Mathematical Introduction to Quantum Information Processing

(growing lecture notes, SS2019)

Michael M. Wolf

June 22, 2019

Contents

These are (incomplete but hopefully growing) lecture notes of a course taught in summer 2019 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 a property that distinguishes Hilbert spaces from almost all other normed spaces and has various applications in the form of a *dimension reduction* argument:

Lemma 1.2 (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*

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

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

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

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

This is often stated and used for real Hilbert spaces, but 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 $l_2(\mathbb{N})$.

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*".
A good way to learn about the mathematics of Hilbert spaces is from Paul Halmos' "*A Hilbert space problem book*".

1.2 Bounded 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 1.3. *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 $\langle e_k, A^* e_l \rangle = \overline{\langle A^* e_l, e_k \rangle} = \overline{\langle e_l, A e_k \rangle}$.

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.

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.

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

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 $|A| = A_+ + A_-$. 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* $\mathcal{B}_0(\mathcal{H}_1, \mathcal{H}_2) := \mathrm{lin}\{|\psi\rangle\langle\varphi| \;\; \big| \psi \in \mathcal{H}_2, \varphi \in \mathcal{H}_1\}$. 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 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* $\mathrm{tr}\,[A] \in [0, \infty]$ is defined as

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

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 1.6 (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 multiples 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 $\mathrm{tr}\,[(A_n - A)B] \to 0\ \forall B \in \mathcal{B}_1(\mathcal{H})$,
- *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 $\mathcal{B}(\mathcal{H})$. The *weak-* topology*, for instance, can be defined as the smallest topology in which all functionals of the form $\mathcal{B}(\mathcal{H}) \ni A \to \mathrm{tr}\,[AB]$ are continuous for any $B \in \mathcal{B}_1(\mathcal{H})$.

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

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

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

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).

Exercise 1.5. 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 1.6. 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 1.7. 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 1.8. 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 1.9. Let $Q \in \mathcal{B}(\mathcal{H})$ be positive and such that $\ker(Q) = \{0\}$. Prove that $(A, B) \mapsto \mathrm{tr}\,[QA^*B]$ defines an inner product on $\mathcal{B}_2(\mathcal{H})$.

Exercise 1.10. Construct a sequence of finite rank operators $A_n \in \mathcal{B}_0(\mathcal{H})$ that converges weakly to zero but not strongly.

1.3 Probabilistic structure of Quantum Theory

Quantum theory can be regarded as a general theoretical framework for physical theories. It consist 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 natures 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 which determines all probability distributions of any possible measurement. It has to be a procedure which, 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 1.4 (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 1.5 (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 1.7 (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

$$\rho = \frac{1}{2}\Big(\mathbb{1} + \sum_{i=1}^{3} r_i \sigma_i\Big). \tag{1.16}$$

The purity is then expressible as $\mathrm{tr}\left[\rho^2\right] = \frac{1}{2}\big(1 + \|r\|^2\big)$. 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.(1.16) is not possible in higher dimensions.

In infinite dimensions, as seen in Exercise 1.10, weak converges can be a rather weak, 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 1.6 (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 $\mathrm{tr}\left[\rho_n\right] \to 1$. 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 $1 - \mathrm{tr}\left[\rho P\right] =: \epsilon$ 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

$$\|\rho - \rho_n\|_1 \leq \|P(\rho - \rho_n)P\|_1 + 2\left\|P(\rho - \rho_n)P^\perp\right\|_1 + \left\|P^\perp(\rho - \rho_n)P^\perp\right\|_1. \tag{1.17}$$

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.(1.17) we first use that $P\rho P^\perp = 0$ and then bound the remaining part via

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

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.(1.17) 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.

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 1.7 (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*

$$\sum_k M(X_k) = \mathbb{1} \tag{1.18}$$

for any 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.(1.18), 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.(1.18) implies $0 \leq M(Y) \leq \mathbb{1}$.[8] Moreover:

Lemma 1.8. *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.(1.18) 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 $0 \leq M(Y) \leq \mathbb{1}$ implies that $M(Y)M(J) = 0$ whenever $Y \cap J = \emptyset$ (cf. Exercise 1.15).

