Quantum Channels & Operations

Guided Tour

Michael M. Wolf

July 5, 2012

These notes are based on a course given at the Niels-Bohr Institute (Copenhagen) in 2008/2009.

Most passages are still cryptic, erroneous, incomplete or actually missing. The notes are occasionally updated. Since they contain also some unpublished observations, please let me know if you want to use them ...

Contents

Chapter 1

(De)constructing quantum mechanics

Quantum mechanics can be regarded as a general theoretical framework for physical theories. It consist out of a mathematical core which becomes a physical theory when adding a set of correspondence rules telling us which mathematical objects we have to use in different physical situations. In contrast to classical physical theories, these correspondence rules are not very intuitive as linear operators on Hilbert spaces are quite far from everyday life. It is truly remarkable that Heisenberg, Schrödinger, Dirac, Bohr, von Neumann together with all the other famous minds of this golden age come up with such a theory.

In this chapter we will briefly review the mathematical formalism of quantum mechanics—abstracting from concrete physical realizations. All quantities will be dimensionless and $\hbar$ is set to one (although one should keep in mind that it is actually $10^{-34} Js$, in other words: really, really small). Moreover, we will with few exceptions restrict ourselves to finite dimensional systems. This suffices to clarify the basic concepts and allows us to achieve this with little more requirements than linear matrix algebra.

1.1 States and Observables

Preparation and Measurement It is often useful to divide 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 ambiguous (see the discussion of time evolution in Sec.1.3) but, fortunately, in the case of quantum mechanics predictions do not depend on this choice.

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 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. This leads to two related conceptual puzzles: (i) we have to specify where this cut takes place, i.e., where facts appear. This choice is ambiguous but fortunately irrelevant (as long as we make this cut at some point). (ii) the transition is intrinsically irreversible and thus in apparent conflict with the usual reversible way of setting up the theory. Variations of these two themes run under the name *measurement problem*.

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 mechanical '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. Although it is quite common, one should neither assign states to single events nor interpret them as elements of reality.

States as density matrices The division of physical experiments into preparation of a state and measurement of an observable quantity (an *observable*) is reflected in the mathematical structure of quantum mechanics. The mathematical representation of the set of observables is given by Hermitian elements A taken from an algebra $\mathcal{A}$ (called *observable algebra*). Just to recall: an *algebra* is a set which is closed under multiplication and addition as well as under multiplication with scalars. Moreover, we will assume that each element has an *adjoint*, i.e., that there is a Hermitian conjugation operation. The algebras are usually represented in terms of bounded[1] linear operators $\mathcal{B}(\mathcal{H})$ acting on a Hilbert space $\mathcal{H}$. When necessary we will write $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ for the bounded linear operators from a Hilbert space $\mathcal{H}_1$ to $\mathcal{H}_2$, i.e., $\mathcal{B}(\mathcal{H}) = \mathcal{B}(\mathcal{H}, \mathcal{H})$. The

[1] At this point you might disagree when having position or momentum operators or certain unbounded Hamiltonians in mind. However, in order to compute observable statistics, only the spectral projectors of these operators are needed and these are bounded again.

algebras used will always admit such a representation

A state in turn corresponds to a linear functional mapping observables onto real numbers—their expectation values $\langle A\rangle$. That is, a state is an element of the *dual* $\mathcal{A}^*$ of $\mathcal{A}$, i.e., the space of continuous, linear functionals over $\mathcal{A}$. In order to get a nice interpretation in terms of probabilities, states are required to be

- *normalized*: $\langle \mathbb{1}\rangle = 1$ (i.e., $\mathcal{A}$ should contain a unit element $\mathbb{1}$) and
- *positive*: $\forall A \in \mathcal{A} : \langle A^\dagger A\rangle \geq 0$, where † is the adjoint operation[2].

Commonly, the words 'state' and 'observable' are used to refer to both the physical concept and the mathematical operator.

We will in the following mainly consider systems of finite dimension d where $A \in \mathcal{M}_d \cong \mathcal{B}(\mathbb{C}^d)$ is a $d \times d$ matrix with a concrete representation on $\mathcal{H} = \mathbb{C}^d$. In general, $\mathcal{A}$ is a C^*-algebra (see Sec.1.6). As inequivalent representations are usually no issue in finite dimensions, we will largely disregard the distinction between $\mathcal{A}$ and its representations.

The set of complex valued matrices $\mathcal{M}_d$ can be upgraded to a Hilbert space equipped with the Hilbert-Schmidt scalar product $\langle B, A\rangle_{HS} := \mathrm{tr}\left[B^\dagger A\right]$.[3] Hence, any linear functional can be identified with a matrix itself, so that every state can be described by a *density matrix* $\rho \in \mathcal{M}_d$ for which normalization and positivity read $\mathrm{tr}\left[\rho\right] = 1$ and $\rho \geq 0$ respectively.[4] The expectation value of an observable is then given by

$$\langle A\rangle = \mathrm{tr}\left[\rho A\right]. \tag{1.1}$$

An important property of the set of density matrices is that it is *convex*, i.e., if ρ_i are density operators and λ_i probabilities, then the *mixture*

$$\rho = \sum_i \lambda_i \rho_i \tag{1.2}$$

is again an admissible density operator. It describes a state which can be prepared by generating the states ρ_i with probability λ_i. If a state ρ has no non-trivial convex decomposition of the form in Eq.(1.2) it is called *pure*, and *mixed* otherwise. Pure state density operators are one-dimensional projectors $\rho = |\psi\rangle\langle\psi|$, so that pure states can equivalently be represented by (normalized) vectors $|\psi\rangle \in \mathcal{H}$ traditionally called *state vectors* or *wave functions*. We will see that various measures can be used in order to quantify how pure/mixed a given state is. The simplest one, often called *purity*, is $\mathrm{tr}\left[\rho^2\right]$ and ranges from $\frac{1}{d}$ for the *maximally mixed state* $\rho = \mathbb{1}/d$ to 1 which is attained iff the state is pure.

[2] OK, we will use a compromise in notation: $A^\dagger$ for the adjoint, and $\bar{c}$ for complex conjugation. Though, I don't have the heart of writing $C^\dagger$-algebra.

[3] The resulting Hilbert space is sometimes called *Hilbert Schmidt* Hilbert space. *Hilbert Schmidt class* operators are those for which the *Hilbert Schmidt norm* $\sqrt{\langle A, A\rangle_{HS}}$ is finite.

[4] In infinite dimensions this is no longer true in general. There are states (i.e., linear positive and normalized functionals on an observable algebra) which cannot be represented by density matrices. These are called *singular states* as opposed to *normal states* for which a density matrix representation exists.

Note that every mixed state can be convexly decomposed into pure states. A straight forward way to achieve this is given by the spectral decomposition $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, where the weights λ_i are eigenvalues and the $|\psi_i\rangle$ the corresponding orthogonal eigenvectors. However, every mixed state has infinitely many different decompositions into pure states reflecting infinitely many possibilities of preparing it by mixing pure states. This is a crucial difference to classical probability theory where the convex set of probability distributions forms a *simplex*, meaning that there is a unique decomposition into 'pure' elements. Remember that by definition of a quantum mechanical state there is no way to distinguish between preparations corresponding to different decompositions. A hypothetical device which is defined by its capability of achieving this task is sometimes call *mixed state analyzer* and often appears together with an individual state interpretation (for which we already learned to keep our hand off it). In fact, we will see later that such an innocent looking mixed state analyzer would immediately lead to a breakdown of *Einstein locality* when applied to entangled states. A closely related pitfall is to think of the occurrence of mixed states in quantum mechanics as a result of ignorance or incomplete knowledge. While this can be correct in specific cases it not true in general and we will soon see that it fails in particular if the mixed state is part of a pure entangled state.

Example 1.1 (Bloch sphere) *Two-level quantum systems,* qubits *($d = 2$), are simpler than higher dimensional ones not just because two is smaller than three, but very often they turn out to be special and allow for a qualitative different treatment. One reason for this are the special properties of the three Pauli matrices*

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.3}$$

which, together with the identity $\mathbb{1}$, form a basis for all 2×2 matrices. For any $M \in \mathcal{M}_2(\mathbb{C})$ we can write $M = \frac{1}{2}(x_0 \mathbb{1} + \vec{x} \cdot \vec{\sigma})$, where $x_0 = \mathrm{tr}[M]$ and $\vec{x} \cdot \vec{\sigma} = \sum_{i=1}^{3} x_i \sigma_i$, $\vec{x} \in \mathbb{C}^3$. M is Hermitian iff x_0 and $\vec{x}$ are real and it is positive iff in addition $||\vec{x}||_2 \leq x_0$. Hence, the set of density matrices

$$\rho = \frac{1}{2}\Big(\mathbb{1} + \vec{x} \cdot \vec{\sigma}\Big) \tag{1.4}$$

can be represented by a ball $||\vec{x}||_2 \leq 1$, the Bloch ball. *In this visualization a state is pure iff $||\vec{x}||_2 = 1$, i.e., it is part of the* Bloch sphere. *Closer to the center states become more and more mixed (the purity is given by $\mathrm{tr}\left[\rho^2\right] = \frac{1}{2}(1 + ||\vec{x}||^2)$) and for $\vec{x} = 0$ we have the maximally mixed state. An orthogonal rotation of the Bloch ball corresponds to a unitary on the level of the density matrix. More precisely,*

$$\rho \mapsto U_{\vec{n},\theta} \rho U_{\vec{n},\theta}^\dagger, \quad \textit{with} \quad U_{\vec{n},\theta} = e^{-i\theta \frac{\vec{n}\cdot\vec{\sigma}}{2}} = \mathbb{1} \cos\frac{\theta}{2} - i\vec{n} \cdot \vec{\sigma} \sin\frac{\theta}{2} \tag{1.5}$$

corresponds to a rotation of the Bloch ball by an angle θ about the $\vec{n}$ axis. Conversely, every 2×2 unitary is, up to a phase factor, of the form in Eq.(1.5). This reflects the isomorphism $SO(3) \cong SU(2)/\{-\mathbb{1}, \mathbb{1}\}$, i.e., every pair $\pm U \in SU(2)$ is in one-to-one correspondence with an element $O \in SO(3)$.

A close relative (if not alter ego) of the Bloch sphere is the Poincaré sphere *describing polarization of light. In this context the north pole, 'front pole' and 'east pole' correspond to perfect horizontal,* 45° *linear and circular polarization. The components of the respective vector are then called* Stokes parameters*.*

Measurements An observable in quantum mechanics should be understood as a measurement procedure in a statistical experiment. Again we consider equivalence classes of measurement procedures yielding equal probability distributions when applied to equal preparations. Similar to the individual state representation we run into trouble if we insist on assigning real entities to observables. They rather describe *how* we measure than *what* we measure.

Let us label distinct measurement outcomes corresponding to an observable A by an index α and denote their values and probabilities by a_α and p_α respectively.[5] With this the expectation value reads $\langle A\rangle = \sum_\alpha a_\alpha p_\alpha$. In quantum mechanics the probabilities p_α are calculated by assigning a positive operator $0 \leq P_\alpha \leq \mathbb{1}$ (called *effect* operator) to each outcome so that

$$p_\alpha = \operatorname{tr}\left[\rho P_\alpha\right]. \tag{1.6}$$

As probabilities sum up to one we have to require $\sum_\alpha P_\alpha = \mathbb{1}$. A set of positive operators fulfilling this normalization condition (being a *resolution of the identity*) is called *positive operator-valued measure* (POVM). POVMs are the most general concept in quantum mechanics for describing measurements if we are only interested in probabilities and not in the state after the measurement. The connection between POVMs and the standard text book notion of observables as self-adjoint operators in Eq.(1.1) is given by the spectral decomposition $A = \sum_\alpha a_\alpha P_\alpha$. Here the P_α are now orthogonal projectors onto the eigenspaces of A with corresponding eigenvalues a_α. Clearly, the spectral projectors form a POVM and the formulas for the expectation value become consistent if one interprets the eigenvalues as measurement outcomes. However, POVMs need not consist out of orthogonal projectors, which is why they are sometimes called *generalized measurements* as opposed to *von Neumann measurements* in the case of spectral projectors.

Two observables given by POVMs $\{P_\alpha\}$, $\{Q_\beta\}$ are *jointly measurable* iff there is a 'finer' POVM $\{R_{\alpha\beta}\}$ which recovers them as 'marginals', i.e., $\sum_\alpha R_{\alpha\beta} = Q_\beta$ and $\sum_\beta R_{\alpha\beta} = P_\alpha$. A set of von Neumann measurements is jointly measurable iff the corresponding Hermitian operators commute pairwise (and can thus be all diagonalized simultaneously).

Problem 1 (What's physical?) *Is, in every given physical context, every observable (i.e., every POVM) principally measurable?*[6] *Or similarly, does every state correspond to a physically possible preparation procedure? Are there well-defined systems where an observable when measured would provide the answer to an undecidable problem? If in some physical context some observable cannot be measured, can we*

[5] To avoid subtleties in measure theory we assume that the set of measurement outcomes is finite. From a practical point of view this is justified by finite resolutions of any apparatus.

[6] Is there more to say than 'it's a matter of energy'?

always get rid of the 'redundant part' in the theoretical description? Is there a way of making sense out of these questions?

Problem 2 (Joint measurability vs. coexistence of effects) *A POVM can be viewed as a mapping from subsets of possible measurement outcomes to the set of positive semidefinite (psd) operators. More precisely, if X is a (not necessarily discrete) measurable set characterizing all possible measurement outcomes, then $X \supseteq x \mapsto P(x)$ is a POVM if it is a consistent assignment of psd operators with $P(x) \geq 0$ and $P(X) = \mathbb{1}$. For the above discrete case this means $P(x) = \sum_{\alpha \in x} P_\alpha$.*

Two POVMs P and Q are said to be coexistent *if there is a POVM R whose range (as a mapping into psd operators) includes the ranges of both P and Q. For von Neumann measurements as well as for the case of P and Q being two-valued, coexistence and joint measurability are known to be equivalent. Which is the general relation?*

Problem 3 (Uniqueness of joint measurements) *Given two jointly measurable POVMs $\{P_\alpha\}$ and $\{Q_\beta\}$. Find necessary and/or sufficient conditions for the joint observable $\{R_{\alpha\beta}\}$ to be unique.*

1.2 Composite systems & entanglement

Quantum theory replaces the Cartesian product, which classical theories use in order to describe composite systems, by the tensor product. The algebra of observables for a composite system is for instance given by the tensor product $\mathcal{A} \otimes \mathcal{B}$ of their observable algebras $\mathcal{A}$ and $\mathcal{B}$. The Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ then becomes a tensor product as well and the number of discrete degrees of freedom increases to $\dim \mathcal{H} = \dim \mathcal{H}_A \dim \mathcal{H}_B$. $\mathcal{H}_A$ is said to correspond to a *subsystem* of $\mathcal{H}$, but we may later loosen this notion to the extent that $\mathcal{H}_a$ is a *subsystem* of $\mathcal{H}$ if

$$\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b \oplus \mathcal{H}_c. \tag{1.7}$$

A *composite* system or *bipartite* system always refers to a tensor product. The simplest states of a composite system are *product states*, i.e., tensor products of individual states. States which do not factorize are said to exhibit *correlations*, as for some observables $\langle A \otimes B \rangle \neq \langle A \otimes \mathbb{1} \rangle \langle \mathbb{1} \otimes B \rangle$. This is usually written in a shorter form by introducing the *reduced density operator* ρ_A defined via $\mathrm{tr}\,[\rho_A A] = \mathrm{tr}\,[\rho(A \otimes \mathbb{1})]$ for all $A \in \mathcal{A}$. The mapping $\mathrm{tr_B} : \rho \mapsto \sum_b \langle b|\rho|b\rangle = \rho_A$ with $\{|b\rangle\}$ an orthonormal basis in $\mathcal{H}_B$ is called *partial trace* and the resulting ρ_A is sometimes called a *marginal* of ρ.

A special type of correlations are those which can be of classical origin: assume a state is given by a convex combination of product states[7]

$$\rho = \sum_x p_x \, \rho_x^{(A)} \otimes \rho_x^{(B)}, \tag{1.8}$$

[7] Here we could choose pure product states as well, since it follows from Caratheodory's theorem that a density matrix of the form (1.8) can always be decomposed into at most $\mathrm{rank}(\rho)^2$ pure product states. Note also that positivity of the p_x's is crucial: if we allow for $p_x \in \mathbb{R}$ then every state admits a representation like (1.8).

then one way of preparing ρ is to draw a classical random variable x from the probability distribution p and to prepare a product state $\rho_x^{(A)} \otimes \rho_x^{(B)}$ conditioned on x. With this motivation a state of the form (1.8) is called *classically correlated* or synonymously *separable* or *unentangled*. *Entanglement* is defined as its negation. An equivalent operational definition of classical correlations (to which we will get back later) is as those which can be generated by means of arbitrary local operations and classical communications.

Talking about 'local operations' and 'communication' already indicates that here and in the following we usually assume that the systems corresponding to different tensor factors are situated at distant locations. Such a *distant location paradigm* is not really necessary. However, it prevents us from misconceptions appearing when the apparently entangled systems are not spatially separated. This also avoids an extra discussion of systems of indistinguishable particles for which otherwise a proper definition of entanglement becomes a more subtle issue.

Obviously, pure states are separable if and only if they are product states. That is, any kind of correlation is then due to entanglement and in this sense non-classical. Since entanglement properties are independent of the choice of local bases, we may bring a pure state into a simple normal form:

Proposition 1.1 (Schmidt decomposition) *For every vector* $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ *with* $d = \min\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\}$ *there exist orthonormal bases* $\{e_j \in \mathcal{H}_A\}$ *and* $\{f_j \in \mathcal{H}_B\}$ *such that*

$$|\psi\rangle = \sum_{j=1}^{d} \sqrt{\lambda_j} |e_j\rangle \otimes |f_j\rangle, \quad \textit{with } \lambda_j \geq 0, \quad \sum_i \lambda_i = ||\psi||^2 \tag{1.9}$$

The coefficients $\{\sqrt{\lambda_j}\}$ are called *Schmidt coefficients* and the number of non-zero λ_j is the *Schmidt rank* of ψ. Prop.1.1 is easily proven by noting that a change of local bases in a general decomposition $|\psi\rangle = \sum_{a,b} C_{a,b} |a\rangle \otimes |b\rangle$ corresponds to left and right multiplication of the coefficient matrix C with two different unitaries. Hence, we can choose unitaries diagonalizing C such that the Schmidt coefficients are the singular values of C.[8] We will briefly summarize some implications of the Schmidt decomposition:

- *Reductions and purifications.* The reduced density operators of ψ are diagonal in the bases $\{e_j\}$ and $\{f_j\}$ and the λ_j's are their non-zero eigenvalues. Conversely, given any density matrix ρ_A with spectral decomposition $\rho_A = \sum_j \lambda_j |e_j\rangle\langle e_j|$ Eq.(1.9) provides a *purification* such that $\rho_A = \mathrm{tr}_B[|\psi\rangle\langle\psi|]$. The *minimal dilation space* $\mathcal{H}_B^{\min}$ has $\dim(\mathcal{H}_B^{\min}) = \mathrm{rank}(\rho_A)$. If $\psi \in \mathcal{H}_A \otimes \mathcal{H}_B^{\min}$ is a purification then all other purifications of ρ_A are of the form $|\psi'\rangle = (\mathbb{1} \otimes V)|\psi\rangle$ with $V \in \mathcal{B}(\mathcal{H}_B^{\min}, \mathcal{H}_B)$ an isometry.

- *Monogamy.* The Schmidt decomposition tells us that pure states cannot be correlated with any other system: a pure state as a reduced density

[8]This makes clear that the bases $\{e_j\}$ and $\{f_j\}$ depend on ψ.

operator means that any possibly larger, composite system is in a product state. Hence, correlations contained in pure states are monogamous and cannot be shared with any other system. This in sharp contrast to classical correlations (1.8). In fact, an equivalent way of characterizing classical correlations is (by the quantum de Finetti theorem) as those which can be symmetrically shared with an arbitrary number of other parties.

- *Mixed states.* As a density operator $\rho \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is itself a 'vector' (although not normalized) in the Hilbert-Schmidt Hilbert space, we can find a decomposition analogous to (1.9) of the form

$$\rho = \sum_j \sqrt{\mu_j} E_j \otimes F_j, \quad \mu_j \geq 0, \tag{1.10}$$

where $\{E_j\}$ and $\{F_j\}$ are sets of orthonormal operators w.r.t. the Hilbert-Schmidt scalar product. Then $\mathrm{tr}\left[\rho^2\right] = \sum_j \mu_j$ and ρ being classically correlated can be shown to imply $\sum_j \sqrt{\mu_j} \leq 1$.

