at least $1-\delta$. That is:

$$\mathbb{P}\big[\rho\in C \mid \text{experiment is described by } \rho\big] \geq \delta. \tag{3.111}$$

Under the assumptions of Prop.3.52 we can derive a crude confidence region starting with the following bound:

$$\mathbb{P}\Big[\|\hat{\rho}-\rho\|_2^2 > \epsilon\Big] \quad \leq \quad \mathbb{P}\Big[\|h^{-1}\|^2 \|f-h(\rho)\|_2^2 > \epsilon\Big]. \tag{3.112}$$

For large n we may approximate the distribution of each f_i by a normal distribution with the same mean and variance. In this way, can replace each $(f_i - h(\rho)_i)^2$ by $X_i^2 \operatorname{Var}[f_i]$, where $X_i \sim \mathcal{N}(0,1)$ for $i = 1, \ldots, m$ are i.i.d. normal random variables. Invoking again the uniform bound $\operatorname{Var}[f_i] \leq \Delta^2/(4n)$, we thus obtain

$$\mathbb{P}\Big[\|\hat{\rho}-\rho\|_2^2 > \epsilon\Big] \leq \mathbb{P}\left[\sum_{i=1}^m X_i^2 > \frac{4n\epsilon}{\Delta^2 \|h^{-1}\|^2}\right]. \tag{3.113}$$

The r.h.s. of Eq.(3.113) is the survival function (i.e., one minus the cumulative distribution function) of the χ^2_m-distribution. An explicit tail bound is for instance given for any $a > m$ by

$$\mathbb{P}\Big[\sum_{i=1}^m X_i^2 > a\Big] \leq \exp\big[-\tfrac{m}{2}\big(\tfrac{a}{m} - 1 - \ln\tfrac{a}{m}\big)\big]. \tag{3.114}$$

Notes and literature Different confidence regions for quantum state tomography are discussed and compared in [127]. Eq.(3.114) is taken from [128]).

Bibliography

[1] S. Narayanan and J. Nelson, "Optimal terminal dimensionality reduction in euclidean space," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, (New York, NY, USA), p. 1064–1069, Association for Computing Machinery, 2019.

[2] D. Hilbert, J. v. Neumann, and L. Nordheim, "Über die Grundlagen der Quantenmechanik," *Mathematische Annalen*, vol. 98, pp. 1–30, Mar 1928.

[3] P. Halmos, *A Hilbert Space Problem Book*. Graduate Texts in Mathematics, Springer, 1982.

[4] K. R. Davidson, "Normal operators are diagonal plus Hilbert Schmidt," *Journal of Operator Theory*, vol. 20, no. 2, pp. 241–249, 1988.

[5] G. Pedersen, *Analysis Now*. Graduate Texts in Mathematics, Springer New York, 2012.

[6] J. Conway, *A Course in Operator Theory*. Graduate studies in mathematics, American Mathematical Society, 2013.

[7] B. Simon, *A Comprehensive Course in Analysis: Operator theory*. A Comprehensive Course in Analysis, American Mathematical Society, 2015.

[8] N. Dunford and J. Schwartz, *Linear Operators, Part 1: General Theory*. Wiley Classics Library, Wiley, 1988.

[9] J. Arazy, "More on convergence in unitary matrix spaces," *Proceedings of the American Mathematical Society*, vol. 83, no. 1, pp. 44–48, 1981.

[10] J. Von Neumann, *Mathematische Grundlagen der Quantenmechanik*. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen, J. Springer, 1932.

[11] G. F. Dell'Antonio, "On the limits of sequences of normal states," *Communications on Pure and Applied Mathematics*, vol. 20, no. 2, pp. 413–429, 1967.

[12] G. Ludwig, *Die Grundlagen der Quantenmechanik*. Grundlehren der mathematischen Wissenschaften, 1954.

[13] E. Davies, *Quantum Theory of Open Systems*. Academic Press, 1976.

[14] A. Nayak, "Optimal lower bounds for quantum automata and random access codes," in *40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039)*, pp. 369–376, 1999.

[15] C. W. Helstrom, "Quantum detection and estimation theory," *Journal of Statistical Physics*, vol. 1, no. 2, pp. 231–252, 1969.

[16] A. Holevo, "Statistical decision theory for quantum systems," *Journal of Multivariate Analysis*, vol. 3, no. 4, pp. 337–394, 1973.

[17] V. P. Belavkin, "Optimum distinction of non-orthogonal quantum signals," *Radio Engineering and Electronic Physics*, vol. 20, pp. 1177–1185, June 1975.