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 1.9 (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*

$$p(Y|\rho, M) = \mathrm{tr}\left[\rho M(Y)\right]. \tag{1.19}$$

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 1.10. *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 1.6a) and the upper bound from Eq.(1.18) applied to the trivial partition of X together with $\mathrm{tr}\left[\rho\right] = 1$. When applying Eq.(1.18) 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.(1.18) since

$$\sum_k p(X_k) = \sum_k \mathrm{tr}\left[\rho M(X_k)\right] = \mathrm{tr}\left[\rho \sum_k M(X_k)\right] = \mathrm{tr}\left[\rho \mathbb{1}\right] = 1. \tag{1.20}$$

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 practise, 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 text-book quantum theory often assume 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: $p_x := \operatorname{tr}[\rho_x M_x]$. Now consider the average probability of success, averaged uniformly over all $x \in X$:

Proposition 1.11. *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 $\|\rho_x\|_\infty \leq 1$:

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

□

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.1.11 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 1.12 (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 an a common measurable space $(X, \mathbb{B})$ and $Y \in \mathbb{B}$. Then*

$$\left|p(Y|\rho, M) - p(Y|\rho', M)\right| \leq \frac{1}{2}\|\rho - \rho'\|_1, \tag{1.21}$$

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

$$\sup_{\rho} \left| p(Y|\rho, M) - p(Y|\rho, M') \right| = \left\| M(Y) - M'(Y) \right\|_\infty . \tag{1.22}$$

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 projections 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 1.6a and Hölder's inequality):

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

where we have used $\|M(Y)\|_\infty \leq 1$, which is a consequence of $0 \leq M(Y) \leq \mathbb{1}$ (cf. Exercise 1.8). 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.(1.22), first note that Hölder's inequality together with $\|\rho\|_1 = 1$ leads to the upper bound

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

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 1.8d). □

The fact that Eq.(1.21) is tight provides an operational interpretation for the trace-norm distance of 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

$$\langle m \rangle = \mathrm{tr}\left[\rho \hat{M} \right], \quad \hat{M} := \int_X m(x)\, dM(x), \tag{1.23}$$

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. Exercise1.15) we can express the variance as

$$\mathrm{var}(m) = \mathrm{tr}\left[\rho \hat{M}^2\right] - \mathrm{tr}\left[\rho \hat{M}\right]^2 =: \mathrm{var}(\hat{M}). \tag{1.24}$$

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

Exercise 1.12. 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 1.13. Prove the Bloch ball representation in Eq.(1.16). (Hint: use the determinant). For a given density operator on $\mathbb{C}^2$, how can the vector r be obtained?

Exercise 1.14. 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.

Exercise 1.15. 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 1.16. 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 1.17. 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 $U\rho U^* = \rho'$.

---

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

1.4 Convexity

Convex sets and extreme points

Definition 1.13. *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 $\mathcal{E}(C)$.*

Theorem 1.14 (Caratheodory)**.** *Let V be a normed space, $C \subseteq V$ a compact convex set 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 analogue 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 1.15 (Krein-Milman)**.** *Let V be a locally convex topological vector space and $C \subseteq V$ compact and convex. Then C is 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:

Theorem 1.16 (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 $\frac{1}{n}(U_1 + \ldots + U_n) = A$.*

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

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 1.17 (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*

$$\left\| \phi - \frac{1}{n} \sum_{i=1}^{n} \psi_i \right\|^2 \leq \frac{b^2}{n}, \tag{1.25}$$

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 1.18). 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 1.8 (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

$$\rho = \sum_k \sqrt{\rho}|e_k\rangle\langle e_k|\sqrt{\rho} = \sum_k p_k|\varphi_k\rangle\langle\varphi_k|, \tag{1.26}$$

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. 1.22 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. 1.5 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*.

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

Definition 1.18 (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*

$$\sum_{i=1}^{k} \lambda_i^\downarrow \leq \sum_{i=1}^{k} \mu_i^\downarrow \quad \forall k. \tag{1.27}$$

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 1.19 (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:*

$$\sum_{j=1}^{d} M_{ij} = \sum_{j=1}^{d} M_{ji} = 1. \tag{1.28}$$

Example 1.9 (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 1.10 (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 1.20. *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:*

(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 $f(0) = 0$:

$$\sum_k f(\lambda_k) \leq \sum_k f(\mu_k). \tag{1.29}$$

When applied to density operators, this gives:

Corollary 1.21. *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 $f(0) = 0$: $\mathrm{tr}\,[f(\rho_1)] \leq \mathrm{tr}\,[f(\rho_2)]$.*

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 1.22. *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*

$$\lambda \succ p. \tag{1.30}$$

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.1.20 $\lambda \succ p$. □

In the context of Example 1.8, this result implies that among all the decompositions into pure states to which Eq.(1.26) 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. 12).