Example 1.2 (Tricks with maximal entanglement) *If all Schmidt coefficients of a pure state ϕ are $\lambda_j = 1/d$, then ϕ is called* maximally entangled[9] *of dimension d. Every maximally entangled state is of the form $\phi = (\mathbb{1} \otimes U)|\Omega\rangle$ where U is any unitary and*

$$|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |jj\rangle. \tag{1.11}$$

Note that by using a basis of Hilbert-Schmidt orthogonal unitaries $\{U_j\}_{j=1,\ldots,d^2}$ we can construct an orthonormal basis of maximally entangled states $\{(U_j \otimes \mathbb{1})|\Omega\rangle\}$ for $\mathbb{C}^d \otimes \mathbb{C}^d$. Ω has a series of useful properties, like $\forall A, B \in \mathcal{B}(\mathbb{C}^d)$:

$$\langle\Omega|A \otimes B|\Omega\rangle = \frac{1}{d}\mathrm{tr}\left[A^T B\right], \quad (A \otimes \mathbb{1})|\Omega\rangle = (\mathbb{1} \otimes A^T)|\Omega\rangle, \tag{1.12}$$

where the transposition has to be taken in the Schmidt basis of Ω. This implies that (i) $(U \otimes \bar{U})|\Omega\rangle = |\Omega\rangle$ for all unitaries U and (ii) every pure state ψ with reduced density operator (ρ_B) can be written as[10]

$$|\psi\rangle = (\mathbb{1} \otimes R)|\Omega\rangle, \tag{1.13}$$

where $R = \sqrt{d\rho_B}V$ and the isometry V takes care of adjusting the different Schmidt bases.[11]

Identifying basis vectors in $\mathcal{H}_A$ and $\mathcal{H}_B$ (assuming they are of the same dimension) leads to a good friend of Ω: the flip *(or swap) operator, defined as $\mathbb{F}|ij\rangle = |ji\rangle$. They*

---

[9] The name 'maximally entangled' is justified for instance by the fact that every other state (of the same dimension) can be obtained with unit probability from a maximally entangled one by means of local operations and classical communication.

[10] Clearly, ρ_A has to be supported by the Schmidt basis of Ω.

[11] Eq.(1.13) is essentially the observation of Schrödinger who at that time was rather puzzled about this consequence of quantum mechanics. In his own words: *"It is rather discomforting that the theory should allow a system to be steered or piloted into one or the other type of state at the experimenter's mercy in spite of his having no access to it"*. The analogue of Eq.(1.13) in quantum field theory is the Reeh-Schlieder theorem where the subsystem is any open, bounded region in Minkowski-space and the vacuum plays the role of Ω.

can be interconverted by partial transposition[12] *as* $d|\Omega\rangle\langle\Omega|^{T_A} = \mathbb{F}$ *and the equivalent of (1.12) is*

$$\mathrm{tr}\left[(A \otimes B)\mathbb{F}\right] = \mathrm{tr}\left[AB\right], \quad (A \otimes \mathbb{1})\mathbb{F} = \mathbb{F}(\mathbb{1} \otimes A). \tag{1.14}$$

A useful representation of $\mathbb{F}$ *is in terms of any orthonormal*[13] *basis of operators* $\{G_i\}_{i=1..d^2}$ *as*

$$\mathbb{F} = \sum_{i=1}^{d^2} G_i \otimes G_i^\dagger. \tag{1.15}$$

Problem 4 (Separability problem) *Consider small bipartite systems associated to* $\mathcal{H} = \mathbb{C}^3 \otimes \mathbb{C}^3$ *or* $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^4$ *(but feel free to be more ambitious). Find an efficient and certifiable algorithmic way of deciding whether any density operator* $\rho \in \mathcal{B}(\mathcal{H})$ *is separable or not. Here, efficiency refers to a run-time which scales polynomially with the length of the specified input and also with the inverse of a reasonably chosen 'error'.*

1.3 Evolutions

Time plays a special role in quantum mechanics. Unlike most other quantities it is in a sense treated classically—described by a parameter rather than an operator[14]. When dealing with time evolution one typically distinguishes a special case: the evolution of *closed systems*. Closed systems (as opposed to *open systems*) are those for which the evolution is physically reversible. One might argue that this concept is at the root of the measurement problem and is therefore disputable, but this is a different story.

As quantum mechanics divides every physical process into preparation and measurement there are different (but in the end equivalent) 'pictures' depicting time evolution: we may treat the evolution as part of the preparation process, as part of the measurement process or split it up between the two. This leads to the *Schrödinger picture*, the *Heisenberg picture* or an *interaction picture*.

Closed systems and reversible evolutions Depending on the chosen picture we will describe the evolution for a given time as a linear transformation on observables $\mathcal{A} \ni A \mapsto u(\mathcal{A})$ (Heisenberg picture) or on states $\rho \mapsto u^*(\rho)$ (Schrödinger picture) where consistency imposes the relation

$$\mathrm{tr}\left[\rho u(A)\right] = \mathrm{tr}\left[u^*(\rho)A\right], \tag{1.16}$$

i.e., u and u^* are mutually *dual* or *adjoint* maps. Physically reversible evolutions should be described by mathematically reversible transformations (the converse

[12] The partial transpose of an operator $C \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is defined w.r.t. a given product basis as $\langle ij|C^{T_A}|kl\rangle := \langle kj|C|il\rangle$.

[13] meaning $\mathrm{tr}\left[G_i^\dagger G_j\right] = \delta_{ij}$.

[14] Pauli gave the following simple argument against the existence of a self-adjoint operator T satisfying $[T, H] = i\mathbb{1}$ with H a Hamiltonian: take T as the generator of a unitary group e^{ihT} and let this act on an eigenstate of H with eigenvalue E, then we get another eigenstate with energy $E + h$. The spectrum of H is thus continuous and in particular it cannot be semi-bounded. This argumentation, however, gets some loopholes when considering issues of the domains of the operators.

is not true). Moreover, the concatenation of evolutions is naturally considered to be associative (i.e., $(uv)w = u(vw)$) and should again lead to an admissible evolution. Consequently, the set of reversible evolutions is described by a *group* of linear transformations. This is in addition supposed to preserve the structure of the algebra $\mathcal{A}$ in the sense that for all $A, B \in \mathcal{A}$: $u(AB) = u(A)u(B)$, $u(A^\dagger) = u(A)^\dagger$ and $||u(A)|| = ||A||$. In other words the set of reversible evolutions is identified with the group $aut(\mathcal{A})$ of linear automorphisms of $\mathcal{A}$.

A subgroup of such automorphisms, which are then called *inner*, is formed by the unitary elements $U \in \mathcal{A}$ via $u(A) = U^\dagger AU$ or equivalently

$$\rho \mapsto U\rho U^\dagger. \tag{1.17}$$

Clearly, if two unitary elements in $\mathcal{A}$ differ only by a phase they give rise to the same automorphism. More generally, this is true iff they differ by a unitary $V \in \mathcal{A}$ which commutes with all $A \in \mathcal{A}$ (i.e., V is in the *center* of $\mathcal{A}$).

For $\mathcal{A} = \mathcal{M}_d$ (the case we are interested in the following) all automorphisms are inner. However, if we either restrict $\mathcal{A}$ or consider an infinite dimensional case this is no longer true in general. Take for instance the algebra of diagonal matrices. Then the only inner automorphism is the identity map $\mathrm{id}(A) = A$. Yet there is obviously a representation for which other automorphisms (like permutations) have the form $u(A) = U^\dagger AU$ with U a unitary, albeit none corresponding to an element in $\mathcal{A}$. Such automorphisms are called *spatial* (w.r.t. a given representation) or *unitarily implementable*.

We say that a reversible evolution is *homogeneous* in discrete time if after n time steps it is described by the n'th power of a unitary U which describes the evolution for a single time step. The pendant in continuous time is a one-parameter group of unitaries U_t satisfying $U_tU_r = U_{t+r}$ for all $t, r \in \mathbb{R}$. Imposing continuity in the form $(U_t - U_0)|\psi\rangle \to 0$ for $t \to 0$ and all $\psi \in \mathcal{H}$, *Stone's theorem* asserts that there is a unique self-adjoint operator H (the *Hamiltonian* or *infinitesimal generator*) such that

$$U_t = \exp{(iHt)}. \tag{1.18}$$

Let us finally have a look at a different approach towards the statement that reversible evolutions are associated to unitaries acting on Hilbert space. Imagine a reversible mapping from the set of normalized pure states $\mathcal{S} := \{\psi := |\psi\rangle\langle\psi|\} \subset \mathcal{B}(\mathcal{H})$ onto itself. Two obvious types of such mappings are $\psi \mapsto U\psi U^\dagger$ with U being a unitary and $\psi \mapsto \psi^T$ a transposition in some basis. The following theorem asserts that these are in fact the only possibilities which preserve the norms of scalar products:

Theorem 1.1 (Wigner's theorem) *Let $S : \mathcal{S} \to \mathcal{S}$ be a bijective mapping such that $\mathrm{tr}\,[S(\psi)S(\phi)] = \mathrm{tr}\,[\psi\phi]$ for all $\psi, \phi \in \mathcal{S}$. Then S falls in one of the two following classes:*

- Unitary. *There exists a unitary U such that $S(\psi) = U\psi U^\dagger$ for all $\psi \in \mathcal{S}$.*

- Antiunitary[15]. *There exists a unitary U such that $S(\psi) = U\psi^T U^\dagger$ for all $\psi \in \mathcal{S}$.*

Wigner's theorem has various nice consequences. A simple one is the following:

Corollary 1.1 (Spectrum preserving maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a linear map which is Hermitian (i.e., $T(X)^\dagger = T(X^\dagger)$ for all X) and spectrum preserving on Hermitian matrices. Then there is a unitary U such that T is either of the form $T(X) = UXU^\dagger$ or $T(X) = UX^T U^\dagger$.*

Proof Note first that preservation of the spectrum as a set implies its preservation counting multiplicities. This follows from continuity of linear maps: suppose the multiplicities are not preserved. Then there is at least one eigenvalue whose multiplicity is larger for the input operator than for its image. By perturbing the input we see that this would contradict continuity—so multiplicities have to be preserved as well.

Consider two Hermitian rank-one projections $\phi, \psi \in \mathcal{M}_d(\mathbb{C})$. The spectrum of the matrix $M := \phi + \psi$ is given by $1 \pm \sqrt{\mathrm{tr}[\phi\psi]}$. Since T is linear, Hermitian and spectrum preserving we obtain from applying T to M that $\mathrm{tr}[T(\phi)T(\psi)] = \mathrm{tr}[\phi\psi]$ so that application of Wigner's theorem completes the proof. □

Open systems and irreversible dynamics Irreversible dynamics is usually regarded as a consequence of considering only part of a larger system which (together with the unobserved part) undergoes reversible evolution. We will for the moment, however, not make use of this viewpoint and rather discuss briefly the general requirements of descriptions of physical evolution—reversible or not. Later we will find that all these evolutions have indeed a mathematical representation in terms of the mentioned reversible system-plus-environment-dynamics.

Consider an evolution which, in the Schrödinger picture, is described by a map $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$. When describing a physically meaningful evolution T should fulfill the following three conditions:

- *Linearity.* This is an inherent quantum mechanical requirement. We will see in Sec.1.4 that it is related to locality, i.e., the fact that a spatially localized action does not instantaneously affect distant regions. Linearity means that

$$\forall A, B \in \mathcal{B}(\mathcal{H}),\ c \in \mathbb{C} : \quad T(cA + B) = cT(A) + T(B). \tag{1.19}$$

- *Preservation of the trace.* T has to map density operators onto density operators. Since every element in $\mathcal{B}(\mathcal{H})$ is a linear combination of density operators we obtain by virtue of linearity

$$\forall A \in \mathcal{B}(\mathcal{H}) : \quad \mathrm{tr}[T(A)] = \mathrm{tr}[A]. \tag{1.20}$$

[15]This name stems from the fact that if we imagine the action of S on $\mathcal{H}$ then it involves complex conjugation, i.e., it is anti-linear rather than linear on $\mathcal{H}$.

- *Complete positivity.* Another consequence of linearity together with asking T to map density operators onto density operators is that it has to be a *positive map*, i.e., $T(A^\dagger A) \geq 0$ for all $A \in \mathcal{B}(\mathcal{H})$. Positivity alone is, however, not sufficient: consider $\mathcal{H}$ as part of a bipartite system so that the evolution of the larger system is described by $T \otimes \mathrm{id}$. That is, the additional system merely plays the role of a spectator as the evolution on this part is the trivial one. Yet $T \otimes \mathrm{id}$ should again be a positive map—a requirement which is stronger than positivity. So the relevant condition is *complete positivity* of T which means positivity of the map $T \otimes \mathrm{id}_n$ for all $n \in \mathbb{N}$ where id_n is the identity map on $\mathcal{M}_n$.

For the map $T^* : \mathcal{B}(\mathcal{H}') \to \mathcal{B}(\mathcal{H})$ describing the same evolution in the Heisenberg picture via $\mathrm{tr}\,[T(\rho)A] = \mathrm{tr}\,[\rho T^*(A)]$ the conditions linearity and complete positivity remain the same, only the trace preserving condition translates to *unitality*

$$T^*(\mathbb{1}) = \mathbb{1}. \tag{1.21}$$

A mapping which fulfills the above three conditions (either in Heisenberg or Schrödinger picture) is called a *quantum channel*. Quantum channels are the most general framework in which general input-output relations (i.e., black box devices) are described within quantum mechanics. It is crucial, however, that the mapping itself does neither depend on the input nor on its history. If such correlations appear, then the above black box description becomes inappropriate and either a larger system (including 'the environments memory') has to be taken into account or a consistent effective description has to be found. When talking about quantum channels in the following we will always mean the *Markovian* or synonymously *memory-less* case in which such correlations or dependencies do not occur.

Unlike positivity, complete positivity of a linear map is rather simple to check in finite dimensions:

Proposition 1.2 (Checking complete positivity) *A linear map $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ is completely positive iff $(T \otimes \mathrm{id}_d)(|\Omega\rangle\langle\Omega|) \geq 0$, where Ω is a maximally entangled state of dimension $d = \dim \mathcal{H}$.*

Proof Necessity of the condition follows from the definition of complete positivity. In order to prove sufficiency let us begin with an arbitrary n and density operator $\rho \in \mathcal{B}(\mathcal{H} \otimes \mathbb{C}^n)$. Then $(T \otimes \mathrm{id}_n)(\rho) \geq 0$ if the same holds true for all pure states appearing in the spectral decomposition of ρ. By Eq.(1.13) we can write every such pure state in the form $(\mathbb{1}_d \otimes R)|\Omega\rangle$ with some $R \in \mathcal{B}(\mathcal{H}, \mathbb{C}^n)$. Together with the fact that positive semidefiniteness of an operator B implies that of $ABA^\dagger$, this concludes the proof. □

We will later on see that the operator $(T \otimes \mathrm{id}_d)(|\Omega\rangle\langle\Omega|)$ called *Jamiolkowski state* (or *Choi matrix* when multiplied by d) contains in fact all information about T.

Probabilistic operations and Instruments The trace preserving condition (or unitality in the Heisenberg picture) reflects the normalization of probabilities. That is, quantum channels (trace preserving, completely positive linear maps) describe quantum operation which succeed with unit probability. When measurements are involved we may, however, relax this condition and consider as well linear completely positive maps T which are *trace non-increasing* in the sense that $\mathrm{tr}\,[T(A)] \leq \mathrm{tr}\,[A]$ for all $A \in \mathcal{B}(\mathcal{H})$ (or equivalently $T^*(\mathbb{1}) \leq \mathbb{1}$). The interpretation of such a map is that it only succeeds upon an input state ρ with probability $p = \mathrm{tr}\,[T(\rho)]$ and in this case yields the output density matrix $T(\rho)/p$. Such maps are sometimes called *probabilistic* or *stochastic* (although this notion is more than doubly occupied) or *filtering operations*. This leads directly to the concept of an *instrument*:

In the Schrödinger picture an *instrument* is a set of completely positive linear maps $\{T_i\}$ whose sum $\sum_i T_i$ is trace preserving (i.e., $\sum_i T_i^*(\mathbb{1}) = \mathbb{1}$). The label i (which might as well be a measurable continuous set) can be interpreted as the outcome of a measurement which occurs with probability $p_i = \mathrm{tr}\,[T_i(\rho)]$ and conditioned on which the density matrix transforms as $\rho \mapsto T_i(\rho)/p_i$.

Instruments encompass both, the concept of a quantum channel (when ignoring i and considering only $T = \sum_i T_i$) and the concept of a POVM (when ignoring the state at the output and only dealing with the set of effect operators $\{T_i^*(\mathbb{1})\}$).

Note that every POVM corresponds to an entire equivalence class of instruments rather than to a single one. This reflects the fact that a POVM alone does not determine the state after the measurement. A simple type of instruments realizing any POVM with effect operators $\{P_i\}$ is the *Lüders instrument* for which

$$T_i(\rho) = \sqrt{P_i}\rho\sqrt{P_i}. \tag{1.22}$$

But obviously the instruments $T_i(\rho) = U\sqrt{P_i}\rho\sqrt{P_i}U^\dagger$ (with U any unitary) or $T_i(\rho) = \mathrm{tr}\,[\rho P_i]$ and many others correspond to the same POVM. The last example may describe situations in which the quantum system 'disappears' after the measurement.

Let us finally see how we can use the concept of an instrument to get a strengthened and operational version of Eq.(1.13):

Proposition 1.3 (Quantum steering) *Let $\rho \in \mathcal{B}(\mathcal{H}_A)$ be a density operator with purification $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. For every convex decomposition $\rho = \sum_i \lambda_i \rho_i$ there is an instrument $\{T_i : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B)\}$ acting on Bob's system such that*

$$\lambda_i \rho_i = \mathrm{tr}_B\left[(\mathrm{id} \otimes T_i)(|\psi\rangle\langle\psi|)\right]. \tag{1.23}$$

Proof W.l.o.g. we can assume that ψ is a minimal purification ($\dim \mathcal{H}_B = \mathrm{rank}[\rho] =: d$) and write $|\psi\rangle = (\sqrt{d\rho} \otimes \mathbb{1})|\Omega\rangle$ where Ω is a maximally entangled state of dimension d with the same Schmidt bases as ψ. Moreover, we exploit that for every linear map T_i:

$$(\mathrm{id} \otimes T_i)(|\Omega\rangle\langle\Omega|) = (\theta T_i^* \theta \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|), \tag{1.24}$$

where $\theta : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B)$ is the matrix transposition w.r.t. the Schmidt basis of Ω.[16] Putting these two things together we obtain for any linear map T_i that

$$\mathrm{tr}_B\left[(\mathrm{id}\otimes T_i)(|\psi\rangle\langle\psi|)\right] = \sqrt{\rho}\,\left(T_i^*(\mathbb{1})\right)^T\sqrt{\rho}. \tag{1.25}$$

Therefore, an instrument will satisfy Eq.(1.23) if it corresponds to a POVM $\{T_i^*(\mathbb{1})\}$ which fulfills $T_i^*(\mathbb{1}) = \lambda_i \rho^{-1/2}\rho_i^T\rho^{-1/2}$ on the range of ρ. One possible choice is thus $T_i^*(A) = K_i A K_i^\dagger$ with $K_i := \rho^{-1/2}\sqrt{\lambda_i\rho_i^T} \oplus (\delta_{1,i}\mathbb{1})$ where the second term in the direct sum acts on the kernel of ρ. □

The interpretation of this is the following: assume Alice and Bob share the state ψ, then Bob can generate not only any single state on Alice side (which is in the support of her local density operator) but he can engineer every ensemble $\{\lambda_i\rho_i\}$ consistent with her reduced density matrix. In order to 'steer' the state ρ_i on Alice's side, Bob can apply the described instrument. For every single event he has to report to Alice whether or not his operation was successful (i.e., he received an outcome i) which happens with probability λ_i. If Alice then performs any statistical experiment using only events for which Bob reported success, her state will be ρ_i. The following provides the maximum probability of success:

Proposition 1.4 (Maximal weight in convex decomposition) *Let ρ and ρ_1 be two density operators acting on the same space. There exists a convex decomposition of the form $\rho = \sum_i \lambda_i\rho_i$ with $\lambda_1 > 0$ iff $\mathrm{supp}(\rho_1) \subseteq \mathrm{supp}(\rho)$ and*

$$\lambda_1 \leq ||\rho^{-1/2}\rho_1\rho^{-1/2}||_\infty^{-1}, \tag{1.26}$$

where the inverse is taken on the range of ρ.