[18] V. P. Belavkin, "Optimal multiple quantum statistical hypothesis testing," *Stochastics*, vol. 1, no. 1-4, pp. 315–345, 1975.

[19] H. Yuen, R. Kennedy, and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 125–134, 1975.

[20] C. A. McCarthy, "Cp," *Israel Journal of Mathematics*, vol. 5, pp. 249–271, Oct 1967.

[21] B. Simon, *Convexity: An Analytic Viewpoint*. Cambridge Tracts in Mathematics, Cambridge University Press, 2011.

[22] R. Rockafellar, *Convex Analysis*. Princeton Landmarks in Mathematics and Physics, Princeton University Press, 1997.

[23] M. Ito and B. F. Lourenço, "A bound on the carathéodory number," *Linear Algebra and its Applications*, vol. 532, pp. 347–363, 2017.

[24] R. V. Kadison and G. K. Pedersen, "Means and convex combinations of unitary operators," *Mathematica Scandinavica*, vol. 57, no. 2, pp. 249–266, 1985.

[25] G. Pisier, "Remarques sur un résultat non publié de B. Maurey," *Séminaire d'Analyse fonctionnelle (dit "Maurey-Schwartz")*, 1980-1981. talk:5.

[26] G. Ivanov, "Approximate carathéodory's theorem in uniformly smooth banach spaces," *Discrete & Computational Geometry*, vol. 66, pp. 273–280, Jul 2021.

[27] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *Journal of Mathematical Physics*, vol. 14, pp. 1938–1941, 1973.

[28] M. E. Shirokov, "Continuity of the von neumann entropy," *Communications in Mathematical Physics*, vol. 296, pp. 625–654, Jun 2010.

[29] B. Simon, *Trace Ideals and Their Applications*. Mathematical surveys and monographs, American Mathematical Society, 2005.

[30] B. Schumacher and M. D. Westmoreland, "Approximate quantum error correction," *Quantum Information Processing*, vol. 1, pp. 5–12, Apr 2002.

[31] A. Winter, "Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints," *Communications in Mathematical Physics*, vol. 347, pp. 291–313, Oct 2016.

[32] T.-C. Lin, I. H. Kim, and M.-H. Hsieh, "An operator extension of weak monotonicity," 2023.

[33] E. A. Carlen and E. H. Lieb, "Bounds for entanglement via an extension of strong subadditivity of entropy," *Letters in Mathematical Physics*, vol. 101, pp. 1–11, Jul 2012.

[34] E. A. Carlen and E. H. Lieb, "Remainder terms for some quantum entropy inequalities," *Journal of Mathematical Physics*, vol. 55, no. 4, 2014.

[35] J. Aczél, B. Forte, and C. T. Ng, "Why the shannon and hartley entropies are 'natural'," *Advances in Applied Probability*, vol. 6, no. 1, p. 131–146, 1974.

[36] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[37] J. Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," *American Journal of Mathematics*, vol. 58, no. 1, pp. 141–163, 1936.

[38] W. Heisenberg, "Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," *Zeitschrift für Physik*, vol. 43, pp. 172–198, Mar 1927.

[39] E. H. Kennard, "Zur quantenmechanik einfacher bewegungstypen," *Zeitschrift für Physik*, vol. 44, pp. 326–352, Apr 1927.

[40] H. P. Robertson, "An indeterminacy relation for several observables and its classical interpretation," *Phys. Rev.*, vol. 46, pp. 794–801, Nov 1934.

[41] J. L. Barnes, *Laplace-Fourier Transformation, the Foundation for Quantum Information Theory and Linear Physics*, pp. 157–174. Princeton: Princeton University Press, 1971.

[42] A. Wigderson and Y. Wigderson, "The uncertainty principle: Variations on a theme," *Bulletin of the American Mathematical Society*, vol. 58, pp. 225–261, Jan. 2021.

[43] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[44] S. Deffner and S. Campbell, "Quantum speed limits: from heisenberg's uncertainty principle to optimal quantum control," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, p. 453001, oct 2017.

[45] "The uncertainty relation between energy and time in non-relativistic quantum mechanics," *J.Phys. (USSR)*, vol. 9, 1945.

[46] N. Margolus and L. B. Levitin, "The maximum speed of dynamical evolution," *Physica D: Nonlinear Phenomena*, vol. 120, no. 1, pp. 188–195, 1998. Proceedings of the Fourth Workshop on Physics and Consumption.

[47] S. Luo and Z. Zhang, "Estimating the first zero of a characteristic function," *Comptes Rendus Mathematique*, vol. 338, no. 3, pp. 203–206, 2004.

[48] D. Reeb and M. M. Wolf, "An improved landauer principle with finite-size corrections," *New Journal of Physics*, vol. 16, p. 103011, oct 2014.