Theorem 1.23. *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 $\mathrm{spec}(A) \subseteq [a,b]$. Then, for every unit vector $\psi \in \mathcal{H}$:*

$$f\big(\langle\psi, A\psi\rangle\big) \leq \langle\psi, f(A)\psi\rangle. \tag{1.31}$$

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 1.24 (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 $F : \mathcal{C}(\mathcal{H}) \to \mathbb{R}$, $F(A) := \mathrm{tr}\,[f(A)] \in [0,\infty]$. 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 $\mathcal{H}$:

$$F(A) \geq \sum_k f\big(\langle e_k, Ae_k\rangle\big). \tag{1.32}$$

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.(1.31), 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.(1.31) 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 1.25. *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*

$$\begin{aligned}
\sum_{i=1}^n \alpha_i f_i(a) g_i(b) &\geq 0 \quad \forall a, b \in I, \text{ then} \\
\sum_{i=1}^n \alpha_i \operatorname{tr}\left[f_i(A) g_i(B)\right] &\geq 0
\end{aligned} \tag{1.33}$$

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

$$\operatorname{tr}\left[f(A) - f(B)\right] \geq \operatorname{tr}\left[(A-B)f'(B)\right]. \tag{1.34}$$

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

$$\operatorname{tr}\left[f(A) - f(B)\right] - \operatorname{tr}\left[(A-B)f'(B)\right] \geq \frac{c}{2}\|A-B\|_2^2. \tag{1.35}$$

Proof. Both inequalities exploit Lemma 1.25. Eq.(1.34) then follows from the fact that any convex function satisfies $f(a) - f(b) \geq (a-b)f'(b)$ and Eq.(1.26) uses the mean-value version of Taylor's theorem, which states that there is a $z \in [a, b]$ such that

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

□

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

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.

On the relevant interval $[0, 1]$ the function h is non-negative, continuous and concave. By Cor.1.24 (i) this implies that the von Neumann entropy S is a non-negative, concave functional on the set of density operators. From Cor. 1.21 we get

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

For finite dimensional Hilbert spaces the von Neumann entropy is continuous, which is implied by 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 1.11 and Exercise 1.21).

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 1.11 (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.

Exercise 1.18. Show that pure states are extreme points of the convex set of density operators.

Exercise 1.19. 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 1.20. 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 1.21. Construct a sequence of density operators of finite entropy that converges in trace-norm to a pure state but has entropy diverging to ∞.

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 1.28 (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 1.29 (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*

$$\langle \varphi_1 \otimes \varphi_2, \psi_1 \otimes \psi_2 \rangle := \langle \varphi_1, \psi_1 \rangle \langle \varphi_2, \psi_2 \rangle, \tag{1.36}$$

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:

$$\begin{aligned} \mathcal{H}_1 \otimes \mathcal{H}_2 &\simeq \mathcal{H}_2 \otimes \mathcal{H}_1, \\ (\mathcal{H}_1 \otimes \mathcal{H}_2) \otimes \mathcal{H}_3 &\simeq \mathcal{H}_1 \otimes (\mathcal{H}_2 \otimes \mathcal{H}_3), \\ \mathcal{H}_1 \otimes (\mathcal{H}_2 \oplus \mathcal{H}_3) &\simeq (\mathcal{H}_1 \otimes \mathcal{H}_2) \oplus (\mathcal{H}_1 \otimes \mathcal{H}_3). \end{aligned} \tag{1.37}$$

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:

$$\left(\sum_{i=1}^{k} \psi_i \right) \otimes \left(\sum_{j=1}^{l} \varphi_j \right) = \sum_{i=1}^{k} \sum_{j=1}^{l} \psi_i \otimes \varphi_j, \tag{1.38}$$

$$(c\psi) \otimes \varphi = c(\psi \otimes \varphi) = \psi \otimes (c\varphi), \quad \text{for } c \in \mathbb{C}. \tag{1.39}$$

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 1.30 (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 $\Psi, \Phi \in \mathcal{H}_1 \otimes \mathcal{H}_2$:*

$$\langle \Phi, \Psi \rangle = \operatorname{tr}\left[\mathcal{I}(\Phi)^* \mathcal{I}(\Psi) \right]. \tag{1.40}$$