Proof The necessity of the condition for the supports is evident since adding positive operators can never decrease the support. The existence of a decomposition is equivalent to $\rho - \lambda_1\rho_1 \geq 0$ which can be rewritten as $\mathbb{1} \geq \lambda_1\rho^{-1/2}\rho_1\rho^{-1/2}$ from which the statement follows. □

1.4 Linearity and locality

Locality is one of the most basic concepts in physics. It runs under various names like *Einstein locality*, *no-signaling condition* or the non-existence of a *Bell telephone*[17]. Its meaning is essentially that a spatially localized action does not (instantaneously) influence distant parts. Evidently, this is a crucial ingredient when we want to talk about small systems without always having to take the entire rest of the universe into account.

[16] Eq.(1.24) is a simple exercise which can be solved using the tools presented in Example 1.2. Of course, it can also be directly seen from a Kraus decomposition of T_i.

[17] Did anybody tell this the Bell telephone company?

Quantum mechanics is by construction a local theory. [18] In order to see this, consider a bipartite system prepared in any initial state $\rho \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and act with a quantum channel T on Bob's system only. The reduced density operator on Alice's side is then given by

$$\mathrm{tr}_B[(\mathrm{id} \otimes T)(\rho)] = \mathrm{tr}_B[\rho], \tag{1.27}$$

where we used that $T^*(\mathbb{1}) = \mathbb{1}$ which together with the tensor product structure implies that Alice is not affected by Bob's action, i.e., ρ_A is independent of T.

In fact, quantum mechanics exhibits a stronger form of locality/independence, sometimes called *Schlieder property*: not only is Bob unable to communicate to Alice by local operations, but Alice is unable to gain any information about Bob's system (when she is constrained to local operations and does not know the overall state). This follows from the simple fact that every reduced density matrix ρ_A is compatible with every other reduced density matrix ρ_B since there is always a joint state (like $\rho_A \otimes \rho_B$) which returns both as marginals.[19]

In the following we will see that locality (in the sense of no-signaling) is strongly connected to linearity of quantum mechanics. The argument for this comes in two flavors: (i) making explicit use of the formalism of quantum mechanics one can show that any sort of non-linearity would imply a breakdown of locality (unless we modify some of the framework, e.g., by imposing additional restrictions). (ii) The existence of certain observable correlations predicted by quantum mechanics allows to rule out specific non-linear devices. So, while the second argument is weaker in the sense that it relates only a particular observable form of non-linearity to non-locality, it is stronger in that it applies to every theory which predicts the required correlations.

Non-linear evolutions and mixed state analyzers Suppose there is a physical device which is characterized by a *non-linear* map N on density operators, i.e., one for which there exists a convex decomposition of a density operator $\rho = \sum_i \lambda_i \rho_i$ such that

$$\sum_i \lambda_i N(\rho_i) \neq N\Big(\sum_i \lambda_i \rho_i\Big). \tag{1.28}$$

Such a device can be used to break the no-signaling condition. In order to see this recall the steering result of Prop.1.3 and assume that ρ is Alice's reduced state of a bipartite pure state ψ. If Bob leaves his part untouched (or equivalently, discards it) and Alice applies her non-linear device, then $N(\rho)$ describes her state. If, however, Bob applies an instrument $\{T_i\}$ tailored to prepare ρ_i on Alice's side with probability λ_i, then Alice's state will be $\sum_i \lambda_i N(\rho_i)$. As this is,

[18] There is a very unfortunate use of the notion 'non-locality' as an attribute of quantum mechanics. This is used in order to express that certain correlations predicted by quantum mechanics do not admit a description within a *local* hidden variable theory.

[19] Usually, this is invoked in the context of algebraic quantum field theory where the statement is that $AB = 0$ implies that either $A = 0$ or $B = 0$ if A and B are elements of observable algebras corresponding to spacelike separated regions.

by the non-linear character of Alice's mysterious machine, different from $N(\rho)$, she is able to distinguish the two cases with a probability of success which is strictly bigger than one half. That is, by only looking at her local system she can gain information about whether or not Bob applied the instrument—something in conflict with the no-signaling condition.

The same type of argument holds for *mixed state analyzers*, i.e., hypothetical devices which are capable of telling us whether the 'true' decomposition of a state is $\rho = \sum_i \lambda_i \rho_i$ or $\rho = \sum_i \tilde{\lambda}_i \tilde{\rho}_i$. If such a machine existed (and Alice possesses it) then we could again think about Bob transmitting information to Alice by choosing one out of two instruments to apply on his subsystem.

The above arguments may not be on firm enough grounds to force us to conclude that quantum mechanics does not allow non-linear processes. However, they show at least that a naive incorporation of such processes comes in conflict with locality unless we impose additional restrictions and/or modify the framework.

A sequence of impossible machines Some non-linear devices can be specified by their observable action and thus without going too much into the details of the theory. An example of this kind is the *cloning device* for which it is usually argued that its impossibility within quantum mechanics follows from the linearity of quantum mechanical evolutions. While this is certainly true, this reasoning misses an important point. After all, the Liouville equation of motion for classical mechanics is linear as well but we would hardly argue that this implies impossibility of making a copy of a classical system. Evidently, the difference is the absence of a theory of measurements in classical mechanics, i.e., the tacit assumption that we can in principle observe any detail of a system in classical mechanics without causing disturbance. A more careful derivation of the quantum *no-cloning theorem* should thus take measurements and re-preparations into account. It turns out, however, that the validity of the no-cloning theorem goes far beyond the particular framework of quantum mechanics and is implied by the no-signaling condition when supplemented by the existence of certain observable correlations. This argument can be nicely embedded in a remarkable chain: a series of hypothetical devices where the existence of each would imply the possibility of all subsequent ones, and where the last in the row is the *Bell telephone* – a device capable of breaking the no-signaling condition. So if we want to retain locality we are forced to accept no-go statements for all the devices of the hierarchy and within all theories which are consistent with the required correlations. The hierarchy is build up of the following devices:

- *Classical teleportation* is the process of converting an unknown system into classical information (a bit string or punchcard) and then re-preparing it again such that no statistical test could distinguish between the original and the re-prepared system.[20] A classical teleportation device thus con-

[20]The difference between classical teleportation and *entanglement-assisted teleportation* (often just called *teleportation*) is that in the latter case the re-preparation requires more than

sists out of two parts: a measurement device and a repreparation machine.

- *Cloning*, i.e., generating statistically indistinguishable duplicates for arbitrary systems, is clearly possible if we are able to do classical teleportation—we just make a copy of the classical bit string and feed it into two repreparation machines. Cloning in turn enables:
- *Measuring without disturbance*, meaning that every type of measurement can be performed without changing the statistics of any other subsequent measurement. A slightly weaker version of this is the possibility of
- *Joint measurements* of any pair of observables. This requires that the statistics of the joint measurement is the same as the one for individual measurements. The asumed existence of a joint measurement device implies the possibility of a
- *Bell telephone*, a hypothetical device which enables two distant parties to break the no-signaling condition.

Whereas all previous implications are fairly obvious, the last one requires an additional ingredient—specific correlations between the two parties—and some background knowledge: suppose Alice is given a joint measurement device which she uses to measure two observables A_1 and A_2 yielding outcomes a_1, a_2 with probability $p(a_1, a_2)$. If Bob, at a distance, measures an observable B_1 with outcome b_1, then they observe in a statistical experiment a joint probability distribution $p(a_1, a_2, b_1|B_1)$ so that[21]

$$p(a_1, a_2) = \sum_{b_1} p(a_1, a_2, b_1|B_1). \tag{1.29}$$

However, in a no-signaling theory this has to be independent of Bob's chosen observable, i.e., a possibly measured $p(a_1, a_2, b_2|B_2)$ has to have the same marginal $p(a_1, a_2)$. Assume that Bob chooses B_1 or B_2 at random so that they measure both triple distributions albeit not simultaneously as they come from disjoint statistical ensembles. Nevertheless we can write down a joint distribution

$$p(a_1, a_2, b_1, b_2) := \frac{p(a_1, a_2, b_1|B_1)p(a_1, a_2, b_2|B_2)}{p(a_1, a_2)}, \tag{1.30}$$

which by construction correctly returns all measured distributions as marginals. As a result, the existence of a joint measurement device implies a joint probability distribution (1.30) if the no-signaling condition is invoked. The point is now that not all pair distributions $p(a_i, b_j|A_i, B_j)$ allow such a joint distribution. The conditions under which a joint distribution exists are called *Bell inequalities*. These are, however, known to be violated by certain measurements

just the classical information, namely the remaining part of an entangled state.

[21] As common in classical probability theory we use the notation $p(\cdot|\cdot)$ where right of the dash is the *condition* which the probability is subject to. In our case this is the observable which is measured.

performed on particular entangled quantum states. Accepting this as a fact, the conclusion must be: either the no-signaling condition is violated or a universal joint measurement device, and with it the entire above hierarchy, is impossible. Even on Sundays such strong implications come rarely so cheap.

The quantum Boltzmann map is a special case of a non-linear map—an analogue of a map Boltzmann introduced in classical statistical mechanics in connection to the 'Stosszahlansatz'. It is defined as $T : \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \rightarrow \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ via

$$T_{QB}(C) = \text{tr}_B[C] \otimes \text{tr}_A[C], \tag{1.31}$$

and sometimes also called *Zwanzig projection*. For a density operator ρ we get $T_{QB}(\rho) = \rho_A \otimes \rho_B$, i.e., T_{QB} 'breaks' all correlations between the two subsystems while retaining the local states. Its appearance in statistical mechanics stems from the fact that it maximizes the global entropy under given local constraints.

1.5 Do we need complex numbers?

No, but there are good reasons to use them.

The algebras and Hilbert spaces used in the formulation of quantum mechanics are almost exclusively over the field of complex numbers. Of course, we may express everything in terms of real and imaginary parts, and thus by real numbers, but can we do so without ever writing 'i'? This is, actually, simple (although in the end not recommended): consider the mapping $\mathcal{M}_d(\mathbb{C}) \ni A \mapsto \tilde{A} \in \mathcal{M}_{2d}(\mathbb{R})$ given by

$$\tilde{A} := R(A \oplus \bar{A})R^\dagger = \frac{1}{2}\begin{pmatrix} \text{Re}A & -\text{Im}A \\ \text{Im}A & \text{Re}A \end{pmatrix} \text{with } R := \frac{1}{\sqrt{2}}\begin{pmatrix} \mathbb{1} & \mathbb{1} \\ -i\mathbb{1} & i\mathbb{1} \end{pmatrix}. \tag{1.32}$$

Sine R is a unitary we obtain for the trace of a product of matrices:

$$\text{tr}\left[\prod_i \tilde{A}_i\right] = 2\text{Re tr}\left[\prod_i A_i\right], \tag{1.33}$$

and if A is Hermitian, positive or unitary[22], $\tilde{A}$ will be so as well. Furthermore, a complex POVM will be mapped onto a real one and by introducing a real density operator $\rho_r := \tilde{\rho}/2$ a unitary time-evolution of the expectation value of a Hermitian operator A can by Eq.(1.33) be written in a purely real form:

$$\text{tr}\left[\rho U A U^\dagger\right] = \text{tr}\left[\rho_r \tilde{U} \tilde{A} \tilde{U}^T\right]. \tag{1.34}$$

If ρ corresponds to a pure state, then ρ_r will have rank two. However, we can safely replace it by any (real) rank-one projection $|\psi\rangle\langle\psi|$ which is in the support

[22] An element of $U(n)$ will be mapped onto a special orthogonal matrix which is symplectic as well. Eq.(1.32) is then nothing but the group isomorphism $U(n) \simeq Sp(2n, \mathbb{R}) \cap SO(2n, \mathbb{R})$.

of ρ_r so that the expectation value in (1.34) wont change. In this way we may even replace the Schrödinger equation by a real counterpart $\partial_t|\psi\rangle = H|\psi\rangle$, where H is real and anti-symmetric. The ambiguity in choosing ψ in the support of ρ_r may be compared to the irrelevance of a global phase in the complex framework.

The main drawback of the discussed real representation is the description of composite systems. Clearly, $\widetilde{A \otimes B} \neq \tilde{A} \otimes \tilde{B}$ so that composite systems are no longer associated to a tensor product—we have to make use of the tensor product in the complex framework. Alternatively we could try to double the dimension for each tensor factor separately (and actually use $\tilde{A} \otimes \tilde{B}$). In this case, however, the total dimension would not only depend on the number of degrees of freedom but, quite inconveniently, also on the way we group them into subsystems (something which might change during a calculation or which we might not want to fix beforehand).

Of course, one can think about other ways of representing the expectation values of complex quantum mechanics by using real Hilbert spaces. In the end we are, however, always faced with the problem that the dimensions of tensor products do not mach unless we invoke some unpleasant book-keeping.

Problem 5 (Quaternionic quantum mechanics) *Real and complex quantum mechanics are equivalent in the sense that they allow the same observable correlations (incl. violations of Bell inequalities). Are complex and quaternionic quantum mechanics equivalent?*[23] *What about no-go theorems such as the impossibility of (complex) quantum bit commitment?*

1.6 The algebraic framework

"I would like to make a confession which may seem immoral: I do not believe in Hilbert space anymore" [J.von Neumann in a letter to Birkhoff, 1935][24]

A glimpse beyond 'Hilbert-space-centrism' Most literature/work on quantum mechanics is based on Hilbert space formalism. This has, in the early days of quantum mechanics, been brought forward by Dirac and (independently and with a rather different flavor) by von Neumann. There is, however, a branch of quantum physics which departs from using a (separable) Hilbert space as starting point and which focuses more on properties of algebras assigned to local observables. This point of view, which actually builds up on von Neumann's work on operator algebras, mainly arose in the third quarter of the 20th century and is called *algebraic quantum (field) theory*. But why change the view point? What's wrong with Hilbert spaces?

First of all, there is hardly any difference between the two points of view when dealing with finite dimensional quantum systems. So everything we need

[23] Using a representation of quaternions in terms of complex valued 2×2 matrices, a simple translation between the quaternionic and complex world can be found for single systems. However, non-cummutativity complicates the description of composite systems.

[24] quoted from Fred Kronz: *von Neumann vs. Dirac* in the *The Stanford Encyclopedia of Philosophy* — actually a nice article.

for our (finite-dimensional) purposes is a little bit of useful terminology and some representation theory for finite dimensional $*$-algebras.

The 'algebraic framework' becomes beneficial in the context of a mathematically rigorous treatment of infinite systems. The 'infinity' may either arise from thermodynamical limits (infinite lattice systems) and/or from the investigation of quantum fields in continuous space(-time). In both cases there is an assignment of an algebra of observables to each region in space(-time). This assignment should be (i) *consistent* in the sense that algebras of subregions correspond to subalgebras and (ii) *local* in the sense that algebras assigned to space-like separated regions commute. For a finite lattice system we may still think about operators acting on a tensor-product of Hilbert spaces (with one tensor factor per site). However, the need to go beyond this framework becomes apparent in continuous space-time where the possibility of considering finer-and-finer subdivisions comes in conflict with representations on separable Hilbert spaces.

Some appealing properties of the algebraic framework are that (i) it treats quantum and classical systems on the same ground (ii) it allows to some extent to circumvent the problem of unitarily inequivalent representations, and (iii) it describes the occurence of super-selection rules as a consequence of reducible representations. Having said this, one has to admit, however, that algebraic quantum (field) theory is, so far, more a principal framework for physical theories than the origin (or even host) of a full-fledged physical theory. It provides solid grounds for discussing axiomatic and foundational questions, but cross-sections and line-widths are a different story.

The following intends to provide a pedestrian introduction to the basic notions around C^*-algebras. The general discussion in the next paragraph is disconnected from the other chapters (so, feel free to skip), but it might be useful for connecting the discussed topics to literature which goes beyond the finite-dimensional case. The focus in the second part will be on the finite-dimensional world again.

Algebras in a nutshell As a guideline it is helpful to think about algebras as 'generalizations of matrices', with the set of matrices as a special case (and our main focus—at least outside the present paragraph). An algebra $\mathcal{A}$ over the field of complex numbers $\mathbb{C}$ is a set which is closed under scalar multiplication, addition and multiplication. That is, if $A, B \in \mathcal{A}$ and $c \in \mathbb{C}$, then cA, $A+B$ and AB are elements of $\mathcal{A}$ as well. We will include associativity and the existence of a unit element ($\mathbb{1}A = A$) in our definition of an algebra. $\mathcal{A}$ is said to be *abelian* or *commutative* if $AB = BA$ for all $A, B \in \mathcal{A}$ and it is said to be a *-*algebra* if there is an involution $^\dagger : \mathcal{A} \to \mathcal{A}$ which fulfills all basic requirements known from the Hermitian adjoint of matrices: $(AB)^\dagger = B^\dagger A^\dagger$, $(cA)^\dagger = \bar{c}A^\dagger$ and $(A+B)^\dagger = A^\dagger + B^\dagger$. An element $A \in \mathcal{A}$ is called *normal* if $AA^\dagger = A^\dagger A$ and *selfadjoint* if $A = A^\dagger$.

A *normed algebra* is equipped with a norm $||\cdot||$ which satisfies, apart from the usual requirements for vector space norms, the product inequality $||AB|| \leq$

$||A||\,||B||$. If $\mathcal{A}$ is complete w.r.t. this norm (i.e., it contains all limits of Cauchy sequences) then it is called a *Banach algebra*. For a *Banach* **-algebra* we require in addition $||A^\dagger|| = ||A||$.

The *spectrum* spec(A) can be defined in purely algebraic terms as the set of all $\lambda \in \mathbb{C}$ for which $(\lambda \mathbb{1} - A)$ is not invertible. It satisfies $\text{spec}(AB)\backslash 0 = \text{spec}(BA)\backslash 0$ and selfadjoint elements for which $\text{spec}(A) \subseteq \mathbb{R}^+$ are called *positive*. The *spectral radius* is defined by $\varrho(A) := \sup\{|\lambda| \mid \lambda \in \text{spec}(A)\}$. In every Banach algebra the spectral radius fulfills $\varrho(A) \leq ||A||$ and

$$\varrho(A) = \lim_{n\to\infty} ||A^n||^{1/n}. \tag{1.35}$$

Conversely, we can define a norm from the spectral radius by $||A||_\infty := \sqrt{\varrho(A^\dagger A)}$. This satisfies

$$||A^\dagger A||_\infty = ||A||_\infty^2, \tag{1.36}$$

which means for selfadjoint elements $||A||_\infty = \varrho(A)$ and thus $||A||_\infty \leq ||A||$. Note that due to Eq.(1.35) every norm wich fulfills the so-called C^**-norm property* in Eq.(1.36) has to coincide with $||\cdot||_\infty$. Hence, a Banach *-algebra norm which fulfills condition (1.36) is unique (and called C^**-norm*).

A Banach *-algebra whose norm satisfies condition (1.36) is called C^**-algebra*. C^*-algebras exhibit a number of useful properties, for example:

- The spectrum of an element A is the same in every C^*-subalgebra containing A.[25]
- Positive elements are precisely those which can be written as $A = B^\dagger B$ for some $B \in \mathcal{A}$. They form a convex cone whose elements have a unique positive square root which can in turn be characterized in a purely algebraic way.
- Homomorphisms are order preserving (i.e., they map positive elements onto positive elements) and norm decreasing (hence, continuous). Isomorphisms are norm preserving and the range of every homomorphism is again a C^*-algebra.
- Every derivation is bounded.

A linear functional $\rho : \mathcal{A} \to \mathbb{C}$ on a C^*-algebra is called *positive* if $\rho(A^\dagger A) \geq 0$ for all $A \in \mathcal{A}$ and *normalized* if $\rho(\mathbb{1}) = 1$. A *state* is a positive, normalized linear functional. It is called *faithful* if $\rho(A^\dagger A) > 0$ unless $A = 0$ and *pure* if it does not admit a non-trivial convex decomposition into other states.