[49] A. Frigerio, V. Gorini, and M. Verri, "The zeroth law of thermodynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 137, no. 3, pp. 573–602, 1986.

[50] M. Fannes, "A continuity property of the entropy density for spin lattice systems," *Communications in Mathematical Physics*, vol. 31, pp. 291–294, Dec 1973.

[51] K. M. R. Audenaert, "A sharp continuity estimate for the von neumann entropy," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, p. 8127, jun 2007.

[52] S. Becker, N. Datta, and M. G. Jabbour, "From classical to quantum: Uniform continuity bounds on entropies in infinite dimensions," *IEEE Transactions on Information Theory*, 2023.

[53] F. Hansen, J. Pecaric, and I. Peric, "Jensen's operator inequality and its converses," *Mathematica Scandinavica*, vol. 100, no. 1, pp. 61–73, 2007.

[54] A. Montanaro, "A lower bound on the probability of error in quantum state discrimination," in *2008 IEEE Information Theory Workshop*, pp. 378–380, 2008.

[55] K. M. R. Audenaert and M. Mosonyi, "Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination," *Journal of Mathematical Physics*, vol. 55, no. 10, p. 102201, 2014.

[56] D. Bacon, I. L. Chuang, and A. W. Harrow, "Efficient quantum circuits for schur and clebsch-gordan transforms," *Phys. Rev. Lett.*, vol. 97, p. 170502, Oct 2006.

[57] K. Li, "Discriminating quantum states: The multiple Chernoff distance," *The Annals of Statistics*, vol. 44, no. 4, pp. 1661 – 1679, 2016.

[58] A. S. Holevo, "Note on optimal quantum measurements," *Probl. Peredachi Inf.*, vol. 10, no. 4, pp. 51–55, 1974.

[59] S. M. Barnett and S. Croke, "On the conditions for discrimination between quantum states with minimum error," *Journal of Physics A: Mathematical and Theoretical*, vol. 42, no. 6, p. 062001, 2009.

[60] P. Hausladen and W. K. Wootters, "A 'pretty good' measurement for distinguishing quantum states," *Journal of Modern Optics*, vol. 41, no. 12, pp. 2385–2390, 1994.

[61] H. Barnum and E. Knill, "Reversing quantum dynamics with near-optimal quantum and classical fidelity," *Journal of Mathematical Physics*, vol. 43, no. 5, pp. 2097–2106, 2002.

[62] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.

[63] T. Ogawa and H. Nagaoka, "Strong converse to the quantum channel coding theorem," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2486–2489, 1999.

[64] M. Ježek, J. Řeháček, and J. Fiurášek, "Finding optimal strategies for minimum-error quantum-state discrimination," *Phys. Rev. A*, vol. 65, p. 060301, Jun 2002.

[65] J. Tyson, "Error rates of Belavkin weighted quantum measurements and a converse to Holevo's asymptotic optimality theorem," *Phys. Rev. A*, vol. 79, p. 032343, Mar 2009.

[66] J. Tyson, "Two-sided estimates of minimum-error distinguishability of mixed quantum states via generalized Holevo–Curlander bounds," *Journal of Mathematical Physics*, vol. 50, no. 3, p. 032106, 2009.

[67] P. M. Hayden, D. W. Leung, and G. Smith, "Multiparty data hiding of quantum information," *Physical Review A*, vol. 71, p. 062339, 2005.

[68] A. W. Harrow and A. Winter, "How many copies are needed for state discrimination?," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 1–2, 2012.

[69] A. Montanaro, "Pretty simple bounds on quantum state discrimination," 2019.

[70] H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations," *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493 – 507, 1952.

[71] T. Ogawa and M. Hayashi, "On error exponents in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1368–1372, 2004.

[72] M. Nussbaum and A. Szkoła, "The chernoff lower bound for symmetric quantum hypothesis testing," *The Annals of Statistics*, vol. 37, no. 2, pp. 1040–1057, 2009.

[73] K. Audenaert, J. Calsamiglia, R. Munoz Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, "Discriminating states: The quantum chernoff bound," *Physical review letters*, vol. 98, p. 160501, 05 2007.

[74] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," *Communications in Mathematical Physics*, vol. 279, pp. 251–283, Apr 2008.

[75] V. Jaksic, Y. Ogata, C.-A. Pillet, and R. Seiringer, "Quantum hypothesis testing and non-equilibrium statistical mechanics," *Reviews in Mathematical Physics*, vol. 24, no. 06, p. 1230002, 2012.