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

$$\mathcal{I} : |\psi\rangle \otimes |\varphi\rangle \mapsto |\varphi\rangle\langle\overline{\psi}| \tag{1.41}$$

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 1.31 (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*

$$\Psi = \sum_{i=1}^{r} s_i \; e_i \otimes f_i. \tag{1.42}$$

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.1.30 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.(1.41) then proves the decomposition in Eq.(1.42). 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 1.12 (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.1.30 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

$$(A_1\otimes A_2)(B_1\otimes B_2) = (A_1B_1)\otimes(A_2B_2). \tag{1.43}$$

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.1.30 and see how it treats tensor products of operators:

Corollary 1.32. *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.(1.41), 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

$$(A\otimes\mathbb{1})\Psi = (\mathbb{1}\otimes B)\Psi. \tag{1.44}$$

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.(1.41), 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.(1.44) 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.(1.44), 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 1.13 (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 1.33 (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 1.34. *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.1.30, 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

$$\Psi_n = n\Big(e + \frac{1}{n} f\Big) \otimes \Big(e + \frac{1}{n} f\Big) \otimes \Big(e + \frac{1}{n} f\Big) - n\, e \otimes e \otimes e. \tag{1.45}$$

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

Example 1.14 (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

$$(AB)_{\alpha\beta} = \sum_{i=1}^r C_{i,\alpha\beta}\ a(A)_i\ b(B)_i. \tag{1.46}$$

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.(1.46) 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 analogue of this marginalizing map. The analogy will then be made clearer in the subsequent paragraph.

Theorem 1.35 (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*

$$\mathrm{tr}\left[B(A \otimes \mathbb{1})\right] = \mathrm{tr}\left[\mathrm{tr}_2[B]A\right], \quad \forall A \in \mathcal{B}(\mathcal{H}_1), \tag{1.47}$$

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

$$\begin{aligned} B \geq 0 \quad &\Rightarrow \quad \mathrm{tr}_2[B] \geq 0, \\ B = B_1 \otimes B_2 \quad &\Rightarrow \quad \mathrm{tr}_2[B] = B_1 \mathrm{tr}\left[B_2\right], \\ \mathrm{tr}\left[\mathrm{tr}_2[B]\right] \quad &= \quad \mathrm{tr}\left[B\right]. \end{aligned} \tag{1.48}$$

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

$$\mathrm{tr}_2[B] := \sum_k \big(\mathbb{1} \otimes \langle e_k|\big) B \big(\mathbb{1} \otimes |e_k\rangle\big). \tag{1.49}$$

According to the subsequent Lemma 1.36, the r.h.s. of this equation converges in trace-norm to a trace-class operator. Hence, tr_2 is well-defined and Eq.(1.47) can be verified by insertion. Uniqueness of the map is implied by the fact that specifying $\mathrm{tr}\left[XA\right]$ for all $A \in \mathcal{B}(\mathcal{H}_1)$ determines X. In particular, the construction in Eq.(1.49) is basis-independent.

The properties summarized in Eq.(1.48) follow immediately from Eq.(1.47). For instance, positivity of $\langle\psi, \mathrm{tr}_2[B]\psi\rangle = \mathrm{tr}\left[B(|\psi\rangle\langle\psi| \otimes \mathbb{1})\right]$ is implied by positivity of B together with $|\psi\rangle\langle\psi| \otimes \mathbb{1} \geq 0$ (cf. Exercise 1.6). □

Finally, we prove the missing Lemma that shows trace-norm convergence of the ansatz in Eq.(1.49). For later use, the formulation is slightly more general.

Lemma 1.36. *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*

$$\left\| B' - \sum_{k=1}^n A_k B A_k^* \right\|_1 \to 0\,, \tag{1.50}$$

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.1.35. 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. $\mathrm{tr}_\Lambda = \prod_{i\in\Lambda} \mathrm{tr}_i$.

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 1.37. *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*

$$\rho_1 = \sum_{i=1}^{r} \lambda_i |e_i\rangle\langle e_i| \quad \textit{and} \quad \rho_2 = \sum_{i=1}^{r} \lambda_i |f_i\rangle\langle f_i|. \tag{1.51}$$

Proof. The statement follows from inserting the Schmidt decomposition into the explicit form of the partial trace in Eq.(1.49). The calculation simplifies if we use the basis of the Schmidt decomposition in the respective partial trace. □

Cor.1.37 leads to some of 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 1.38 (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.(1.51), 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. 1.37 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.