So far, our discussion remained abstract—before any concrete representation. Two paradigmatic examples of concrete C^*-algebras are (i) the space of complex-valued continuous functions on a compact space, and (ii) the space of bounded operators acting on a Hilbert space. A result by Gelfand shows that every commutative C^*-algebra is isomorphic to one of type (i). (ii) instead sets

[25] In general, if a $\mathcal{B}$ is a subalgebra of $\mathcal{A}$ and $A \in \mathcal{B}$, then the spectrum of A depends on whether we regard A as an element of $\mathcal{A}$ or $\mathcal{B}$, though $\text{spec}_{\mathcal{A}}(A) \subseteq \text{spec}_{\mathcal{B}}(A)$ holds.

the stage for general representations. In fact, every C^*-algebra has a faithful (i.e., one-to-one) representation within bounded operators on some (not necessarily separable) Hilbert space. The C^*-norm then becomes the operator norm on the level of the representation. That is, if we denote the representation by $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$, then

$$||A||_\infty = \sup_{\psi\in\mathcal{H}} \frac{||\pi(A)\psi||}{||\psi||}, \quad \text{with} \quad ||\psi|| = \sqrt{\langle\psi|\psi\rangle}. \tag{1.37}$$

The *Gelfand-Neumark-Segal (GNS) theorem* constructs a $\mathcal{B}(\mathcal{H})$-representation of a C^*-algebra starting from a state $\rho : \mathcal{A} \to \mathbb{C}$. The main idea is to define a scalar product via $\langle A, B\rangle := \rho(A^\dagger B)$. If ρ is not faithful, one has to define the elements of the sought Hilbert space in terms of equivalence classes and consider the set $\mathcal{A}/\mathcal{J}$ with $\mathcal{J} = \{A \in \mathcal{A}|\rho(A^\dagger A) = 0\}$. By completing $\mathcal{A}/\mathcal{J}$ we get a Hilbert space $\mathcal{H}_\rho$. The action of any $A \in \mathcal{A}$ on vectors in $\mathcal{H}_\rho$ is then represented by $\pi_\rho(A)|B\rangle := |AB\rangle$. In this way we can write

$$\rho(A) = \langle \mathbb{1}|\pi_\rho(A)|\mathbb{1}\rangle, \tag{1.38}$$

and interpret $|\mathbb{1}\rangle$ as the vector state which represents ρ in $\mathcal{H}_\rho$. The so constructed GNS-representation is irreducible iff ρ is pure. Physically, the GNS construction is closely related to constructing a Hilbert space from a 'vacuum state' by acting on it with creation operators (and thereby adding particles).

From now on let us assume that a faithful representation has been chosen, i.e., that $\mathcal{A}$ is a *concrete* algebra of bounded operators acting on some Hilbert space $\mathcal{H}$. The commutant $\mathcal{A}' := \{C \in \mathcal{B}(\mathcal{H})|\forall A \in \mathcal{A} : AC = CA\}$ is then the set of operators which commute with every element of $\mathcal{A}$. The *center* of $\mathcal{A}$ is the intersection $\mathcal{C} := \mathcal{A} \cap \mathcal{A}'$.

Sometimes C^*-algebras are too general and more structure is either required or provided. In fact, C^*-algebras can exhibit some 'unpleasant' properties: they may not contain projections other then 0 and $\mathbb{1}$, the *bicommutant* $\mathcal{A}'' := (\mathcal{A}')'$ may be larger than $\mathcal{A}$, and a tensor product of C^*-algebras can be ambiguous since several norms may have the C^*-property.[26] All these things a 'cured' by restricting to a subset of C^*-algebras called *von Neumann algebras*. Von Neumann algebras can be defined as bicommutant of a $^*-$algebra $\tilde{\mathcal{A}} \subseteq \mathcal{B}(\mathcal{H})$ or, equivalently, as the weak or strong closure of $\tilde{\mathcal{A}}$. A von Neumann algebra is generated (and classified) by its projections.

A state on a von Neumann algebra $\mathcal{A}$ is called *normal* if it admits a density operator representation, i.e., with some abuse of notation $\rho(A) = \mathrm{tr}\,[\rho A]\,, \forall A \in \mathcal{A}$. $\mathcal{A}$ is called a *factor* if its center contains only multiples of the identity. These can be regarded as basic building blocks as every von Neumann algebra admits a decomposition into factors.

[26] No, there isn't a contradiction with the uniqueness of the C^*-norm since that requires that the algebra is complete w.r.t. the considered norm—this is not fulfilled by the algebraic tensor product of two C^*-algebras unless one of them is finite dimensional.

Finite dimensional $*$-algebras and conditional expectations Since the topologies which make the difference between C^* and von Neumann algebras coincide in finite dimensions, these types of algebras coincide there, too. Similarly, every normed space becomes a Banach space as finite dimensional vector spaces are automatically complete. In this way every finite dimensional normed *-algebra is automatically a von Neumann/C^* algebra w.r.t. the norm induced by the spectral radius.

Every finite dimensional von Neumann algebra is isomorphic to a direct sum of full matrix algebras. More precisely, if a von Neumann algebra $\mathcal{B} \subseteq \mathcal{A} \simeq \mathcal{M}_d(\mathbb{C})$ is a subalgebra of a finite dimensional matrix algebra, then there is a unitary U such that

$$\mathcal{B} = U\Big(0 \oplus \bigoplus_{k=1}^{K} \mathcal{M}_{d_k} \otimes \mathbb{1}_{m_k}\Big)U^\dagger. \tag{1.39}$$

The K-fold direct sum corresponds to a decomposition into K *factors*. Each factor is isomorphic to a full matrix algebra of dimension d_k^2 which appears with multiplicity m_k. Hence, the commutant of the k'th factor is a full matrix algebra of dimension m_k^2 and multiplicity d_k. We will denote the decomposition of the Hilbert space to which Eq.(1.39) leads to as $\mathbb{C}^d = \mathcal{H}_0 \oplus \bigoplus_{k=1}^K \mathcal{H}_k$ with $\mathcal{H}_k = \mathcal{H}_{k,1} \otimes \mathcal{H}_{k,2} \simeq \mathbb{C}^{d_k} \otimes \mathbb{C}^{m_k}$. Furthermore, we will denote by $V_k : \mathbb{C}^d \to \mathcal{H}_k$ the corresponding isometries which are mutually orthogonal in the sense that $V_k V_l^\dagger = \delta_{kl} \mathbb{1}_{\mathcal{H}_k}$. Note that $\mathcal{B}$ is unital iff $\sum_k V_k V_k^\dagger = \mathbb{1}$, i.e., iff there is no $\mathcal{H}_0$.

The representation in Eq.(1.39) shows us how a subalgebra $\mathcal{B}$ can be embedded in a larger matrix algebra $\mathcal{A} \supseteq \mathcal{B}$ and it provides a simple form for projective (in the sense of idempotent and surjective) mappings $E : \mathcal{A} \to \mathcal{B}$ which allow us to 'restrict' every element of $\mathcal{A}$ to $\mathcal{B}$ and act as the identity on $\mathcal{B}$. If we require E to be positive, then the structure in Eq.(1.39) dictates the form of E:

Proposition 1.5 (Conditional expectations) *Consider a unital *-subalgebra $\mathcal{B} \subseteq \mathcal{A} \simeq \mathcal{M}_d(\mathbb{C})$. If a positive linear map $E : \mathcal{A} \to \mathcal{B}$ satisfies $E(b) = b$ for all $b \in \mathcal{B}$, then (using the above notation) there exist density operators $\{\rho_k \in \mathcal{B}(\mathcal{H}_{k,2})\}$ such that E has the form*

$$E(A) = \sum_{k=1}^{K} V_k^\dagger \left[\mathrm{tr}_{k,2}\left[(V_k A V_k^\dagger)(\mathbb{1}_{d_k} \otimes \rho_k)\right] \otimes \mathbb{1}_{m_k}\right] V_k. \tag{1.40}$$

Proof Define projections $P_k := V_k V_k^\dagger \in \mathcal{B}$ and a 'pinching' map $E_0 : \mathcal{A} \to \mathcal{A}$ via $E_0(A) := \sum_{k=1}^K P_k A P_k$. Since E_0 acts as the identity on $\mathcal{B}$ and the image of E is in $\mathcal{B}$ we have $E_0 E = E$. In order to see that $E E_0 = E$ holds as well, consider an operator of the form $A = \lambda P_k + \lambda^{-1} P_l + Q + Q^\dagger$ with $k \neq l$, $\lambda > 0$ and Q such that $P_k Q P_l = Q$. If $||Q||_\infty \leq 1$, then $A \geq 0$ and therefore $E(A) \geq 0$. However, $E(A) = \lambda P_k + \lambda^{-1} P_l + E(Q + Q^\dagger)$ can only be positive for all $\lambda > 0$ and properly normalized Q if $E(Q+Q^\dagger) = 0$. That is, the image of E is independent of 'off-diagonal blocks' in the pre-image so that indeed $E = EE_0$ and therefore

$$E(A) = \sum_{kl,=1}^{K} P_l E\big(P_k A P_k\big) P_l. \tag{1.41}$$

Next, we want to argue that all summands in Eq.(1.41) with $k \neq l$ vanish. For that recall that every matrix can be written as a complex linear combination of Hermitian rank-one projections so that a linear map is determined by its action on the latter. Consider $\psi \otimes \phi_\alpha \in \mathcal{H}_{k,1} \otimes \mathcal{H}_{k,2}$ with $\{\phi_\alpha\}$ any orthonormal basis in $\mathcal{H}_{k,2}$ and note that (i) $E(V_k^\dagger |\psi\rangle\langle\psi| \otimes |\phi_\alpha\rangle\langle\phi_\alpha| V_k) \geq 0$ and (ii) $\sum_\alpha E(V_k^\dagger |\psi\rangle\langle\psi| \otimes |\phi_\alpha\rangle\langle\phi_\alpha| V_k) = V_k^\dagger |\psi\rangle\langle\psi| \otimes \sum_\alpha |\phi_\alpha\rangle\langle\phi_\alpha| V_k$ since $E(b) = b$ for all $b \in \mathcal{B}$. Hence, the image of every summand must be supported within the support space of P_k — in other words $k \neq l$ gives no contribution in Eq.(1.41). So we can finally focus on the action of E restricted to the factors of $\mathcal{B}$ where we refer to it as $E_k : \mathcal{B}(\mathcal{H}_k) \to \mathcal{B}(\mathcal{H}_k)$, i.e.,

$$E(A) = \sum_{k=1}^{K} V_k^\dagger E_k\big(V_k A V_k^\dagger\big) V_k. \tag{1.42}$$

Each E_k has to be a positive linear map satisfying $E_k(X \otimes Y) = X' \otimes \mathbb{1}$ for some $X'(X, Y)$ and $E_k(|\psi\rangle\langle\psi| \otimes \mathbb{1}) = |\psi\rangle\langle\psi| \otimes \mathbb{1}$ for all ψ. Since E_k is positive, and thus order-preserving (see Sec.5.1), this implies that for $Y \leq \mathbb{1}$ we have $E_k(|\psi\rangle\langle\psi| \otimes Y) \leq |\psi\rangle\langle\psi| \otimes \mathbb{1}$ and therefore $X'(X, Y) \propto X$. The proportionality factor has to be a linear, normalized functional of Y such that this together with positivity requires that $E_k(X \otimes Y) = X \otimes \mathbb{1}\mathrm{tr}\,[\rho_k Y]$ for some density operator ρ_k. Inserting into Eq.(1.42) yields Eq.(1.40). □

A map E with the properties stated in Prop.1.5 is called a *conditional expectation*. In other words a conditional expectation is a positive map projecting onto a unital *-subalgebra. Note that for conditional expectations positivity implies complete positivity and that they are self-dual, i.e., $E = E^*$, if $\rho_k \propto \mathbb{1}$ for all k (see below for a more detailed discussion of duality). Moreover, they satisfy

$$\forall b_1, b_2 \in \mathcal{B},\ \forall a \in \mathcal{A} : E(b_1 a b_2) = b_1 E(a) b_2. \tag{1.43}$$

This is often used as the defining property, since every positive linear map $E : \mathcal{A} \to \mathcal{B}$ fulfilling Eq.(1.43) satisfies the requirements of Prop.1.5.

Note that there are two basic building blocks for conditional expectations: (i) partial traces and (ii) *pinchings* (i.e., restrictions to block diagonal matrices).

Duality and (complete) positivity Let us now switch the focus to *duality* and *positivity*. In finite dimensions every linear functional $f : \mathcal{A} \to \mathbb{C}$ can be written as

$$f(A) = \mathrm{tr}\,[FA] \ \text{ for some } F \in \mathcal{A}, \tag{1.44}$$

so that we can identify $\mathcal{A}$ with its dual $\mathcal{A}^*$. If $T : \mathcal{A} \to \mathcal{B}$ is any linear map between two finite dimensional C^*-algebras, then its dual map $T^* : \mathcal{B} \to \mathcal{A}$ is defined via $\mathrm{tr}\,[AT^*(B)] = \mathrm{tr}\,[T(A)B]$ for all A, B. A map T is positive if it maps positive elements of $\mathcal{A}$ to positive elements of $\mathcal{B}$. Recall that positivity (meaning positive semidefiniteness) of $A \in \mathcal{A}$ can be expressed in different equivalent ways: (i) there is an $a \in \mathcal{A}$ such that $A = a^\dagger a$, (ii) for all positive elements $f \in \mathcal{A}^* : f(A) \geq 0$, or (iii) $A^\dagger = A$ and $A + \lambda\mathbb{1}$ is invertible for all $\lambda > 0$ (i.e., A has no negative eigenvalues). A frequently used implication is that $BAB^\dagger \geq 0$ for all $B \in \mathcal{A}$ whenever $A \geq 0$.

Every matrix $X \in \mathcal{M}_d(\mathbb{C})$ can be expressed as a linear combination of two Hermitian matrices via $2X = (X + X^\dagger) - i(iX - iX^\dagger)$ and (by further decomposing each Hermitian term into positive and negative spectral part) as a linear combination of four positive matrices. In fact, the *polarization identity*

$$B^\dagger A = \frac{1}{4}\sum_{k=0}^{3} i^k (A + i^k B)^\dagger (A + i^k B) \tag{1.45}$$

holds for all bounded operators A and B.

To every linear map $T : \mathcal{A} \to \mathcal{B}$ we can assign a sequence of maps $T_n : \mathcal{M}_n(\mathcal{A}) \to \mathcal{M}_n(\mathcal{B})$ defined as $T_n([A_{ij}]) = [T(A_{ij})]$.[27] By identifying $\mathcal{M}_n(\mathcal{A})$ with $\mathcal{A}\otimes\mathcal{M}_n$ this can be written more conveniently as $T_n = (T\otimes\mathrm{id}_n)$. Attributes of T (like positivity, contractivity, boundedness) are then said to be *complete* if they hold for all $T_n, n \in \mathbb{N}$. Putting things together we get that T is completely positive iff

$$\mathrm{tr}\left[b^\dagger b (T \otimes \mathrm{id}_n)(a^\dagger a)\right] = \mathrm{tr}\left[(T^* \otimes \mathrm{id}_n)(b^\dagger b) a^\dagger a\right] \geq 0, \tag{1.46}$$

for all $a \in \mathcal{A} \otimes \mathcal{M}_n$ and $b \in \mathcal{B} \otimes \mathcal{M}_n$. In other words T is completely positive if T^* is. While complete positivity is in general a stronger requirement than positivity, they become the same if either $\mathcal{A}$ or $\mathcal{B}$ is abelian:

Proposition 1.6 (Complete positivity from positivity) *Let $T : \mathcal{A} \to \mathcal{B}$ be positive linear map between unital C^*-algebras. If either $\mathcal{A}$ or $\mathcal{B}$ is commutative then T is completely positive.*

Proof We will prove the finite-dimensional version although the proposition holds in general. W.l.o.g. we can restrict to the case where $\mathcal{B}$ is commutative. The case where the domain is commutative then follows by invoking that T is completely positive iff T^* is. We have to show that $\sum_{ij=1}^{n}\langle\psi_i|T(a_{ij})|\psi_j\rangle \geq 0$ for all $\{|\psi_i\rangle\}$ is implied by $[a_{ij}] \geq 0$. As the $T(a_{ij})$ mutually commute we can choose a basis where they are diagonal and denote the components of $|\psi_j\rangle$ in this basis by $\psi_{j,\beta}$. Then

$$\sum_{ij=1}^{n}\langle\psi_i|T(a_{ij})|\psi_j\rangle = \sum_\beta \langle\beta|T\Big(\underbrace{\sum_{ij}\bar{\psi}_{i,\beta}a_{ij}\psi_{j,\beta}}_{\geq 0}\Big)|\beta\rangle \geq 0, \tag{1.47}$$

[27] With $[A_{ij}]$ we mean the block matrix whose blocks are labeled by i and j.

due to positivity of T and $[a_{ij}]$. □

Sometimes a linear map might only be specified on a subspace $S \subseteq \mathcal{A}$ of a C^*-algebra. If this space is closed under the adjoint and contains the unit element, then it is called an *operator system*. Note for instance that in the proof of Prop.1.6 the algebra structure of $\mathcal{A}$ was never used, so that complete positivity also holds for positive maps which map an operator system to any commutative C^*-algebra. Every such map can then be extended to one with a larger domain (extending at the same time the utility of Prop.1.6):

Proposition 1.7 (Extending cp maps from operator systems) *Let $\mathcal{A}, \mathcal{B}$ be two finite dimensional C^*-algebras and $T : S \to \mathcal{B}$ a completely positive linear map from an operator system $S \subseteq \mathcal{A}$. Then there is a completely positive map $\tilde{T} : \mathcal{A} \to \mathcal{B}$ which coincides with T on S.*

Proof We want to use the Hahn-Banach theorem, so we need a functional: define $\tau(A) := \langle\Omega|(T \otimes \mathrm{id}_n)(A)|\Omega\rangle$ with Ω a maximally entangled state of dimension n which is in turn chosen such that $\mathcal{B}$ is a subalgebra of $\mathcal{M}_n(\mathbb{C})$. So τ is a positive linear functional on $S \otimes \mathcal{M}_n$. Note that complete positivity can be expressed as $\tau\big(\mathbb{1} - A/||A||_\infty\big) \geq 0$ for all $A = A^\dagger$ or equivalently

$$\tau(A) \leq f(A) := ||A||_\infty \tau(\mathbb{1}). \tag{1.48}$$

Since f is sublinear we can now invoke the Hahn-Banach theorem in order to extend τ from the Hermitian subspace of S to that of $\mathcal{A}$ and (by linearity) further to the entire algebra $\mathcal{A}$. The extension fulfills Eq.(1.48) which, by the Choi-Jamiolkowski correspondence (Prop.2.1), implies complete positivity of the extension $\tilde{T}$. By Prop.2.1 $\tilde{T}$ is in one-to-one correspondence with τ. □

1.7 Literature

For a general introduction to quantum mechanics we refer to the excellent text books of Ballentine [?], Galindo and Pascal [?] and Peres [?]. A closer look on quantum mechanics from the point of view of mathematical physics can be found in the books of Thirring [?] and Holevo [?]. A superb introduction to quantum computing and quantum information theory can be found in Nielsen and Chuang [?] the lecture notes by Preskill [?] and the review article of Keyl [?].

Chapter 2

Representations

In this chapter we will discuss the most important ways of parameterizing and representing a quantum channel: (i) in terms of a bipartite quantum state, leading to the state-channel duality introduced by Jamiolkowski and Choi, (ii) as the reduced dynamics of a larger (unitarily evolving) system as expressed by the representation theorems of Kraus, Stinespring, and Neumark (for POVMs) and (iii) as a linear map represented in terms of a 'transfer matrix'. This will be supplemented with a brief discussion about normal forms.

2.1 Jamiolkowski and Choi

We saw in Prop.1.2 that complete positivity of a linear map T is equivalent to positivity of the operator $\tau := (T \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$ which is obtained by letting T act on half of a maximally entangled state Ω. In fact, the operator τ obtained in this way encodes not only complete positivity but every property of T. This reflects the simple equivalence $\mathcal{B}(\mathcal{M}_d, \mathcal{M}_{d'}) \simeq \mathcal{M}_{dd'}$ which was certainly observed and used at various places, but in the context of quantum channels the credit goes to Choi and Jamiolkowski. $d\tau$ is often called *Choi matrix*, and if T is a trace-preserving quantum channel, then τ is called the corresponding *Jamiolkowski state*.