[76] V. Kargin, "On the Chernoff bound for efficiency of quantum hypothesis testing," *The Annals of Statistics*, vol. 33, no. 2, pp. 959 – 976, 2005.

[77] J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acín, and E. Bagan, "Quantum chernoff bound as a measure of distinguishability between density matrices: Application to qubit and gaussian states," *Physical Review A*, vol. 77, p. 032311, 2008.

[78] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99 – 114, 1991.

[79] T. Ogawa and H. Nagaoka, "Strong converse and stein's lemma in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2428–2433, 2000.

[80] M. Hayashi, "Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding," *Phys. Rev. A*, vol. 76, p. 062301, Dec 2007.

[81] H. Nagaoka, "The converse part of the theorem for quantum Hoeffding bound," 2006.

[82] S. Slussarenko, M. M. Weston, J.-G. Li, N. Campbell, H. M. Wiseman, and G. J. Pryde, "Quantum state discrimination using the minimum average number of copies," *Phys. Rev. Lett.*, vol. 118, p. 030502, Jan 2017.

[83] E. Martínez Vargas, C. Hirche, G. Sentís, M. Skotiniotis, M. Carrizo, R. Muñoz Tapia, and J. Calsamiglia, "Quantum sequential hypothesis testing," *Phys. Rev. Lett.*, vol. 126, p. 180502, May 2021.

[84] Y. Li, V. Y. F. Tan, and M. Tomamichel, "Optimal adaptive strategies for sequential quantum hypothesis testing," in *2021 IEEE Information Theory Workshop (ITW)*, p. 1–6, IEEE Press, 2021.

[85] K. Gustafson, "The Toeplitz-Hausdorff theorem for linear operators," *Proc. Amer. Math. Soc.*, no. 25, pp. 203–204, 1970.

[86] A. Brown and C. Pearcy, "Spectra of tensor products of operators," *Proceedings of the American Mathematical Society*, vol. 17, no. 1, pp. 162–166, 1966.

[87] R. C. Thompson, "Proof of a conjectured exponential formula," *Linear and Multilinear Algebra*, vol. 19, no. 2, pp. 187–197, 1986.

[88] A. Uhlmann, "The transition probability for states of *-algebras," *Annalen der Physik*, vol. 497, no. 4-6, pp. 524–532, 1985.

[89] R. Duan, Y. Feng, and M. Ying, "Perfect distinguishability of quantum operations," *Phys. Rev. Lett.*, vol. 103, p. 210501, Nov 2009.

[90] V. Paulsen, *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2003.

[91] A. Acín, "Statistical distinguishability between unitary operations," *Phys. Rev. Lett.*, vol. 87, p. 177901, Oct 2001.

[92] R. Duan, Y. Feng, and M. Ying, "Entanglement is not necessary for perfect discrimination between unitary operations," *Phys. Rev. Lett.*, vol. 98, p. 100503, Mar 2007.

[93] J. Antezana, G. Larotonda, and A. Varela, "Optimal paths for symmetric actions in the unitary group," *Communications in Mathematical Physics*, vol. 328, pp. 481–497, Jun 2014.

[94] J. Antezana, G. Larotonda, and A. Varela, "Thompson-type formulae," *Journal of Functional Analysis*, vol. 262, no. 4, pp. 1515–1528, 2012.

[95] R. Duan, C. Guo, C.-K. Li, and Y. Li, "Parallel distinguishability of quantum operations," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2259–2263, 2016.

[96] Z. Puchała, Ł. Pawela, A. Krawiec, and R. Kukulski, "Strategies for optimal single-shot discrimination of quantum measurements," *Phys. Rev. A*, vol. 98, p. 042103, Oct 2018.

[97] Z. Puchała, Ł. Pawela, A. Krawiec, R. Kukulski, and M. Oszmaniec, "Multiple-shot and unambiguous discrimination of von Neumann measurements," *Quantum*, vol. 5, p. 425, Apr. 2021.

[98] A. Krawiec, Ł. Pawela, and Z. Puchała, "Discrimination of povms with rank-one effects," *Quantum Information Processing*, vol. 19, p. 428, Nov 2020.

[99] S. Weigert, "Pauli problem for a spin of arbitrary length: A simple method to determine its wave function," *Physical review. A*, vol. 45, pp. 7688–7696, 07 1992.

[100] J.-P. Amiet and S. Weigert, "Reconstructing a pure state of a spin s through three stern-gerlach measurements," *Journal of Physics A: Mathematical and General*, vol. 32, pp. 2777–2784, jan 1999.

[101] J. Finkelstein, "Pure-state informationally complete and "really" complete measurements," *Phys. Rev. A*, vol. 70, p. 052107, Nov 2004.