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 1.8) 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 states 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 1.39 (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 $I(A:B) := S(\rho\|\rho_A \otimes \rho_B)$.*

A crucial properties of both quantities is positivity together with the fact that they are zero only in the obvious case:

Corollary 1.40 (Pinsker inequality)**.** *The relative entropy and the mutual information as defined in Def.1.39 satisfy:*

$$S(\rho\|\sigma) \geq \frac{1}{2}\|\rho - \sigma\|_1^2, \tag{1.52}$$

$$I(A:B) \geq \frac{1}{2}\|\rho_{AB} - \rho_A \otimes \rho_B\|_1^2. \tag{1.53}$$

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. (1.52,1.53) for $\|\cdot\|_2$ instead of for $\|\cdot\|_1$. Clearly, the trace-norm bound is the stronger result and we refer to ... for its proof.

By definition of the mutual information, Eq.(1.53) is a consequence of Eq.(1.52). In order to arrive at Eq.(1.52), we use the fact that $f(x) := x \log x$ is strongly convex on $[0,1]$ with $f''(x) = 1/x \geq 1$. So we can apply Eq.(1.26) from which the result then follows instantly. □

Exercise 1.22. 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 1.23 (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 1.24 (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 1.25 (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).

1.6 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 $C\operatorname{tr}[\rho]$. 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 1.41. *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 $\operatorname{tr}[T(A)] = \operatorname{tr}[A]$,*

- 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 1.15 (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 1.23).

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.[12] 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

[12] 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.

Heisenberg picture. We will comment on the more subtle converse direction below.

Theorem 1.42 (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 $\forall A \in \mathcal{B}(\mathcal{H}_2), \rho \in \mathcal{B}_1(\mathcal{H}_1)$:*

$$\operatorname{tr}\left[T(\rho)A\right] = \operatorname{tr}\left[\rho\, T^*(A)\right]. \tag{1.54}$$

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.

Proof. Consider the map $f : \mathcal{B}_1(\mathcal{H}_1) \to \mathbb{C}$ defined by $f(B) := \operatorname{tr}\left[T(B)A\right]$ for fixed $A \in \mathcal{B}(\mathcal{H}_2)$. Due to 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 $f(B) = \operatorname{tr}\left[BT^*(A)\right]$, which verifies Eq.(1.54). As the l.h.s. of Eq.(1.54) 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 $\operatorname{tr}\left[\rho T^*(A)\right]$ 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

$$\operatorname{tr}\left[T(|\psi\rangle\langle\psi|)A\right] = \langle\psi, T^*(A)\psi\rangle. \tag{1.55}$$

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.(1.54) we derive the equation

$$\operatorname{tr}\left[T(B) - B\right] = \operatorname{tr}\left[B\big(T^*(\mathbb{1}) - \mathbb{1}\big)\right]. \tag{1.56}$$

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

$$\|B\|_1 = \sup_{\|A\|_\infty=1} \big|\mathrm{tr}\,[AB]\,\big|, \quad \text{and} \quad \|A\|_\infty = \sup_{\|B\|_1=1} \big|\mathrm{tr}\,[AB]\,\big|. \tag{1.57}$$

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

Corollary 1.43. *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 $B \in \mathcal{B}_1(\mathcal{H}_1), A \in \mathcal{B}(\mathcal{H}_2)$:*

$$\|T(B)\|_1 \leq \|B\|_1 \quad \textit{and} \quad \|T^*(A)\|_\infty \leq \|A\|_\infty\,. \tag{1.58}$$

Proof. Using the defining relation between T and T^* and Eq.(1.57) we obtain

$$\|T^*\| = \sup_{\|A\|_\infty=1} \sup_{\|B\|_1=1} \big|\underbrace{\mathrm{tr}\,[BT^*(A)]}_{=\mathrm{tr}[T(B)A]}\big| = \|T\|\,. \tag{1.59}$$

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.1.16) and the former by the Schmidt-decomposition (Eq.(1.8)). Thus

$$\|T^*\| = \sup_U \sup_{\psi,\varphi} \big|\langle\varphi, T^*(U)\psi\rangle\big|, \tag{1.60}$$