Proposition 2.1 (Choi-Jamiolkowski representation of maps) *The following provides a one-to-one correspondence between linear maps $T \in \mathcal{B}(\mathcal{M}_d, \mathcal{M}_{d'})$ and operators $\tau \in \mathcal{B}(\mathbb{C}^{d'} \otimes \mathbb{C}^d)$:*

$$\tau = (T \otimes \mathrm{id}_d)(|\Omega\rangle\langle\Omega|), \qquad \mathrm{tr}\,[AT(B)] = d\,\mathrm{tr}\left[\tau A \otimes B^T\right], \tag{2.1}$$

for all $A \in \mathcal{M}_{d'}, B \in \mathcal{M}_d$ and $\Omega \in \mathbb{C}^d \otimes \mathbb{C}^d$ being a maximally entangled state as in Eq.(1.11). The maps $T \mapsto \tau$ and $\tau \mapsto T$ defined by (2.1) are mutual inverses and lead to the following correspondences:

- Hermiticity*:* $\tau = \tau^\dagger$ *iff* $T(B^\dagger) = T(B)^\dagger$ *for all* $B \in \mathcal{M}_d$*,*[1]
- Complete positivity*:* T *is completely positive iff* $\tau \geq 0$*,*
- Doubly-stochastic*:* $T(\mathbb{1}) \propto \mathbb{1}$ *and* $T^*(\mathbb{1}) \propto \mathbb{1}$ *iff* $\mathrm{tr}_A[\tau] \propto \mathbb{1}$ *and* $\mathrm{tr}_B[\tau] \propto \mathbb{1}$*.*
- Unitality*:* $T(\mathbb{1}) = \mathbb{1}$ *iff* $\mathrm{tr}_B[\tau] = \mathbb{1}_{d'}/d$*.*
- Preservation of the trace: $\mathrm{tr}_A[\tau] = T^*(\mathbb{1})^T/d$*, i.e.,* $T^*(\mathbb{1}) = \mathbb{1}$ *iff* $\mathrm{tr}_A[\tau] = \mathbb{1}_d/d$*.*
- Normalization*:* $\mathrm{tr}\,[\tau] = \mathrm{tr}\,[T^*(\mathbb{1})]\,/d$*.*

Proof The complete positivity correspondence was proven in Prop.1.2 and the other equivalences follow from Eqs.(2.1) by direct inspection. So, the only piece to prove is that the two relations in (2.1) are mutual inverses which then establishes the claimed one-to-one correspondence. To this end let us start with the rightmost expression in (2.1) and insert $\tau = (T \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$:

$$d\,\mathrm{tr}\left[\tau A \otimes B^T\right] = \mathrm{tr}\left[\mathbb{F}^{T_B} T^*(A) \otimes B^T\right] \tag{2.2}$$

$$= \mathrm{tr}\left[\mathbb{F}\, T^*(A) \otimes B\right] = \mathrm{tr}\left[A T(B)\right], \tag{2.3}$$

where we have used the basic tools from Example 1.2, in particular that $d|\Omega\rangle\langle\Omega| = \mathbb{F}^{T_B}$. Eqs.(2.2,2.3) show that indeed $T \to \tau \to T$. So the two maps in (2.1) are mutual inverses if $T \to \tau$ is surjective. This is, however, easily seen by decomposing τ into a linear combination of rank-one operators $|\psi\rangle\langle\psi'|$ and using $|\psi\rangle = (X \otimes \mathbb{1})|\Omega\rangle$ from Eq.(1.13). □

When writing out the mapping $\tau \to T$ from Eq.(2.1) in terms of a product basis we get

$$T(B) = d \sum_{ijkl} \langle ij|\tau|kl\rangle\ |i\rangle\langle j|\ B\ |l\rangle\langle k|. \tag{2.4}$$

If T is a quantum channel in the Schrödinger picture, then τ is a density operator with reduced density matrix $\tau_B = \mathbb{1}/d$. This means that the set of quantum channels corresponds one-to-one to the set of bipartite quantum state which have one reduced density matrix maximally mixed. By replacing Ω in the construction of τ by any other pure state we can easily establish a similar *state-channel-duality* with respect to states with different reduced density matrices. Such a correspondence will be one-to-one iff the respective reduced state has full rank.

If d is countable infinite then Ω looses its meaning. However, we can partly restore the above correspondence by using either its unnormalized counterpart $\sum_{ij} |ii\rangle\langle jj|$ or any non-maximally entangled state which has reduced density matrix with full rank instead.

The correspondence between T and τ enables us to show that every linear map admits a decomposition into at most four completely positive maps:

[1] Note that this is in turn equivalent to T being a Hermiticity preserving map, i.e., $T(X) = T(X)^\dagger$ for all $X = X^\dagger$.

Proposition 2.2 (Decomposition into completely positive maps) *Every linear map $T \in \mathcal{B}(\mathcal{M}_d, \mathcal{M}_{d'})$ can be written as a complex linear combination of four completely positive maps. If T is Hermitian (i.e., $T(B^\dagger) = T(B)^\dagger$ for all B), it can be written as a real linear combination of two of them.*

Proof We use that the mapping $T \leftrightarrow \tau$ in Prop.2.1 is one-to-one and linear so that it suffices to decompose τ. On this level the existence of such a decomposition is rather obvious: we can always decompose τ into a Hermitian and anti-Hermitian part and each of them into a positive and negative part. More explicitly, $\tau = (\tau+\tau^\dagger)/2 + i(i\tau^\dagger - i\tau)/2$ provides a complex linear combination of two Hermitian parts and each of the latter can, by invoking the spectral decomposition, be written as a real linear combination of two positive semi-definite matrices. □

As a first simple application of this (extremely useful and later extensively used) correspondence we show the following:

Proposition 2.3 (No information without disturbance) *Consider an instrument represented by a set of completely positive maps $\{T_\alpha : \mathcal{M}_d \to \mathcal{M}_d\}$. If there is no disturbance on average, i.e., $T = \sum_\alpha T_\alpha$ satisfies $T = \mathrm{id}$, then $T_\alpha \propto \mathrm{id}$ for every α and the probability for obtaining an outcome α (given by $\mathrm{tr}\,[T_\alpha(\rho)]$) is independent of the input ρ (hence, no information gain).*

Proof On the level of the Jamiolkowski states the decomposition $\mathrm{id} = \sum_\alpha T_\alpha$ reads $|\Omega\rangle\langle\Omega| = \sum_\alpha \tau_\alpha$. Since $\tau_\alpha \geq 0$ (due to complete positivity of T_α) this corresponds to a convex decomposition of the density operator for the maximally entangled state. Since the latter is pure there is only the trivial decomposition, so that $\tau_\alpha = c_\alpha |\Omega\rangle\langle\Omega|$ for some constants $c_\alpha \geq 0$. Consequently, $T_\alpha = c_\alpha \mathrm{id}$ so that $\mathrm{tr}\,[T_\alpha(\rho)] = c_\alpha$ independent of ρ. □

Implementation via teleportation If T is a quantum channel then its Jamiolkowski state τ can operationally be obtained by letting the channel act on a maximally entangled state. What about the converse? If τ is given, does this help us to implement T as an action on an arbitrary input ρ? The answer to this involves some form of *teleportation*: assume that the bipartite state τ is shared by two parties, Alice and Bob, so that Bob has the maximally mixed reduced state. Suppose that Bob has an additional state ρ and that he performs a measurement on his composite system using a POVM which contains the maximally entangled state $\omega := |\Omega\rangle\langle\Omega|$ as an effect operator. We claim now that Alice's state is given by $T(\rho)$ whenever Bob has obtained a measurement outcome corresponding to ω. In order to show this denote Alice's reduce density matrix (in case of Bob's success) by ρ_A, the success probability by p and compute the expectation value with an arbitrary operator A:

$$\begin{aligned} p\,\mathrm{tr}\,[A\rho_A] &= \mathrm{tr}\,[(\tau\otimes\rho)(A\otimes\omega)] && (2.5)\\ &= \mathrm{tr}\left[(\tau\otimes\rho^T)(A\otimes\mathbb{F})\right]/d && (2.6)\\ &= \mathrm{tr}\left[\tau(A\otimes\rho^T)\right]/d \;=\; \mathrm{tr}\,[AT(\rho)]/d^2, && (2.7)\end{aligned}$$

where we used the basic ingredients from Example1.2 again and, in the last step, the r.h.s. of Eq.(2.1). This shows that the described protocol is indeed successful with probability $p = 1/d^2$. In some cases Alice and Bob can, however, do much better: assume that there is a set of local unitaries $\{V_i \otimes U_i\}_{i=1..N}$ w.r.t. which $\tau = (V_i \otimes U_i)\tau(V_i \otimes U_i)^\dagger$ is invariant and which are orthogonal in the sense that $\mathrm{tr}\left[U_i U_j^\dagger\right] = d\delta_{ij}$. Due to the latter condition Bob can use a POVM which contains all $(\mathbb{1} \otimes U_i)^\dagger \omega (\mathbb{1} \otimes U_i)$ as effect operators. When Bob obtains one of the corresponding outcomes i and he reports this to Alice, she can 'undo' the action of the unitary by applying V_i with the consequence that $\rho_A = T(\rho)$ with probability $p = N/d^2$. So if the U_i's form a basis (i.e., $N = d^2$) the two protagonists can implement T with unit probability by 'teleporting' through τ. If $T = \mathrm{id}$ and therefore $\tau = |\Omega\rangle\langle\Omega|$ this reduces to the standard protocol for entanglement-assisted teleportation.

Note that by linearity the implemented action of T does not only hold for uncorrelated states but it also works properly if ρ is part of a larger system.

Problem 6 (Implementation via teleportation) *Suppose Alice and Bob are given the Jamiolkowski state τ corresponding to a given quantum channel T. Which is the largest probability at which Bob can teleport an unknown state ρ to Alice (using τ, local operations and classical communication as resource) so that Alice ends up with the state $T(\rho)$?*

2.2 Kraus, Stinespring and Neumark

The Choi-Jamiolkowski state-channel duality allows us to translate between properties of bibipartite states and quantum channels. One immediate implication is a more specific and very useful representation of quantum channels which corresponds, on the level of the Jamiolkowski state, to a convex (or spectral) decomposition into rank-one operators:

Theorem 2.1 (Kraus representation) *A linear map $T \in \mathcal{B}(\mathcal{M}_d, \mathcal{M}_{d'})$ is completely positive iff it admits a representation of the form*

$$T(A) = \sum_{j=1}^{r} K_j A K_j^\dagger. \tag{2.8}$$

This decomposition has the following properties:

1. *Normalization: T is trace preserving iff $\sum_j K_j^\dagger K_j = \mathbb{1}$ and unital iff $\sum_j K_j K_j^\dagger = \mathbb{1}$.*

2. Kraus rank:[2] *The minimal number of* Kraus operators $\{K_j \in \mathcal{B}(\mathbb{C}^d, \mathbb{C}^{d'})\}_{j=1..r}$ *is $r = \mathrm{rank}(\tau) \leq dd'$.*

[2]We call $r = \mathrm{rank}(\tau)$ *Kraus rank* or *Choi rank* in order not to confuse it with the rank of T as a linear map. Take for instance $T = \mathrm{id}$ the ideal channel. As this is obviously invertible it has full rank as a linear map. However, its Kraus rank is $r = 1$. $T(B) = \mathrm{tr}[B]$ instead has rank one but Kraus rank equal to d, the dimension of the input space.

3. Orthogonality: *There is always a representation with* $r = \text{rank}(\tau)$ *Hilbert-Schmidt orthogonal Kraus operators (i.e.,* $\text{tr}\left[K_i^\dagger K_j\right] \propto \delta_{ij}$*).*

4. Freedom: *Two sets of Kraus operators* $\{K_j\}$ *and* $\{\tilde{K}_l\}$ *represent the same map* T *iff there is a unitary* U *so that* $K_j = \sum_l U_{kl}\tilde{K}_l$ *(where the smaller set is padded with zeros).*

Proof Assume T is completely positive. By Prop.2.1 this is equivalent to saying that $\tau \geq 0$ which allows for a decomposition of the form

$$\tau = \sum_{j=1}^{r} |\psi_j\rangle\langle\psi_j| = \sum_{j=1}^{r}(K_j \otimes \mathbb{1})|\Omega\rangle\langle\Omega|(K_j \otimes \mathbb{1})^\dagger, \tag{2.9}$$

where the first step uses $\tau \geq 0$ and the second Eq.(1.13). Comparing the r.h.s. of Eq.(2.9) with the definition $\tau := (T \otimes \text{id})(|\Omega\rangle\langle\Omega|)$ and recalling that $T \leftrightarrow \tau$ is one-to-one leads to the desired decomposition in (2.8). It also shows that $r \geq \text{rank}(\tau)$ where equality can be achieved and if the ψ_j's are in addition chosen to be orthogonal, then the Kraus operators are orthogonal (w.r.t. the Hilbert-Schmidt inner product) as well.

Conversely, if a map is of the form in (2.8) then $\tau \geq 0$ which implies complete positivity. The conditions for unitality and preservation of the trace are straight forward. It remains to show that the freedom in the representation is precisely given by unitary linear combinations. This is a direct consequence of Eq.(2.9) and the subsequent proposition. □

Proposition 2.4 (Equivalence of ensembles) *Two ensembles of (not necessarily normalized) vectors* $\{\psi_j\}$ *and* $\{\tilde{\psi}_l\}$ *satisfy*

$$\sum_j |\psi_j\rangle\langle\psi_j| = \sum_l |\tilde{\psi}_l\rangle\langle\tilde{\psi}_l| \tag{2.10}$$

iff there is a unitary U *such that* $|\psi_j\rangle = \sum_l U_{jl}|\tilde{\psi}_l\rangle$ *(where the smaller set is padded with zero vectors).*

Proof W.l.o.g. we may think of the mixture in (2.10) as a given density matrix ρ. From both ensembles we can construct a purification of $\rho = \text{tr}_B[|\Psi\rangle\langle\Psi|]$, of the form $|\Psi\rangle = \sum_j |\psi_j\rangle\otimes|j\rangle$ where $\{|j\rangle\}$ is an orthonormal basis which we use as well in $|\tilde{\Psi}\rangle = \sum_l |\tilde{\psi}_l\rangle\otimes|l\rangle$. It follows from the Schmidt decomposition (Prop.1.1) that two purifications differ by a unitary (or isometry if the dimensions do not match), i.e., $|\Psi\rangle = (\mathbb{1}\otimes U)|\tilde{\Psi}\rangle$. By taking the scalar product with a basis vector $\langle j|$ on the second tensor factor this leads to $|\psi_j\rangle = \sum_l U_{jl}|\tilde{\psi}_l\rangle$ which proves necessity of the condition.[3] Sufficiency follows by inspection from unitarity of U. □

[3]If the U relating the two purifications is an isometry it can always be embedded into a unitary, just by completing the set of orthonormal row or column vectors to an orthonormal basis.

Note that the number of linearly independent Kraus operators is $r = \text{rank}(\tau)$ independent of the representation. The Kraus representations of a completely positive map T and its dual T^* are related via interchanging $K_j \leftrightarrow K_j^\dagger$. Moreover, $T^* = T$ iff there is a representation with Hermitian Kraus operators (a simple exercise using the unitary freedom).

Stinespring and the open-system point of view A common perspective is to regard a quantum channel as a description of 'open system dynamics' which originates from considering only parts of a unitarily evolving system. Using the Kraus representation we will now see that indeed every quantum channel can be represented as arising in this way. Before we discuss this in the Schrödinger picture (which might be the more familiar version for physicists), we will state the equivalent theorem in the Heisenberg picture (more common to operator algebraists):

Theorem 2.2 (Stinespring representation) *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a completely positive linear map. Then for every $r \geq \text{rank}(\tau)$ (recall that $\text{rank}(\tau) \leq dd'$) there is a $V : \mathbb{C}^d \to \mathbb{C}^{d'} \otimes \mathbb{C}^r$ such that*

$$T^*(A) = V^\dagger (A \otimes \mathbb{1}_r) V, \qquad \forall A \in \mathcal{M}_{d'}. \tag{2.11}$$

V is an isometry (i.e., $V^\dagger V = \mathbb{1}_d$) iff T is trace preserving.

Proof This is a simple consequence of the Kraus representation $T^*(A) = \sum_{j=1}^r K_j^\dagger A K_j$: construct $V := \sum_j K_j \otimes |j\rangle$ with $\{|j\rangle\}$ any orthonormal basis in $\mathbb{C}^r$. Then Eq.(2.11) holds by construction, r is at least the Kraus rank (hence minimally $r = \text{rank}(\tau)$, see Thm.2.1) and $V^\dagger V = \sum_j K_j^\dagger K_j = \mathbb{1}$ just reflects the trace preserving condition. □

Note from the proof that a representation of the form (2.11) exists for T as well (and not only for T^*). However, the trace-preserving condition is easier expressed in terms of T^* where it becomes unitality.

The ancillary space $\mathbb{C}^r$ is usually called *dilation space*. In the same way in which we constructed V from the set of Kraus operators, we can obtain the latter from V as $K_j = (\mathbb{1}_{d'} \otimes \langle j|)V$. As $r = \text{rank}(\tau)$ is the smallest number of Kraus operators, it is also the least possible dimension for a representation of the form (2.11). Dilations with $r = \text{rank}(\tau)$ are called *minimal*. From the unitary freedom in the choice of Kraus operators (Thm.2.1) we obtain that for minimal dilations V is unique up to the obvious unitary freedom $V \to (\mathbb{1}_{d'} \otimes U)V$. All other dilations can be obtained by an isometry U.

An alternative, but equivalent, way of characterizing minimal dilations is the identity

$$\mathbb{C}^{d'} \otimes \mathbb{C}^r = \{(A \otimes \mathbb{1}_r)V|\psi\rangle \mid A \in \mathcal{M}_{d'},\ \psi \in \mathbb{C}^d\}, \tag{2.12}$$

i.e., the requirement that the set on the r.h.s. spans the entire space.

There is a natural partial order in the set of completely positive maps: we write $T_2 \geq T_1$ iff $T_2 - T_1$ is completely positive. Due to the Choi-Jamiolkowski representation (Prop.2.1) this is equivalent to $\tau_2 \geq \tau_1$. Such an order relation is reflected in the Stinespring representation in the following way:

Theorem 2.3 (Relation between ordered cp-maps) *Let $T_i : \mathcal{M}_{d'} \to \mathcal{M}_d$, $i = 1, 2$ be two completely positive linear maps with $T_2 \geq T_1$. If $V_i : \mathbb{C}^d \to \mathbb{C}^{d'} \otimes \mathbb{C}^{r_i}$ provide Stinespring representations $T_i(A) = V_i^\dagger(A \otimes \mathbb{1}_{r_i})V_i$, then there is a contraction $C : \mathbb{C}^{r_2} \to \mathbb{C}^{r_1}$ such that $V_1 = (\mathbb{1}_{d'} \otimes C)V_2$. If V_2 belongs to a minimal dilation, then C is unique (for given V_1 and V_2).*

Proof We exploit the fact that $T_2 \geq T_1$ is equivalent to $\tau_2 \geq \tau_1$, i.e., the analogous order relation between the corresponding Choi-Jamiolkowski operators. Define $W_i := (\mathbb{1}_{r_i} \otimes \langle\Omega|)(V_i \otimes \mathbb{1}_{d'}) \in \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^{d'}, \mathbb{C}^{r_i})$. Then for all $\psi \in \mathbb{C}^d \otimes \mathbb{C}^{d'}$

$$||W_2|\psi\rangle||^2 = \langle\psi|\tau_2|\psi\rangle \geq \langle\psi|\tau_1|\psi\rangle = ||W_1|\psi\rangle||^2. \tag{2.13}$$

Hence, there is a contraction (meaning $C^\dagger C \leq \mathbb{1}$) $C : \mathbb{C}^{r_2} \to \mathbb{C}^{r_1}$ such that $W_1 = CW_2$. Since the map $V_i \to W_i$ is one-to-one[4] this implies $V_1 = (\mathbb{1}_{d'} \otimes C)V_2$. Moreover, if $r_2 = \mathrm{rank}(\tau_2)$ (i.e., $\mathbb{C}^{r_2}$ is a minimal dilation space), then W_2 must be surjective so that C becomes uniquely defined. □

As a simple corollary of this result one obtains a Radon-Nikodym type theorem for instruments:

Theorem 2.4 (Radon-Nikodym for quantum instruments) *Let $\{T_i\}$ be a set of completely positive linear maps such that $\sum_i T_i = T \in \mathcal{B}(\mathcal{M}_{d'}, \mathcal{M}_d)$ with Stinespring representation $T(A) = V^\dagger(A \otimes \mathbb{1}_r)V$. Then there is a set of positive operators $P_i \in \mathcal{M}_r$ wich sum up to $\mathbb{1}_r$ such that $T_i(A) = V^\dagger(A \otimes P_i)V$.*

Now let us turn to the Schrödinger picture again and see how to represent a quantum channel from a system-plus-environment point of view:

Theorem 2.5 (Open-system representation) *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a completely positive and trace-preserving linear map. Then there is a unitary $U \in \mathcal{M}_{dd'^2}$ and a normalized vector $\varphi \in \mathbb{C}^{d'} \otimes \mathbb{C}^{d'}$ such that*

$$T(\rho) = \mathrm{tr}_E \left[U(\rho \otimes |\varphi\rangle\langle\varphi|)U^\dagger \right], \tag{2.14}$$

where tr_E denotes the partial trace over the first two tensor factors of the involved Hilbert space $\mathbb{C}^d \otimes \mathbb{C}^{d'} \otimes \mathbb{C}^{d'}$.