[102] S. T. Flammia, A. Silberfarb, and C. M. Caves, "Minimal informationally complete measurements for pure states," *Foundations of Physics*, vol. 35, pp. 1985–2006, Dec 2005.

[103] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, "Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators," *New Journal of Physics*, vol. 14, p. 095022, sep 2012.

[104] T. Heinosaari, L. Mazzarella, and M. M. Wolf, "Quantum tomography under prior information," *Communications in Mathematical Physics*, vol. 318, pp. 355–374, Mar 2013.

[105] M. Kech, P. Vrana, and M. M. Wolf, "The role of topology in quantum tomography," *Journal of Physics A: Mathematical and Theoretical*, vol. 48, p. 265303, jun 2015.

[106] M. Kech and M. M. Wolf, "Constrained quantum tomography of semi-algebraic sets with applications to low-rank matrix recovery," *Information and Inference: A Journal of the IMA*, vol. 6, pp. 171–195, 12 2016.

[107] H. Whitney, "Differentiable manifolds," *Annals of Mathematics*, vol. 37, no. 3, pp. 645–680, 1936.

[108] R. Mañé, "On the dimension of the compact invariant sets of certain non-linear maps," in *Dynamical Systems and Turbulence, Warwick 1980* (D. Rand and L.-S. Young, eds.), (Berlin, Heidelberg), pp. 230–242, Springer Berlin Heidelberg, 1981.

[109] J. C. Robinson, *Dimensions, Embeddings, and Attractors*. Cambridge Tracts in Mathematics, Cambridge University Press, 2010.

[110] C. Foias and E. Olson, "Finite fractal dimension and hölder-lipschitz parametrization," *Indiana University Mathematics Journal*, vol. 45, no. 3, pp. 603–616, 1996.

[111] A. Margaris and J. C. Robinson, "Embedding properties of sets with finite box-counting dimension," *Nonlinearity*, vol. 32, pp. 3523–3547, aug 2019.

[112] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic fourier series," *Transactions of the American Mathematical Society*, vol. 72, no. 2, pp. 341–366, 1952.

[113] J. J. Benedetto and M. Fickus, "Finite normalized tight frames," *Advances in Computational Mathematics*, vol. 18, pp. 357–385, Feb 2003.

[114] P. Lemmens and J. Seidel, "Equiangular lines," *Journal of Algebra*, vol. 24, no. 3, pp. 494–512, 1973.

[115] P. Delsarte, J. Goethals, and J. Seidel, "Bounds for systems of lines, and jacobi polynomials," in *Geometry and Combinatorics* (D. Corneil and R. Mathon, eds.), pp. 193–207, Academic Press, 1991.

[116] B. C. Stacey, "Maximal sets of equiangular lines," 2020.

[117] S. Hoggar, "t-designs in projective spaces," *European Journal of Combinatorics*, vol. 3, no. 3, pp. 233–254, 1982.

[118] V. Levenshtein, "On designs in compact metric spaces and a universal bound on their size," *Discrete Mathematics*, vol. 192, no. 1, pp. 251–271, 1998.

[119] G. Zauner, "Quantum designs: Foundations of a noncommutative design theory," *International Journal of Quantum Information*, vol. 09, no. 01, pp. 445–507, 2011.

[120] A. J. Scott and M. Grassl, "Symmetric informationally complete positive-operator-valued measures: A new computer study," *Journal of Mathematical Physics*, vol. 51, no. 4, p. 042203, 2010.

[121] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, "Symmetric informationally complete quantum measurements," *Journal of Mathematical Physics*, vol. 45, no. 6, pp. 2171–2180, 2004.

[122] A. J. Scott, "Tight informationally complete quantum measurements," *Journal of Physics A: Mathematical and General*, vol. 39, pp. 13507–13530, oct 2006.

[123] S. Waldron, *An Introduction to Finite Tight Frames*. Applied and Numerical Harmonic Analysis, Springer New York, 2019.

[124] J. A. Smolin, J. M. Gambetta, and G. Smith, "Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise," *Phys. Rev. Lett.*, vol. 108, p. 070502, Feb 2012.

[125] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp, "Fast state tomography with optimal error bounds," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, p. 204001, apr 2020.

[126] A. Acharya, T. Kypraios, and M. Guţă, "A comparative study of estimation methods in quantum tomography," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, p. 234001, may 2019.

[127] J. O. de Almeida, M. Kleinmann, and G. Sentís, "Comparison of confidence regions for quantum state tomography," 2023.

[128] M. Ghosh, "Exponential tail bounds for chisquared random variables," *Journal of Statistical Theory and Practice*, vol. 15, p. 35, Mar 2021.