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.(1.60) leads to

$$\begin{aligned} \|T^*\| &\leq \sup_E \sup_{\psi,\varphi} \sum_k \big|\langle\varphi, T^*\big(|e_k\rangle\langle e_k|\big)\psi\rangle\big|, &(1.61)\\ &= \sup_E \sup_{\psi} \sum_k \langle\psi, T^*\big(|e_k\rangle\langle e_k|\big)\psi\rangle = \|T^*(\mathbb{1})\|_\infty\,. &(1.62) \end{aligned}$$

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.(1.58) 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.(1.60) 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.1.42 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.1.42 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 1.44 (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.

$$T(\rho) = \mathrm{tr}_{\mathcal{K}}\big[U(\rho \otimes \sigma)U^*\big], \tag{1.63}$$

(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.

$$T(\rho) = \mathrm{tr}_{\mathcal{K}}\big[W(\rho \otimes |\psi\rangle\langle\psi|)W^*\big], \tag{1.64}$$

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

$$T(\rho) = \mathrm{tr}_{\mathcal{K}}\big[V\rho V^*\big], \tag{1.65}$$

(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

$$T(\rho) = \sum_{k=1}^{r} A_k \rho A_k^*. \tag{1.66}$$

Remark: The A_k's are called *Kraus-operators* and Eq.(1.66) the *Kraus representation* of T. As we have seen in Lemma 1.36, 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.(1.66).

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.1.38, 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.(1.49), we see that Eq.(1.66), after insertion of the A_k's, becomes Eq.(1.64). 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.(1.63) 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 1.8) and, in a similar vein, one can show the following:

Proposition 1.45 (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 1.46 (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.1.44.

Example 1.16 (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:

$$\begin{aligned} \rho \quad &\mapsto \quad \begin{pmatrix} \rho_{00} & (1-\lambda)\rho_{01} \\ (1-\lambda)\rho_{10} & \rho_{11} \end{pmatrix} = \sum_{k=1}^{3} A_k \rho A_k^* \qquad (1.67) \\ \text{with} \quad & A_1 := \sqrt{1-\lambda}\, \mathbb{1}, \quad A_2 := \sqrt{\lambda}\, |0\rangle\langle 0|, \quad A_2 := \sqrt{\lambda}\, |1\rangle\langle 1|. \end{aligned}$$

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 1.17 (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 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 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 1.47 (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*[13]*. 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:

Lemma 1.48 (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.(1.41) (w.r.t. the same basis that defines Ω). Then*

$$|\psi\rangle = (\mathbb{1} \otimes A)|\Omega\rangle. \tag{1.68}$$

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.(1.68) 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 1.49 (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

$$\mathrm{tr}\left[T(A)B\right] = \mathrm{tr}\left[C(A^T \otimes B)\right], \quad \forall A \in \mathcal{B}(\mathcal{H}_1), B \in \mathcal{B}(\mathcal{H}_2), \tag{1.69}$$

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

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

[13]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.

Proof. (i) Note that via Eq.(1.69) T and C mutually determine each other so that Eq.(1.69) 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

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

Here we have used the property of the flip operator from Exercise 1.23 (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 is 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 1.48, 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 $\operatorname{tr}\big[T\big(|i\rangle\langle j|\big)\big] = \langle j|T^*(\mathbb{1})|i\rangle$ the claim follows from $\operatorname{tr}_2[C] = \sum_{ij}|i\rangle\langle j|\operatorname{tr}\big[T\big(|i\rangle\langle j|\big)\big] = T^*(\mathbb{1})^T$.

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

Part (iii) of Thm.1.49 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.1.42), 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_1d_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 1.50 (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 $r\leq d_1d_2$ such that*

$$T(\cdot) = \sum_{k=1}^{r} A_k\cdot B_k^*. \tag{1.70}$$

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 1.48. 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 1.48 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.(1.70) then follows from the fact that there is a unique T corresponding to C (Thm.1.49 (i)). If T is completely positive, then C is positive (Thm.1.49(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 1.51 (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 instruments 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

$$I_Y(\rho) := \sum_{i\in Y} M(Y)^{1/2} \rho M(Y)^{1/2}. \tag{1.71}$$

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.(1.71) is said to be an *ideal measurement* or an *ideal instrument*. Traditionally, these are the ones that are used in quantum mechanics text-books 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 $Y \in \mathbb{B}$.

Commuting dilations 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 1.52 (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 $Y \in \mathbb{B}$:*

$$V^* M'(Y) V = M(Y). \tag{1.72}$$

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 $x \in X$.