Proof By expressing Stinespring's representation theorem in the Schrödinger picture we get $T(\rho) = \mathrm{tr}_{\mathbb{C}^r}[V\rho V^\dagger]$. Let us choose a dilation space of dimension $r = dd'$. In this way we can embed V into a unitary which acts on a tensor product and write $V = U(\mathbb{1}_d \otimes |\varphi\rangle)$ for some $\varphi \in \mathbb{C}^{d'} \otimes \mathbb{C}^{d'}$ so that Eq.(2.14) follows. □

Notice the possible departure from the dimension of the minimal dilation space in Thm.2.5: if d is not a factor of $d'\mathrm{rank}(\tau)$, then the construction fails

[4] The inverse is $V_i = d'^2(W_i \otimes \mathbb{1}_{d'})(\mathbb{1}_d \otimes |\Omega\rangle)$.

and we have to use a space of larger dimension (e.g., $r = dd'$ which is always possible). In this case the type of freedom in the representation is less obvious.

The physical interpretation of Thm.2.5 is clear: we couple a system to an environment, which is initially in a state φ, let them evolve jointly according to a unitary U and finally disregard (i.e., trace out) environmental degrees of freedom. This way of representing quantum channels nicely reminds us of the fundamental assumption used when we express an evolution in terms of a linear, completely positive, and trace preserving map: the initial state of the system has to be independent of the 'environment'—in other words T itself must not depend on the input ρ.

Let us finally revisit Thm.2.4 from the system-plus-environment point of view:

Proposition 2.5 (Environment induced instruments) *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a completely positive and trace-preserving linear with a system-plus-environment representation as in Eq.(2.14). For every decomposition of the form $T = \sum_\alpha T_i$ into completely positive maps T_i there is a POVM $\{P_i \in \mathcal{M}_{dd'}\}$ such that*

$$T_i(\rho) = \operatorname{tr}_E\big[(P_i \otimes \mathbb{1}_{d'})U(\rho \otimes |\varphi\rangle\langle\varphi|)U^\dagger\big]. \tag{2.15}$$

The Kraus rank k_i of T_i satisfies $k_i \leq \operatorname{rank}(P_i)$.

Proof .. wouldn't be necessary, as this can essentially be considered a rewriting of Thm.2.4. However, we want to provide a simple alternative proof which is not based on Thm.2.4. Consider the Jamiolkowski state τ of T and its purification $|\psi\rangle := (\mathbb{1}_d \otimes U)(|\Omega\rangle \otimes |\varphi\rangle)$ so that $\tau = \operatorname{tr}_E[|\psi\rangle\langle\psi|]$. The decomposition of T corresponds to a decomposition $\tau = \sum_i \tau_i$ on the level of Choi-Jamiolkowski operators. As this is a convex decomposition of the reduced density matrix of ψ we can invoke the quantum steering result in Prop.1.3 (together with the one-to-one correspondence between T_i and τ_i) in order to arrive at Eq.(2.15). From this the bound on the Kraus rank follows by utilizing $k_i = \operatorname{rank}(\tau_i)$. □

Neumark's representation of POVMs In the same way as Stinespring's theorem enables us to regard a quantum channel as part of larger, unitarily evolving system, one can represent a POVM as a von Neumann measurement performed in an extended space. To simplify matters let us assume that the effect operators all have rank one (which can always be achieved by spectral decomposition).

Theorem 2.6 (Neumark's theorem) [5] *Consider a POVM with effect operators $\{|\psi_i\rangle\langle\psi_i|\}_{i=1..n}$ acting on $\mathbb{C}^d$, i.e., $\sum_i |\psi_i\rangle\langle\psi_i| = \mathbb{1}_d$. There exists an orthonormal basis $\{\phi_i\}_{i=1..n}$ in $\mathbb{C}^n \supseteq \mathbb{C}^d$ such that each ψ_i is the restriction of ϕ_i to $\mathbb{C}^d$.*

[5] As always we restrict ourselves to the case of finite dimensions and finite outcomes. Neumark's theorem holds, however, similarly if the set of measurement outcomes is characterized by any regular, positive, $\mathcal{B}(\mathcal{H})$-valued measure on a compact Hausdorff space. The general theorem can be viewed as a corollary of Stinespring's representation theorem.

PROOF Take an orthonormal basis $\{|j\rangle \in \mathbb{C}^n\}_{j=1..n}$ so that $\mathbb{C}^d$ is embedded as a subspace spanned by the vectors $|1\rangle, \ldots, |d\rangle$. Then the matrix $\Psi \in \mathcal{M}_{n,d}$ with components $\Psi_{i,j} := \langle j|\psi_i\rangle$ satisfies $\Psi^\dagger\Psi = \mathbb{1}_d$. That is, Ψ is an isometry and can be extended to a unitary in $\mathcal{M}_n$ by completing the set of orthonormal column vectors to an orthonormal basis. The vectors ϕ_j are then obtained from the extended $\Psi \in \mathcal{M}_n$ by $\langle j|\phi_i\rangle := \Psi_{i,j}$. □

Note that in order to implement a general POVM with effect operators $\{P_i\}_{i=1..n}$ as von Neumann measurement the construction leading to Thm.2.6 requires a space of dimension $\sum_i \text{rank}(P_i)$. Sometimes we can, however, do better by using rank-one projections which appear in the decomposition of more than one P_i.

Problem 7 (Minimal Neumark dilation) *Which is the minimal dimension required to implement a given POVM as von Neumann measurement in an extended space if post-processing of the measurement outcomes (such as introducing randomness) is allowed?*

Neumark's theorem shows that one can always implement a POVM (at least in principle) as a von Neumann/sharp measurement in a larger space. That is, we embed the system which is, say, described by a density matrix ρ into a larger space where the additional levels are just not populated so that the state becomes a direct sum (i.e., block diagonal) $\rho \oplus 0$ and we obtain the identity

$$\langle\psi_i|\rho|\psi_i\rangle = \langle\phi_i|\rho \oplus 0|\phi_i\rangle. \tag{2.16}$$

For actual practical implementations a tensor product would be more convenient than a direct sum structure, since this would allow a realization by coupling the system to an ancillary system. Such an 'ancilla representation' of a POVM can easily be obtained from Neumark's theorem by further enlarging the space (if necessary). Let us extend the space until the dimension is a multiple of d, say $dd' \geq n$. As the zero-block now has dimension $(d'-1)d$, this enables us to write $\rho \oplus 0 = \rho \otimes |1\rangle\langle 1|$ (with $|1\rangle$ an element of the computational basis in $\mathbb{C}^{d'}$). Embedding the vectors ϕ_i into this space, such that $|\Phi_i\rangle := |\phi_i\rangle \oplus 0$ becomes part of an orthonormal basis in $\mathbb{C}^d \otimes \mathbb{C}^{d'}$, we get

$$\langle\psi_i|\rho|\psi_i\rangle = \langle\Phi_i|(\rho \otimes |1\rangle\langle 1|)|\Phi_i\rangle. \tag{2.17}$$

2.3 Linear maps as matrices

The last representation, which we discuss, is the simplest one—linear maps represented as matrices in a way that concatenation of maps corresponds to multiplication of the respective matrices.

To this end, we use that the space $\mathcal{M}_{d,d'}(\mathbb{C})$ of complex valued $d \times d'$ matrices is a vector space, i.e., it is closed under linear combinations and scalar multiplication. As a vector space $\mathcal{M}_{d,d'}(\mathbb{C})$ is isomorphic to $\mathbb{C}^{dd'}$. We can upgrade it to a Hilbert space by equipping it with a scalar product. A common choice is

$$\langle A, B\rangle := \text{tr}\left[PA^\dagger B\right], \qquad A, B \in \mathcal{M}_{d,d'}, \tag{2.18}$$

where $P \in \mathcal{M}_{d'}$ is any positive definite operator. Each set of dd' operators which is orthonormal w.r.t. this scalar product ($\langle A_i, A_j\rangle = \delta_{ij}, i, j = 1, \ldots, dd'$) forms a basis of $\mathcal{M}_{d,d'}$. Such orthonormal bases lead to simple completely positive maps of the form

$$\sum_{i=1}^{dd'} A_i^\dagger \rho A_i = \mathrm{tr}\,[\rho]\, P^{-1}, \qquad \forall \rho \in \mathcal{M}_d. \tag{2.19}$$

If $P = \mathbb{1}$ in Eq.(2.18) the scalar product is called *Hilbert-Schmidt inner product* and the respective space *Hilbert-Schmidt Hilbert space*. Unless otherwise stated we will in the following use the Hilbert-Schmidt inner product.

The space of linear maps of the form $T : \mathcal{M}_d \rightarrow \mathcal{M}_{d'}$ is isomorphic to $\mathcal{M}_{d'^2,d^2}$. A concrete matrix representation, which we denote by $\hat{T} \in \mathcal{M}_{d'^2,d^2}$ and refer to as *transfer matrix*, is then obtained in given orthonormal bases $\{G_\beta \in \mathcal{M}_d\}_{\beta=1..d^2}$ and $\{F_\alpha \in \mathcal{M}_{d'}\}_{\alpha=1..d'^2}$ via

$$\hat{T}_{\alpha,\beta} := \mathrm{tr}\left[F_\alpha^\dagger T(G_\beta)\right]. \tag{2.20}$$

By construction a concatenation of maps, e.g., $T''(\rho) := T'\big(T(\rho)\big)$, then corresponds to a matrix multiplication $\hat{T}'' = \hat{T}'\hat{T}$. With some abuse of notation we usually write the composed map as a formal product $T'' = T'T$ as well.

From Eq.(2.20) we can see that the matrix representation of T^* is given by the Hermitian conjugate matrix $\hat{T}^\dagger$ if T is a Hermitian map (i.e., $T(X^\dagger) = T(X)^\dagger$) and by the transpose matrix $\hat{T}^T$ if the bases are Hermitian.

Proposition 2.6 (Self-dual channels) *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a completely positive linear map. The following are equivalent:*

1. $T = T^*$,
2. $\hat{T} = \hat{T}^\dagger$ *if identical bases* $F_\alpha = G_\alpha$ *are chosen,*
3. *there exists a set of Hermitian Kraus operators for* T.

Proof $1 \leftrightarrow 2$ follows from direct inspection. Also $3 \rightarrow 1$ is obvious as the Kraus operators of T and T^* differ by Hermitian conjugation. In order to show $1 \rightarrow 3$ we can thus write $T(X) = \sum_j K_j X K_j^\dagger = \frac{1}{2}\sum_j (K_j X K_j^\dagger + K_j^\dagger X K_j)$. Every pair $(K_j, K_j^\dagger)/\sqrt{2}$ of Kraus operators in the last representation can, however, be transformed into a pair of Hermitian operators $(K_j + K_j^\dagger, i(K_j - K_j^\dagger))/2$ by a unitary linear combination. This provides a different representation of the same map (see Prop.2.1, item 4.) by Hermitian Kraus operators. □

Let us in the following have a closer look at different operator bases. The simplest basis which is orthonormal w.r.t. the Hilbert-Schmidt inner product (i.e., $\mathrm{tr}\left[G_\beta^\dagger G_{\beta'}\right] = \delta_{\beta,\beta'}$) is given by *matrix units* of the form $G_\beta = |k\rangle\langle l|$ where β is identified with the index pair $(k, l) =: \beta$ and $k, l = 1, \ldots, d$. Making this identification explicit leads to the isomorphism $\mathcal{M}_d \simeq \mathbb{C}^{d^2}$ via the correspondence $|k\rangle\langle l| \leftrightarrow |k,l\rangle$. When applied to an arbitrary linear map represented as $T(\cdot) = \sum_j L_j \cdot R_j$ this gives

$$\hat{T} = \sum_j L_j \otimes R_j^T. \tag{2.21}$$

That is, in particular for completely positive maps with Kraus representation $T(\cdot) = \sum_j K_j \cdot K_j^\dagger$ we obtain the simple expression $\hat{T} = \sum_j K_j \otimes \bar{K}_j$, which is independent of the particular choice of Kraus operators (as the unitaries, relating the different sets of Kraus operators, cancel). Another advantage of using matrix units as a basis is a simple mapping between $\hat{T}$ and the Choi-Jamiolkowski operator $\tau = (T \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$ for arbitrary linear maps:

$$\hat{T} = d\,\tau^\Gamma, \tag{2.22}$$

where $\tau \mapsto \tau^\Gamma$ is an involution defined by $\langle m,n|\tau^\Gamma|k,l\rangle := \langle m,k|\tau|n,l\rangle$.

More operator bases Despite the useful relations in Eqs.(2.21,2.22) it is sometimes advantageous to use operator bases other than matrix units. The most common ones can be regarded as generalization of the 2×2 Pauli matrices (incl. identity matrix) to higher dimensions. Since the two-dimensional case is rather special one has to drop some of the nice properties—for instance generalize to Hermitian *or* unitary bases.

Problem 8 (Generalizing Pauli matrices) *Construct Hilbert-Schmidt orthogonal bases of operators in $\mathcal{M}_d$ (different from tensor products of Pauli matrices) which are unitary* and *Hermitian.*

Hermitian operator bases: The set of Hermitian matrices forms a real vector space. Thus, an orthonormal basis of Hermitian operators helps to have this property nicely reflected in a concrete representation. A simple example for such a basis can be constructed by embedding normalized Pauli matrices $\sigma_x/\sqrt{2}$ and $\sigma_y/\sqrt{2}$ as 2×2 principal submatrices into $\mathcal{M}_d$ (so that only two entries are non-zero). This leads to d^2-d orthonormal matrices. In order to complete the basis we add d diagonal matrices which we construct from any orthogonal matrix $M \in \mathcal{M}_d(\mathbb{R})$ by choosing the diagonal entries of the k'th diagonal matrix equal to the k'th column vector of M. This yields a complete Hilbert-Schmidt orthonormal basis. If in addition one column of M leads to $\mathbb{1}_d$, then the other d^2-1 matrices are traceless and generate the Lie-algebra $su(d)$, i.e., they provide a complete set of infinitesimal generators of $SU(d)$. A popular example of this kind are the 8 Gell-Mann matrices in $\mathcal{M}_3$ (and, of course, the 3 Pauli matrices in $\mathcal{M}_2$).

Unitary operator bases play an important role in quantum information theory as they correspond one-to-one to bases of maximally entangled states (see example 1.2). In $\mathcal{M}_2$ all orthogonal bases of unitaries are of the form

$$U_k = e^{i\varphi_k} V_1 \sigma_k V_2, \qquad k = 0, \ldots, 3, \tag{2.23}$$

where V_1 and V_2 are unitaries and σ_k denote the Pauli matrices (with $\sigma_0 = \mathbb{1}$). That is, in $\mathcal{M}_2$ all orthogonal bases of unitaries are essentially equivalent to the set of Pauli matrices. In higher dimensions a similar characterization of all unitary operator bases is not known. A priori, it is quite remarkable that such orthogonal bases exist in every dimension. Note that this depends crucially on the chosen scalar product: if we choose the A_i's in Eq.(2.19) proportional to unitaries (and thus $d = d'$), then $\rho = \mathbb{1}$ shows that they can only form an orthogonal basis if $P \propto \mathbb{1}$, corresponding to the Hilbert-Schmidt scalar product.

Example 2.1 (Discrete Weyl system and GHZ bases) *One of the most important unitary operator bases and also a very inspiring example regarding the construction of others comes from a discrete version of the Heisenberg-Weyl group. Consider a set* $\{U_{kl} \in \mathcal{M}_d\}_{k,l=0..d-1}$ *of* d^2 *unitaries defined by*

$$U_{kl} := \sum_{r=0}^{d-1} \eta^{rl} |k + r\rangle\langle r|, \quad \eta := e^{\frac{2\pi i}{d}}, \tag{2.24}$$

where addition inside the ket is modulo d*. This set has the following nice properties:*

- *It forms a basis of operators in* $\mathcal{M}_d$ *which is orthogonal w.r.t. the Hilbert-Schmidt scalar product, i.e.,* $\operatorname{tr}\left[U_{ij}^\dagger U_{kl}\right] = d\ \delta_{ik}\delta_{jl}$*. Since* $U_{00} = \mathbb{1}$ *we have in particular* $\operatorname{tr}[U_{kl}] = 0$ *for all* $(k, l) \neq (0, 0)$*.*
- *It is a* discrete Weyl system *since* $U_{ij}U_{kl} = \eta^{jk} U_{i+k,j+l}$ *(with addition modulo* d*). Thus* $U_{kl}^{-1} = \eta^{kl} U_{-k,-l}$*.*
- *The set is generated by* U_{01} *and* U_{10} *via* $U_{kl} = (U_{10})^k (U_{01})^l$*.*
- *If* d *is odd, then* $U_{kl} \in SU(d)$*. For* d *even* $\det U_{kl} = (-1)^{k+l}$*.*
- *For* $d = 2$ *the set reduces to the set of Pauli matrices with identity, i.e.,* $(\sigma_x, \sigma_y, \sigma_z) = (U_{10}, iU_{11}, U_{01})$*.*
- *The group generated by the unitaries* U_{kl} *is isomorphic to the discrete Heisenberg-Weyl group*

$$\left\{ \begin{pmatrix} 1 & l & m \\ 0 & 1 & k \\ 0 & 0 & 1 \end{pmatrix} \Big|\, k, l, m \in \mathbb{Z}_d \right\}. \tag{2.25}$$

Among the many applications of this set we will briefly discuss the construction of bases of entangled states. As indicated in example 1.2, $(U_{kl} \otimes \mathbb{1})|\Omega\rangle$ *is an orthonormal basis of maximally entangled states in* $\mathbb{C}^d \otimes \mathbb{C}^d$*. For more than two, say* n*, parties a similar construction can be made by exploiting some group structure. Consider the goup of local unitaries*

$$G = \left\{ \bigotimes_{k=1}^{n} U_{i,j_k} \Big| \sum_{k=1}^{n} j_k \bmod d = 0 \right\}, \tag{2.26}$$

where $i \in \mathbb{Z}_d$*,* $j \in \mathbb{Z}_d^n$ *and one of the components of* j*, say* j_1*, depends on the others by the additional constraint. Utilizing the above properties of the set* $\{U_{kl}\}$ *it is readily* *verified that G is an abelian Group with d^n elements and that it spans its own commutant G'. Since the algebra G' is therefore abelian as well, it forms a simplex, i.e., each element in G' has a unique convex decomposition into minimal projections. The latter turn out to be one-dimensional and can be parametrized by $\omega \in \mathbb{Z}_d^n$ such that they correspond to vectors of the form*

$$|\Psi_\omega\rangle := \frac{1}{\sqrt{d}} \sum_{l=0}^{d-1} \eta^{l\omega_1} \bigotimes_{k=1}^{n} |\omega_k + l\rangle. \tag{2.27}$$

These vectors form an orthonormal basis of the entire Hilbert space $\mathbb{C}^{d\otimes n}$ and they can all be obtained by local unitaries from the GHZ state $|\Psi_0\rangle = \sum_k |k\rangle^{\otimes n}/\sqrt{d}$. For $n = d = 2$ we get $G = \{\mathbb{1}, \sigma_x \otimes \sigma_x, -\sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$ (which is the Klein four group[6]*) and the states in Eq.(2.27) become the four* Bell states.