We will provide an elementary proof for the case of finitely many measurement outcomes, and sketch later how 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 1.53 (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 $Y \in \mathbb{B}$:*

$$\operatorname{tr}\left[\rho M(Y)\right] = \operatorname{tr}\left[\big(\rho \otimes |\psi\rangle\langle\psi|\big) M'(Y)\right] \qquad \forall \rho \in \mathcal{B}_1(\mathcal{H}). \tag{1.73}$$

Conversely, if ψ and M' are as specified, then Eq.(1.73) 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.1.52) 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 $\mathcal{K}$. Uder these assumptions, by copying the argument of the proof of Thm.1.44, 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.(1.73) 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'. □

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 1.54 (Commuting Hermitian dilations)**.** *Let $H_1, \dots, H_n \in \mathcal{B}(\mathcal{H})$ 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 $K_1, \dots, K_n \in \mathcal{B}(\mathcal{K})$ 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 part. With $c := \left\| \sum_{i=1}^{2n} B_i \right\|_\infty$ define $A_i := B_i/c$ and $A_0 := \mathbb{1} - \sum_{i=1}^{2n} A_i$. Then $A_0, \dots, A_{2n}$ 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 $i \in \{1, \dots, n\}$. 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 1.55 (Commuting dilations)**.** *For any finite sequence of operators $A_0, \dots, A_{n-1} \in \mathcal{B}(\mathcal{H})$ there exist pairwise commuting operators $K_0, \dots, K_{n-1} \in \mathbb{B}(\mathbb{C}^n \otimes \mathcal{H})$ and a unit vector $|0\rangle \in \mathbb{C}^n$ s.t.*

$$A_k = \big(\langle 0| \otimes \mathbb{1}\big) K_k \big(|0\rangle \otimes \mathbb{1}\big) \qquad \forall k. \tag{1.74}$$

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.(1.74). To see that this leads to a commuting set of operators note that

$$K_{k_1} K_{k_2} = \sum_{i_1, j_2} |i_1\rangle\langle j_2| \otimes \sum_{j_1} A_{i_1 - j_1 + k_1} A_{j_1 - j_2 + k_2}. \tag{1.75}$$

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 1.26 (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 1.27 (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 1.28 (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 1.29 (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 1.30 (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 $\rho \in \mathcal{B}(\mathbb{C}^d)$.

1.7 Unbounded operators and spectral measures

In this section 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

$$\langle \varphi, A\psi\rangle = \langle A^*\varphi, \psi\rangle \quad \forall \psi \in \mathcal{D}(A), \varphi \in \mathcal{D}(A^*). \tag{1.76}$$

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

$$A = \int_{\mathbb{R}} \lambda \, dP(\lambda). \tag{1.77}$$

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.(1.77) leads directly to a functional calculus. For any measurable function $f : \mathbb{R} \to \mathbb{C}$ we can define

$$\begin{aligned} f(A) \quad &= \quad \int_{\mathbb{R}} f(\lambda) \, dP(\lambda) \\ &\text{on} \quad \mathcal{D}\big(f(A)\big) := \left\{ \varphi \in \mathcal{H} \middle| \int_{\mathbb{R}} |f(\lambda)|^2 \, d\langle \varphi, P(\lambda)\varphi\rangle < \infty \right\}. \end{aligned} \tag{1.78}$$

If f is bounded, then Eq.(1.78) gives rise to a bounded operator.

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 implementation.
- *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 $\langle X \rangle := \operatorname{tr}[\rho X]$ for any $X \in \mathcal{B}(\mathcal{H})$. Then the $n \times n$* covariance matrix[1] *$V_{kl} := \frac{1}{2}\big\langle \{H_k - \langle H_k\rangle, H_l - \langle H_l\rangle\}_+ \big\rangle$ and the* commutator matrix *$\sigma_{kl} := \frac{i}{2}\big\langle [H_k, H_l] \big\rangle$ 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, which allows to map $V \mapsto SVS^T$ to the same block-diagonal structure while keeping 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, $\langle X \rangle := \operatorname{tr}[\rho X]$ for any $X \in \mathcal{B}(\mathcal{H})$, and $\operatorname{var}(A) := \langle A^2\rangle - \langle A\rangle^2$. Then*

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

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. The point that requires additional care in this case are the domains of all involved operators. For instance, in Robertson's uncertainty relation, if $\rho = |\psi\rangle\langle\psi|$ and

[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 $H_k - \langle H_k\rangle \mathbb{1}$.

if $\mathcal{D}(H_kH_l)$ is the domain of H_kH_l, then we need $\psi \in \bigcap_{kl} \mathcal{D}(H_kH_l)$. 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

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

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 (Cor1.38). 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 (1.37) 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 $M : \mathcal{B} \to \mathcal{B}(\mathcal{H})$ 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 1.7). 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. □

2.2 Information-disturbance

No information without disturbance

Theorem 2.5 (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

$$K_i^* K_j = \sigma_{ij} \mathbb{1} \quad \forall i, j. \tag{2.4}$$

(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:

$$I_Y^*(\mathbb{1}) \propto \mathbb{1} \quad \forall Y \in \mathbb{B}. \tag{2.5}$$

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.1.49) 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.(2.4) 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.(2.5) 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. 1.45) 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.(2.4). 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.1.45) again. It allows to choose 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.1.45)). 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 they 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 2.1 (Quantum error correcting codes)**.** The condition in Eq.(2.4) 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

$$V^* F V \propto \mathbb{1} \tag{2.6}$$

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.(2.4) so that Thm.2.5 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 2.6 (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*

$$\Delta(H)\Delta(A) \geq \frac{1}{2} \left| \frac{d\langle A \rangle}{dt} \right|. \tag{2.7}$$

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.(2.7) when evaluated at $t = \tau$.

Remark: Note that $\Delta(H)$ is time-independent. With the necessary care concerning the domain, Eq.(2.7) also holds for an unbounded Hermitian Hamiltonian H.

Proof. With $\frac{d}{dt}\langle A \rangle = i\langle [H, A] \rangle$ Eq.(2.7) 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 $\langle \{A, H - \langle H \rangle\}_+ \rangle|_{t=\tau} = 0$. Moreover, $A\psi(\tau) = i(H - \langle H \rangle \mathbb{1})\psi(\tau)$ so that equality holds in the Robertson-Schrödinger uncertainty relation and therefore also in Eq.(2.7). □

Corollary 2.7 (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

$$\Delta(H)\, t \quad \geq \quad \arccos\left(\sqrt{p(t)}\right), \quad \textit{so that} \tag{2.8}$$

$$\Delta(H) t_{1/2} \geq \frac{\pi}{4} \quad , \quad \Delta(H) t_0 \geq \frac{\pi}{2}, \tag{2.9}$$

where $t_{1/2}$ and t_0 are the shortest times (according to Eq.(2.8)) for $p(t)$ to drop to $1/2$ and 0, respectively.

Proof. We apply the Mandelstam-Tamm uncertainty relation in Eq.(2.7) 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.(2.9) then follow from Eq.(2.8) 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 $p(t)$.

Evolution to orthogonal states In Cor.2.7 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 2.8 (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 $\chi(t) := \int_{R_+} e^{-it\lambda}\, d\mu(\lambda)$ and its p'th moment for any $p > 0$ by $m_p := \int_{R_+} \lambda^p\, d\mu(\lambda)$. Then*

$$t_0 := \inf\left\{t > 0 | \chi(t) = 0\right\} \geq \frac{\pi}{(2m_p)^{1/p}}. \tag{2.10}$$

With this we can prove the following:

Theorem 2.9 (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*

$$t_0\, \langle\psi, H^p\psi\rangle^{1/p} \geq \frac{\pi}{2^{1/p}}. \tag{2.11}$$

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 2.8. □

For $p = 2$ Eq.(2.11) is similar to the consequence that we obtained in Cor.2.7 from the Mandelstam-Tamm inequality. In fact, at first glance, Eq.(2.11) looks even stronger since there is a missing factor $1/\sqrt{2}$. Note, however, that Eq.(2.11) requires an additional assumption, namely positivity of the Hamiltonian.

For $p = 1$ Eq.(2.11) 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 2.10 (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 2.11 (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*

$$\nu \geq \max\{0, 2p-1\}, \tag{2.12}$$

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

$$\left|\sum_k p_k e^{-i\lambda_k t}\right| \;\geq\; p - \left|\sum_{k>1} p_k e^{i(\lambda_1 - \lambda_k)t}\right| \;\geq\; p - \sum_{k>1} p_k = 2p - 1, \tag{2.13}$$

thus proving the inequality in Eq.(2.12).

For proving that equality holds if the λ_k's are independent over $\mathbb{Q}$ we want to use Lemma 2.10. 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 2.10 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 2.12 (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.2.11, 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.2.11 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.