Example 2.1 exhibits two properties which suggest possible starting points for the construction of other unitary operator bases: (i) the group structure of the set, and (ii) the property that every element is generated by two elementary actions: shifting the computational basis and multiplying with phases. The latter approach leads to a fairly general construction scheme. Bases of unitaries obtained in this way are said to be of *shift-and-multiply* type, and they are of the general form

$$U_{kl} = \sum_{r=0}^{d-1} H_{kr}^{(l)} \; |\Lambda_{lr}\rangle\langle r|, \tag{2.28}$$

where $\{H^{(l)} \in \mathcal{M}_d\}$ is a set of (not necessarily distinct) complex Hadamard matrices and $\Lambda \in \mathcal{M}_d(\mathbb{Z}_d)$ is a Latin square. That is, $|H_{kr}^{(l)}| = 1$ are phases so that $H^{(l)}H^{(l)\dagger} = d\mathbb{1}$, and Λ is such that each row and column contains every number from 0 to $d-1$ exactly once. Obviously, the set in Eq.(2.24) is just the simplest example of a shift-and-multiply basis. Its group property (the fact, that the elements form a group of order d^2, up to phases) is not shared by all shift-and-multiply bases. For small dimensions (at least for $d < 6$), every unitary basis with such a group structure (sometimes called *nice error basis*) is of shift-and-multiply type. For higher dimensions ($d = 165$ being the first known counter example) this is no longer valid.

Overcomplete sets, frames and SIC-POVMs We have seen that operator bases for $\mathcal{M}_d$ can be constructed from unitary as well as from Hermitian operators. What about positive semidefinite operators or Hermitian rank-one projections? This turns out not to be possible if one demands orthogonality w.r.t. the Hilbert-Schmidt scalar product (see Prop.2.7). However, relaxing this condition leads to bases with remarkable properties—and our discussion to the framework of *frames*. A frame of a vector space, say $\mathbb{C}^d$, is a set of vectors $\{\phi_i\}_{i=1..n}$ for which there are constants $0 < a \leq b < \infty$ such that for all $\psi \in \mathbb{C}^d$

$$a||\psi||^2 \leq \sum_i |\langle\psi|\phi_i\rangle|^2 \leq b||\psi||^2. \tag{2.29}$$

[6] ok, 'Vierergruppe' sounds better

Hence, the ϕ_i's have to span the entire vector space. If $a = b$, the frame is called *tight* and if in addition $n = d$ it is nothing but an orthonormal basis (with norm equal to $\sqrt{a}$). Since for tight frames $\sum_i |\phi_i\rangle\langle\phi_i| = a\mathbb{1}$, they correspond one-to-one to rank-one POVMs.

A POVM is called *informationally complete* if its measurement outcomes allow for a complete reconstruction of an arbitrary state. The existence of such POVMs is fairly obvious: just use properly normalized spectral projections of a unitary or Hermitian operator basis as effect operators. A simple physical implementation in $\mathbb{C}^d$ would be to first draw a random number from $1, \ldots, d^2$ and then measure the corresponding element of a Hermitian operator basis. Clearly, a POVM acting on $\mathbb{C}^d$ is informationally complete iff it contains d^2 linearly independent effect operators.

Now let us try to construct a positive semidefinite operator basis for $\mathcal{M}_d$ which is 'as orthogonal as possible' and see how this relates to tight frames and informationally complete POVMs:

Proposition 2.7 (SIC POVMs) *Let $\{P_i\}_{i=1..n}$ be a set of positive semidefinite operators in $\mathcal{M}_d$, $d \leq n$, which are normalized w.r.t. the Hilbert-Schmidt scalar product, i.e., $\langle P_i, P_i\rangle = \operatorname{tr}\left[P_i^2\right] = 1$. Then*

$$\sum_{i\neq j} |\langle P_i, P_j\rangle|^2 \geq \frac{(n-d)^2 n}{(n-1)\, d^2}, \tag{2.30}$$

with equality iff all P_i are rank-one projections fulfilling $\langle P_i, P_j\rangle = \frac{n-d}{(n-1)d}$ for all $i \neq j$, and $\sum_i P_i = \frac{n}{d}\mathbb{1}$. In case of equality the P_i's are linearly independent.

Proof Define $Q := \sum_i P_i$. Invoking Cauchy-Schwarz inequality twice we get

$$\sum_{i\neq j} |\langle P_i, P_j\rangle|^2 \;\geq\; \frac{1}{n^2-n}\Big(\sum_{i\neq j}\langle P_i, P_j\rangle\Big)^2 \tag{2.31}$$

$$= \frac{1}{n^2-n}\left(\operatorname{tr}\left[Q^2\right] - n\right)^2 \geq \frac{\left(\operatorname{tr}\left[Q\right]^2/d - n\right)^2}{n^2-n}, \tag{2.32}$$

where the first inequality holds iff $\langle P_i, P_j\rangle = const$ for all $i \neq j$ and the second one iff $Q \propto \mathbb{1}$. Since due to normalization $\operatorname{tr}\left[P_i\right] \geq 1$ and thus $\operatorname{tr}\left[Q\right] \geq n$ with equality iff P_i is a rank-one projection we can further bound (2.32) from below leading to the r.h.s. of (2.30). Collecting the conditions for equality completes the proof of the first part. In order to show linear independence assume $\sum_j c_j P_j = 0$ and set $c := \langle P_i, P_j\rangle$ for $i \neq j$. Then for all i: $0 = \sum_j c_j\langle P_i, P_j\rangle = c_i(1-c) + c\sum_j c_j$, so c_i is independent of i and therefore $c_i = 0$. □

As claimed before, Prop.2.7 shows that positive semidefinite operators cannot form a Hilbert-Schmidt orthogonal operator basis (for that the l.h.s. of Eq.(2.30) would have to be zero, which is in conflict with $n = d^2$). Whether the conditions for equality in Prop.2.7 are vacuous or actually have a solution depends on n and d. Whenever there exists a solution, it forms a tight frame.

While for $n = d$ the solutions are exactly all orthonormal bases of $\mathbb{C}^d$, for $n > d^2$ no solution exists (as the P_i's become necessarily linear dependent then).

Remarkably, there exist solutions for some $n = d^2$, i.e., tight frames which are informationally complete POVMs with the additional symmetry $\mathrm{tr}\,[P_i P_j] = (d+1)^{-1}$ for all $i \neq j$. These are called *symmetric informationally complete* (SIC) POVMs. For $d = 2$ all SIC POVMs are obtained by choosing points on the Bloch sphere which correspond to vertices of a tetrahedron. For $d = 3$ an example can be constructed from the shift-and-multiply basis in Eq.(2.24) by taking projections onto the vectors $|\xi_{kl}\rangle := U_{kl}|\xi_{00}\rangle$ with $|\xi_{00}\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$. In other words, in this example the vectors onto which the P_i's project are *generalized coherent states* w.r.t. the discrete Heisenberg-Weyl group. Analogous to continuous coherent states, all SIC POVMs provide (due to the operator basis property) a representation of arbitrary operators $\rho \in \mathcal{M}_d$:

$$\rho = \frac{1}{d} \sum_i \Big((d+1)\mathrm{tr}\,[P_i \rho] - \mathrm{tr}\,[\rho] \Big)\, P_i. \tag{2.33}$$

In the context of coherent states this is called *diagonal representation* or (in quantum optics for the continuous Heisenberg-Weyl group) *P-function* representation.

Using the effect operators of a SIC-POVM as Kraus operators we get a quantum channel of the form

$$\frac{1}{d} \sum_{i=1}^{d^2} P_i \rho P_i = \frac{\mathbb{1}\mathrm{tr}\,[\rho] + \rho}{d+1}, \qquad \forall \rho \in \mathcal{M}_d. \tag{2.34}$$

Both Eq.(2.33) and Eq.(2.34) are readily verified by using the properties (in particular the operator basis property) of the P_i's.

Problem 9 (SIC POVMs) *Determine the pairs of (n, d) for which equality in Eq.(2.30) can be achieved.*

2.4 Normal forms

It is sometimes useful to cut some trees in order to get a better overview of the forest. Depending on the situation will later encounter various 'normal forms' for the representations discussed previously—in particular, decompositions based on spectral, convex or semi-group properties. This section is devoted to normal forms of completely positive maps w.r.t. invertible maps with Kraus rank one. More specifically, given a completely positive map $T : \mathcal{M}_{d_1} \to \mathcal{M}_{d_2}$ we are interested in in simple representatives of the equivalence class

$$T \sim \Phi_2 T \Phi_1, \tag{2.35}$$

where $\Phi_i \in \mathcal{B}(\mathcal{M}_{d_i})$ is of the form $\Phi_i(\cdot) := X_i \cdot X_i^\dagger$ with $X_i \in SL(d_i, \mathbb{C})$, i.e., complex valued matrices with unit determinant. Such maps often run under the name *filtering operations*. On the level of Choi-Jamiolkowski operators $\tau \in \mathcal{B}(\mathbb{C}^{d_2} \otimes \mathbb{C}^{d_1})$ the transformation in (2.35) corresponds to $\tau \mapsto \tau' := (X_2 \otimes X_1)\tau(X_2 \otimes X_1)^\dagger$ (up to an irrelevant transposition for X_1). In order to construct a normal form w.r.t. such transformations we utilize the following optimization problem:

$$p := \inf_{X_i \in SL(d_i,\mathbb{C})} \operatorname{tr}[\tau']. \tag{2.36}$$

Recall that we assume complete positivity which by Prop.2.1 means $\tau \geq 0$. Clearly, $p \leq \operatorname{tr}[\tau]$ (as we can choose $X_i = \mathbb{1}$) and $p \geq ||X_i||_\infty^2 \lambda_{\min}(\tau)$ with $\lambda_{\min}(\tau)$ the smallest eigenvalue of τ. Our aim is to choose as a representative of the equivalence class one which minimizes (2.36):

Proposition 2.8 (Normal form for generic τ) *Let $\tau \in \mathcal{B}(\mathbb{C}^{d_2} \otimes \mathbb{C}^{d_1})$ be positive definite. Then there exist $X_i \in SL(d_i,\mathbb{C})$ which attain the infimum in (2.36) so that the respective optimal $\tau' = (X_2 \otimes X_1)\tau(X_2 \otimes X_1)^\dagger$ is such that both partial traces are proportional to the identity matrix.*

Proof From the upper and lower bound on p we know that we can w.l.o.g. restrict to $||X_i||_\infty^2 \leq \operatorname{tr}[\tau]/\lambda_{\min}(\tau)$. These X_i form a compact set (since $\tau > 0$ by assumption) so that the infimum is indeed attained.

Eq.(2.36) can now be minimized by a simple iteration in X_1 and X_2: denote by $\tau_1 = \operatorname{tr}_2[\tau]$ a partial trace of τ so that $\tau_1' := X_1 \tau_1 X_1^\dagger$ is the corresponding one for $\tau' := (\mathbb{1} \otimes X_1)\tau(\mathbb{1} \otimes X_1)^\dagger$. Choosing $X_1 := \det(\tau_1)^{1/2d_1}\tau_1^{-1/2}$ will lead to $\tau_1' \propto \mathbb{1}$. Moreover, the arithmetic-geometric mean inequality gives

$$\operatorname{tr}[\tau'] = \det(\tau_1)^{1/d_1} d_1 \leq \operatorname{tr}[\tau_1] = \operatorname{tr}[\tau], \tag{2.37}$$

with equality iff $\tau_1 \propto \mathbb{1}$. Iterating this step w.r.t. X_1 and X_2 will thus decrease the trace while both reduced density matrices converge to something proportional to the identity. As this holds true in particular for the optimal τ', it has to have $\tau_i' \propto \mathbb{1}$. □

Together with the Choi-Jamiolkowski correspondence in Prop.2.1 this has an immediate corollary on the level of completely positive maps:

Proposition 2.9 (Normal form for generic cp maps) *Let $T : \mathcal{M}_{d_1} \to \mathcal{M}_{d_2}$ be a completely positive map with full Kraus rank. There exist invertible completely positive maps $\Phi_i \in \mathcal{B}(\mathcal{M}_{d_i})$ with Kraus rank one such that $T' := \Phi_2 T \Phi_1$ is* doubly-stochastic, *i.e., both $T(\mathbb{1})$ and $T^*(\mathbb{1})$ are proportional to the identity matrix.*

The normal form in Prop.2.8 is unique[7] up to local unitaries $\tau' \mapsto (U_2 \otimes U_1)\tau'(U_2 \otimes U_1)$, which corresponds to unitary channels $\Phi_i(\cdot) = U_i \cdot U_i^\dagger$ in Prop.2.9.

[7] This is an exercise about non-negative matrices: assume $\tilde{\tau} := (D_2 \otimes D_1)\tau'(D_2 \otimes D_1)$ is a second normal form with D_i positive diagonal matrices (we can always use the unitary freedom to choose them like this). Then $M_{ij} := \langle ij|\tau'|ij\rangle$ (and analogously $\tilde{M}$) is a doubly stochastic non-negative matrix, i.e., $M_{ij} \geq 0$ and all rows as well as well as all columns have the same sum. Moreover, $\tilde{M} = M * H$ is a Hadamard product with $H_{ij} := [D_2]_{ii}[D_1]_{jj}$. The only H which achieves such a mapping is, however, a multiple of $H_{ij} = 1$.

This implies that the algorithmic procedure used in the proof of Prop.2.8 in order to obtain the normal form by minimizing $\mathrm{tr}\,[\tau']$ eventually converges to the global optimum.

In cases where τ has a kernel the infimum in (2.36) may either be zero or not be attained for finite X_i. An exhaustive investigation of general normal forms w.r.t. the equivalence class (2.35) has been performed for qubit channels.

Qubit maps We continue the discussion about normal forms w.r.t. Kraus-rank one operations for completely positive maps $T : \mathcal{M}_2 \rightarrow \mathcal{M}_2$. Choosing normalized Pauli matrices as operator basis (with $\sigma_0 = \mathbb{1}$) we can represent T as a 4×4 matrix $\hat{T}_{ij} := \mathrm{tr}\,[\sigma_i T(\sigma_j)]/2$. If T is Hermitian, then $\hat{T}$ is real, and if T is trace preserving, then $(T_{1,j}) = (1, 0, 0, 0)$. A quantum channel in the Schrödinger picture is thus represented by

$$\hat{T} = \begin{pmatrix} 1 & 0 \\ v & \Delta \end{pmatrix}, \tag{2.38}$$

where Δ is a real 3×3 matrix and $v \in \mathbb{R}^3$. The corresponding Jamiolkowski state is given by $\tau = \frac{1}{4}\sum_{ij} \hat{T}_{ij}\sigma_i \otimes \sigma_j^T$ so that v describes its reduced density operator and Δ its correlations. The conditions for $\hat{T}$ to correspond to a completely positive map have to be read off from $\tau \geq 0$. The complexity of this condition is the drawback of this representation. One of its advantages is a nice geometric interpretation: parameterizing a density operator via $\rho = (\mathbb{1} + \sum_{k=1}^{3} x_k\sigma_k)/2$, i.e., in terms of a vector $x \in \mathbb{R}^3$ within the Bloch ball $||x|| \leq 1$ (see example1.1), the action of T is a simple affine transformation

$$x \mapsto v + \Delta x. \tag{2.39}$$

From here conditions for T to be positive are readily derived (as the vector has to stay within the Bloch ball), and we will discuss them in greater detail in ... . The following is a useful proposition in this context:

Proposition 2.10 *Let $T : \mathcal{M}_2(\mathbb{C}) \rightarrow \mathcal{M}_2(\mathbb{C})$ be a Hermiticity preserving linear map with $\Delta_{ij} := \mathrm{tr}\,[\sigma_i T(\sigma_j)]/2$ the lower-right submatrix (i.e., $i, j = 1, 2, 3$) of the matrix representation $\hat{T}$. Then $\hat{T}' := \hat{T}_{00} \oplus \Delta$ represents a (completely) positive map iff $\hat{T}$ does so.*

Proof We use that $D := \mathrm{diag}(1, -1, -1, -1)$ is the matrix representation of time-reversal (see Sec.3.3), i.e., matrix transposition in some basis. Following the relation in Eq.(1.24) the map $D\hat{T}D$ is (completely) positive if $\hat{T}$ is. The same holds thus for the convex combination $(\hat{T} + D\hat{T}D)/2 = \hat{T}'$. □

Suppose now that we act with a unitary before and another one after applying the channel T so that the overall action is $\rho \mapsto U_2 T(U_1 \rho U_1^\dagger)U_2^\dagger$. The transfer matrix corresponding to this concatenation is then given by the product $(1 \oplus O_2)\hat{T}(1 \oplus O_1)$, where $O_i \in SO(3)$ are real rotations of the Bloch sphere. This reflects the two-to-one group homomorphism $SU(2) \rightarrow SO(3)$ (see examples 1.1 and 2.2). We may use this in order to diagonalize $\Delta \rightarrow \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$

so that $\lambda_1 \geq \lambda_2 \geq |\lambda_3|$ and for positive, trace-preserving maps necessarily $1 \geq \lambda_1$. Up to a possibly remaining sign this diagonalization is nothing but a real singular value decomposition. Expressed in terms of the λ_i's a necessary condition for complete positivity is that

$$\lambda_1 + \lambda_2 \leq 1 + \lambda_3. \tag{2.40}$$

This becomes sufficient if the map is unital, i.e., if $v = 0$ (see also Exp.2.3).

Extending the freedom in the transformations from $SU(2)$ to $SL(2)$ enables us to further simplify $\hat{T}$ and to bring it to a normal form (as stated in Prop.2.9 for the full rank case):

Proposition 2.11 (Lorentz normal form) *For every qubit channel T there exist invertible completely positive maps Φ_1, Φ_2, both of Kraus rank one, such that the concatenation $\Phi_2 T \Phi_1 =: T'$ (characterized by v and Δ) is a qubit channel of one of the following three forms:*

1. Diagonal: *T' is unital ($v = 0$) with Δ diagonal. This is the generic case (proved in Prop.2.9).*
2. Non-diagonal: *T' has $\Delta = \mathrm{diag}(x/\sqrt{3}, x/\sqrt{3}, 1/3)$, $0 \leq x \leq 1$ and $v = (0, 0, 2/3)$. These channels have Kraus rank 3 for $x < 1$ and Kraus rank 2 for $x = 1$.*
3. Singular: *T' has $\Delta = 0$ and $v = (0, 0, 1)$. This channel has Kraus rank 2 and is singular in the sense that it maps everything onto the same output.*

Example 2.2 (Lorentz group and spinor representation) *Our aim is to understand the action of $T \rightarrow \Phi_2 T \Phi_1$ as an equivalence transformation on $\hat{T}$ by elements from the Lorentz group. In order to see how this arises, consider the space $\mathcal{M}_2^\dagger(\mathbb{C})$ of complex Hermitian 2×2 matrices. Every such matrix can be expanded as $M = \sum_{i=0}^3 x_i \sigma_i$ with $x \in \mathbb{R}^4$. Since $\det(M) = x_0^2 - x_1^2 - x_2^2 - x_3^2 = \langle x, \eta x\rangle$ with $\eta := \mathrm{diag}(1, -1, -1, -1)$ we can identify $\mathcal{M}_2^\dagger(\mathbb{C})$ with* Minkowski space *such that the determinant provides the Minkowski metric. If $X \in SL(2, \mathbb{C})$, the map*

$$M \mapsto XMX^\dagger \tag{2.41}$$

in this way becomes a linear isometry in Minkowski space, i.e., a Lorentz transformation. *In fact, the group $SL(2, \mathbb{C})$ is a double cover of the special orthochronous Lorentz group*

$$SO^+(1,3) := \{L \in \mathcal{M}_4(\mathbb{R}) \mid \det(L) = 1,\ L\eta L^T = \eta,\ L_{00} > 0\} \tag{2.42}$$

in very much the same way as $SU(2)$ is a double cover of $SO(3)$. The map $SL(2, \mathbb{C}) \rightarrow SO^+(1,3)$ constructed above is sometimes called spinor map. *It is two-to-one since $\pm X$ have the same effect in (2.41). Due to this equivalence the transfer matrix of the channel $\Phi_2 T \Phi_1$ becomes*

$$L_2 \hat{T} L_1, \quad \textit{with} \quad L_i \in SO^+(1,3). \tag{2.43}$$

The normal form in Prop.2.11 can thus be regarded as a normal form w.r.t. special ortochronous Lorentz transformations.

Every $L \in SO^+(1,3)$ can be decomposed into a 'boost' and a spatial rotation. This decomposition can be obtained from the polar decomposition $X = PU$ *in* $SL(2,\mathbb{C})$, *where* $P > 0$ *and* $UU^\dagger = \mathbb{1}$. *In order to make this more explicit, define generators of spatial rotations by* $R_i := \sum_{jk=1}^3 \epsilon_{ijk}|k\rangle\langle j| \in \mathcal{M}_4$, *and generators of boosts by* $B_i := |0\rangle\langle i| + |i\rangle\langle 0| \in \mathcal{M}_4$ *with* $i = 1,2,3$. *Then with* $\vec{n}, \vec{m} \in \mathbb{R}^3$ *the mapping* $SL(2,\mathbb{C}) \to SO^+(1,3)$ *takes the form*

$$U = e^{-i\vec{n}\cdot\vec{\sigma}/2} \to e^{-\vec{n}\cdot\vec{R}}, \qquad P = e^{\vec{m}\cdot\vec{\sigma}/2} \to e^{\vec{m}\cdot\vec{B}}. \tag{2.44}$$

Decomposing P *further according to the spectral decomposition leads to a diagonal matrix which corresponds to a boost in* z*-direction (i.e., it acts in the 0-3–plane as* $\mathbb{1}\cosh m_3 + \sigma_x \sinh m_3$*).*

Example 2.3 (Bell diagonal states and Pauli channels) ...

2.5 Literature

Chapter 3

Positive, but not completely

In this chapter we will discuss linear maps which are positive (i.e., they preserve positive semidefiniteness) but not necessarily completely positive. Although such operations cannot directly be implemented in a physical device they play an important role as mathematical tools, in particular in entanglement theory (see Sec.3.2).

Recall that a linear map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ is *positive* if for every positive semidefinite matrix $A \in \mathcal{M}_d(\mathbb{C})$ (we write $A \geq 0$) we have $T(A) \geq 0$. T is positive iff the adjoint map T^* is so, and every positive map is Hermitian in the sense that $T(X^\dagger) = T(X)^\dagger$ for all $X \in \mathcal{M}_d(\mathbb{C})$. By Prop.2.2 every positive map is a difference of two completely positive maps. Since this is, however, true for every Hermitian map it does, unfortunately, not provide a simple parametrization of positive maps.

The paradigm of a map, which is positive but not completely positive, is the usual matrix transposition $A \mapsto \theta(A) := A^T$ which is defined with respect to a given orthogonal basis by $\langle i|A^T|j\rangle = \langle j|A|i\rangle$. The transposition is trace preserving and positive: if $UDU^\dagger$ is the spectral decomposition of any Hermitian matrix (i.e., U is unitary and D diagonal), then $(UDU^\dagger)^T = \bar{U}DU^T$ has the same eigenvalues. The fact that transposition is not completely positive can be easily seen by applying it to half of a maximally entangled state $|\Omega\rangle = \sum_{i=1}^d |ii\rangle/\sqrt{d}$ (see example 1.2):

$$\big(\theta \otimes \mathrm{id}\big)\,\big(|\Omega\rangle\langle\Omega|\big) = \frac{1}{d}\,\mathbb{F}\,, \qquad \mathbb{F} = \sum_{i,j=1}^{d} |ij\rangle\langle ji|\,. \tag{3.1}$$

Since the "flip operator" $\mathbb{F}$ has eigenvalues ± 1 (with multiplicities $d(d \pm 1)/2$ corresponding to symmetric and antisymmetric eigenvectors respectively) θ fails to be completely positive. For the *partial transpose* $(\theta \otimes \mathrm{id})(X)$ we will occasionally write X^{T_A} or X^{T_1} referring to a transposition on Alice's/the first tensor factor.

Transposition plays an important role in the discussion of the structure of positive maps and isometries. We came across one of its appearances already in Wigner's theorem1.1 and we will see similar relations in Prop,3.6 and ... . Among all positive maps transposition gets physically distinguished by a simple interpretation—as time reversal operation (see Sec.3.3).

The simplest class of positive maps are those which are build up of a matrix transposition together with completely positive maps T_1 and T_2. A positive map which has such a decomposition in the form[1]

$$T = T_1 + T_2\theta \tag{3.2}$$

is thus called *decomposable*. Note that the order of maps is irrelevant here, since $\theta T_2 \theta$ is completely positive iff T_2 is. From $\mathcal{M}_d$ to $\mathcal{M}_d$ with $d = 2$ every positive map is decomposable, whereas for larger dimensions this is no longer true. Both facts have far-ranging consequences in entanglement theory (see Sec.3.2). The existence of *indecomposable* positive maps can, with some imagination and the willingness to make a little detour, be traced back to the history of Hilbert's 17'th problem. Since this is a detour of the nicer kind, we will follow it in Sec.3.4. We will start, however, by discussing notions of positivity in between bare positivity and complete positivity:

3.1 n-positivity

A linear map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ is called *n-positive* (with $n \in \mathbb{N}$) if $T \otimes \mathrm{id}_n$ is positive. That is, positivity of T is preserved if we add an 'innocent bystander' of dimension at most n. For given dimensions, the set $\mathcal{T}_n$ of n-positive maps forms a closed convex cone, i.e., if $T_i \in \mathcal{T}_n$ then $\sum_i p_i T_i \in \mathcal{T}_n$ for all positive $p_i \in \mathbb{R}$ and $\mathcal{T}_n$ contains its limit points. Moreover, the dual map T^* is n-positive if and only if T is. Clearly, n-positivity implies m-positivity for all $m \leq n$. Only if $n = d$, then (by Prop.1.2) m-positivity holds for all $m \in \mathbb{N}$, i.e., the map is completely positive. For different n we thus get a chain of inclusions

$$\mathcal{T}_{cp} := \mathcal{T}_d \subseteq \mathcal{T}_{d-1} \subseteq \cdots \subseteq \mathcal{T}_2 \subseteq \mathcal{T}_1, \tag{3.3}$$

from completely positive maps ($n = d$) to positive maps ($n = 1$). As we will see later all these inclusions are strict. To this end the following characterization is useful:

Proposition 3.1 (Choi-Jamiolkowski for n-positive maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ be a linear map and $\tau := (T \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|) \in \mathcal{M}_{d'} \otimes \mathcal{M}_d$ its corresponding Choi-Jamiolkowski operator. For $n \leq d$ the following are equivalent:*

1. *T is n-positive.*

2. *$(\mathbb{1} \otimes P)\tau(\mathbb{1} \otimes P) \geq 0$ for all Hermitian projections $P \in \mathcal{M}_d(\mathbb{C})$ with $\mathrm{rank}(P) = n$.*

3. *$\langle\psi|\tau|\psi\rangle \geq 0$ for all $\psi \in \mathbb{C}^{d'} \otimes \mathbb{C}^d$ with Schmidt-rank n.*

[1] Here $T_2\theta$ is a concatenation of maps, i.e., $(T_2\theta)(X) = T_2(X^T)$.

PROOF T is n-positive iff $(T \otimes \mathrm{id}_n)(|\phi\rangle\langle\phi|) \geq 0$ for all $\phi \in \mathbb{C}^d \otimes \mathbb{C}^n$. Embedding $\mathbb{C}^n$ in $\mathbb{C}^d$ this becomes equivalent to $(T \otimes \mathrm{id}_d)(|\phi\rangle\langle\phi|) \geq 0$ for all $\phi \in \mathbb{C}^d \otimes \mathbb{C}^d$ with Schmidt rank n. The latter can be parameterized by $|\phi\rangle = (\mathbb{1} \otimes X)|\Omega\rangle$ where $X \in \mathcal{M}_d$ has rank n. In this way n-positivity becomes equivalent to

$$(\mathbb{1} \otimes X)\tau(\mathbb{1} \otimes X)^\dagger \geq 0, \quad \forall X \in \mathcal{M}_d(\mathbb{C}) : \mathrm{rank}(X) = n. \tag{3.4}$$

From here equivalence with 2. follows by expressing $X = \tilde{X}P$ in terms of the projector P onto its support space and an invertible $\tilde{X}$. Equivalence between 3. and Eq.(3.4) is seen by taking the expectation value of the l.h.s. of (3.4) with an arbitrary vector $|\tilde{\psi}\rangle$ and noting that $(\mathbb{1} \otimes X)^\dagger|\tilde{\psi}\rangle =: |\psi\rangle$ has Schmidt rank n. □

In particular the last item motivates to investigate extremal overlaps with pure states of a given Schmidt rank:

Lemma 3.1 (Overlap with fixed Schmidt rank) *Let $\phi \in \mathbb{C}^{d'} \otimes \mathbb{C}^d$ be a normalized vector with reduced density matrix $\mathrm{tr}_2|\phi\rangle\langle\phi| =: \rho$. Then the maximal overlap with all normalized vectors ψ of Schmidt rank n is given by*

$$\sup_\psi |\langle\phi|\psi\rangle|^2 = ||\rho||_{(n)}, \tag{3.5}$$

where $||\cdot||_{(n)}$ denotes the Ky-Fan norm, i.e., the sum over the n largest singular values (here, eigenvalues).

PROOF Using the Schmidt decomposition of both vectors we can rewrite the optimization in the form

$$\sup_\psi |\langle\phi|\psi\rangle|^2 = \sup_{U,V,\mu} \left|\mathrm{tr}\left[U\sqrt{\mu}V\sqrt{\lambda}\right]\right|^2, \tag{3.6}$$

where U, V are unitaries, λ a diagonal matrix containing the eigenvalues of ρ and $\sqrt{\mu}$ a diagonal matrix containing the Schmidt coefficients of ψ. The maximum in Eq.(3.6) is attained for $U = V = \mathbb{1}$ (which can for instance be seen by exploiting a standard singular value inequality[2]). The remaining supremum is then of the form $\sup_\mu |\sum_{i=1}^n \sqrt{\mu_i\lambda_i}|^2$ with $\{\lambda_1, \ldots, \lambda_n\}$ the n largest eigenvalues of ρ. This can easily be solved using a Lagrange multiplier for the normalization constraint $\sum_{i=1}^n \mu_i = 1$ which gives $\mu_i = c\lambda_i$ with c determined by normalization. Putting things together we arrive at (3.5). □

Utilizing this Lemma together with item 3. in Prop.3.1 we can now provide a simple criterion for n-positivity in terms of the spectral decomposition $\tau = \sum_i \nu_i|\phi_i\rangle\langle\phi_i|$ (with $||\phi_i|| = 1$). Denoting by ρ_i the reduced density operator of ϕ_i we obtain:

[2]For two matrices A, B the ordered singular values satisfy $\sum_i s_i(AB) \leq \sum_i s_i(A)s_i(B)$; see Thm. IV.2.5 in ... [Bhatia] .

Proposition 3.2 (Criterion for n-positivity) *Consider a Hermitian operator $\tau \in \mathcal{M}_{d'} \otimes \mathcal{M}_d$ (e.g., the Choi-Jamiolkowski operator of a Hermitian map) whose smallest positive non-zero eigenvalue is ν_0. Then with the above notation*

$$\inf_{\psi} \langle\psi|\tau|\psi\rangle \;\geq\; \nu_0 + \sum_{i:\nu_i \leq 0} (\nu_i - \nu_0)||\rho_i||_{(n)}, \tag{3.7}$$

if the infimum is taken over all normalized vectors of Schmidt rank n. Conversely, if all but one eigenvalue ν_- of τ are strictly positive, and ν_∞ is the largest positive eigenvalue, then:

$$\inf_{\psi} \langle\psi|\tau|\psi\rangle \leq \nu_\infty + (\nu_- - \nu_\infty)||\rho_-||_{(n)}. \tag{3.8}$$

Proof By separating positive and negative parts of τ we obtain

$$\tau \;\; \geq \;\; \nu_0 \sum_{i:\nu_i > 0} |\phi_i\rangle\langle\phi_i| \; + \sum_{j:\nu_j \leq 0} \nu_j |\phi_j\rangle\langle\phi_j|, \tag{3.9}$$

$$\;\; = \;\; \nu_0 \mathbb{1} + \sum_{i:\nu_i \leq 0} (\nu_i - \nu_0)|\phi_i\rangle\langle\phi_i|, \tag{3.10}$$

from which Eq.(3.7) follows by exploiting Lemma 3.1. The inequality (3.9) is reversed if we replace ν_0 by ν_∞, which then leads to Eq.(3.8). □

Now, let us apply this to the map $T_\eta : \mathcal{M}_d \to \mathcal{M}_d$ defined by

$$T_\eta(\rho) = \mathbb{1}_d \mathrm{tr}\,[\rho] - \frac{\rho}{\eta}, \qquad \eta \in \mathbb{R}^+. \tag{3.11}$$

For $\eta \geq d$ this map is completely positive. For $\eta < d$ all positive eigenvalues of the corresponding Choi-Jamiolkowski operator are equal to $1/d$ and there is a single negative eigenvalue $\nu_- = (1/d - 1/\eta)$ with $||\rho_-||_{(n)} = n/d$. Hence, the upper and lower bound in Prop.3.2 actually coincide and we get that T_η is n-positive iff $\eta \geq n$.[3] This proves that all the inclusions in the chain (3.3) are strict. When T_η is positive (i.e., for $\eta \geq 1$) it is decomposable as $T_\eta \theta$ is completely positive.

3.2 Positive maps and entanglement theory

A map, which is positive but not completely positive, like the transposition, does not correspond to a physically implementable operation. Nevertheless, such maps, and in particular the transposition and the map T_1 in Eq.(3.11), have become important mathematical tools in the theory of entanglement. Applied as $(T \otimes \mathrm{id})$ to a bipartite density matrix they are powerful tools for "detecting" entanglement as well as for recognizing "useful" (in the sense of distillable) entanglement.

[3] For the example T_η this can be proven in a simpler way (by averaging a Schmidt rank n vector w.r.t. the group $\{U \otimes \bar{U}\}$), without invoking Prop.3.2. We nevertheless stated the proposition as it can be applied to situations where τ does not have such a symmetry.

Detecting entanglement Suppose that $\rho \in \mathcal{M}_d \otimes \mathcal{M}_{d'}$ is a separable state as defined in Eq.(1.8), then positivity of T implies that $(T \otimes \mathrm{id})(\rho)$ remains positive semi-definite, since tensor products and convex combinations of positive operators are again positive. Hence, for any positive map T a negative eigenvalue of $(T \otimes \mathrm{id})(\rho)$ proves that ρ is entangled:

$$T \text{ positive, and } \langle\Psi|(T \otimes \mathrm{id})(\rho)|\Psi\rangle < 0 \quad \Rightarrow \quad \rho \text{ is entangled.} \tag{3.12}$$

We may go one step further and use an n-positive T. This gives the generalization

$$T \text{ } n\text{-positive, and } \langle\Psi|(T \otimes \mathrm{id})(\rho)|\Psi\rangle < 0 \quad \Rightarrow \quad \begin{array}{l}\text{Every pure state decomposition}\\ \text{of } \rho \text{ contains at least one vector}\\ \text{of Schmidt rank larger than } n.\end{array}$$

The smallest n such that ρ has a convex decomposition into pure states of Schmidt rank at most n is called the *Schmidt number* of ρ. With this terminology separable states are precisely those with Schmidt number one.

In order to prove the converse of the above implications it is useful to introduce the concept of *witnesses* for (Schmidt number $n+1$) entanglement.

Proposition 3.3 (Entanglement witnesses) *A density operator $\rho \in \mathcal{M}_d(\mathbb{C}) \otimes \mathcal{M}_{d'}(C)$ has Schmidt number at least $n+1$ iff there exists a Hermitian operator $W \in \mathcal{M}_d(\mathbb{C}) \otimes \mathcal{M}_{d'}(C)$ such that $\mathrm{tr}[W\rho] < 0$ but $\langle\psi|W|\psi\rangle \geq 0$ for all vectors $\psi \in \mathbb{C}^d \otimes \mathbb{C}^{d'}$ with Schmidt rank n.*

Proof Egg–hyperplane–done. In order to be a bit more precise, let us regard the space of Hermitian operators as a real vector space equipped with the Hilbert-Schmidt inner product. Since the set $\mathcal{S}_n$ of density operators with Schmidt number at most n is compact and convex, every $\rho \notin \mathcal{S}_n$ can be separated from it by a hyperplane of Hermitian operators H satisfying $\langle H, \tilde{W}\rangle = c$ for some constant c and some $\tilde{W} = \tilde{W}^\dagger$. That is, $\langle\rho, \tilde{W}\rangle < c$ whereas $\langle\rho', \tilde{W}\rangle \geq c$ for all $\rho' \in \mathcal{S}_n$. Setting $W := \tilde{W} - c\mathbb{1}$ then completes the 'only if' part of the proof. The 'if' part follows from the fact that $\mathcal{S}_n$ is the closure of the convex hull of the set of pure states with Schmidt rank n. □

An operator W with the properties mentioned in Prop.3.3 is called *entanglement witness* (for Schmidt number $n+1$ entanglement). We will follow the literature and use 'entanglement witness' referring to the case $n = 1$ unless otherwise specified. Invoking the Choi-Jamiolkowski correspondence (Prop.2.1) we can assign a positive map to each witness. This leads to the following characterization:

Proposition 3.4 (Positive maps and entanglement) *A density operator $\rho \in \mathcal{M}_d(\mathbb{C}) \otimes \mathcal{M}_{d'}(C)$ has Schmidt number at most n iff for all n-positive maps $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_{d'}(\mathbb{C})$ it holds that $(T \otimes \mathrm{id})(\rho) \geq 0$.*

Proof We only have to prove the 'if' part. For that assume that $\rho \notin \mathcal{S}_n$. Then by Prop.3.3 there is an entanglement witness for which $\mathrm{tr}[W\rho] < 0$. This defines, by the Choi-Jamiolkowski correspondence, a linear map $T^* : \mathcal{M}_{d'}(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ via

$$W =: (T^* \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|). \tag{3.13}$$

By item 3 in Prop.3.1 T^* and with it its adjoint T is n-positive and by construction

$$\mathrm{tr}\,[W\rho] = \langle\Omega|(T \otimes \mathrm{id})(\rho)|\Omega\rangle < 0. \tag{3.14}$$

Hence, $(T \otimes \mathrm{id})(\rho)$ cannot be positive. □

We see from Eq.(3.14) that the positive map constructed from a witness is actually more powerful, in the sense that it detects more entangled states, than the witness we started with: after all $(T \otimes \mathrm{id})(\rho)$ might have a negative eigenvalue without $\langle\Omega|(T \otimes \mathrm{id})(\rho)|\Omega\rangle$ being negative. A closer look reveals that a positive map T constructed from a witness W 'detects' an entangled state ρ iff it is detected by one of the witnesses from the orbit $\{(\mathbb{1} \otimes X)W(\mathbb{1} \otimes X)^\dagger\}$ with $X \in \mathcal{M}_{d'}(\mathbb{C})$.

Decomposable positive maps correspond to *decomposable witnesses* which are of the form

$$W = P_1 + P_2^{T_1}, \quad \text{with} \quad P_i \geq 0. \tag{3.15}$$

By duality, the existence of indecomposable witnesses for given dimensions is equivalent to the existence of entangled states satisfying $\rho^{T_1} \geq 0$:

Proposition 3.5 (PPT entangled states) [4] *There exists an indecomposable positive map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ iff there is an entangled state $\rho \in \mathcal{M}_d \otimes \mathcal{M}_{d'}$ with positive partial transpose $\rho^{T_1} \geq 0$.*

Proof We exploit the equivalence between indecomposable maps and witnesses. That is, we show that the existence of an indecomposable positive map implies that there is an entangled state with $\rho^{T_1} \geq 0$. The converse follows from Prop3.3. Denote by $\mathcal{W} \subseteq \mathcal{M}_d \otimes \mathcal{M}_{d'}$ the set of all entanglement witnesses, i.e., Hermitian operators with $\langle\phi \otimes \psi|W|\phi \otimes \psi\rangle \geq 0$ for all ϕ, ψ, and by $\mathcal{W}_T \subseteq \mathcal{W}$ all decomposable witnesses. Both $\mathcal{W}$ and $\mathcal{W}_T$ are closed convex sets. If there is (by assumption) a $W \in \mathcal{W} \setminus \mathcal{W}_T$ we can separate it by a hyperplane characterized by a Hermitian operator R such that $\mathrm{tr}\,[WR] < 0 \leq \mathrm{tr}\,[W_T R]$ for all $W_T \in \mathcal{W}_T$. Since $\mathcal{W}_T$ contains all positive operators R must be positive so that $\rho := R/\mathrm{tr}\,[R]$ is a density matrix. By construction $\rho^{T_1} \geq 0$ while ρ is entangled due to $\mathrm{tr}\,[W\rho] < 0$. □

We will see shortly that entangled PPT states exists iff the dimension is larger than $\mathbb{C}^2 \otimes \mathbb{C}^3$.

Example 3.1 (Some positive maps and entanglement witnesses)

Transposition *is a positive map. It corresponds to the set of witnesses $W = (A \otimes B)\mathbb{F}(A \otimes B)^\dagger$ for arbitrary A, B. The separability criterion $(\theta \otimes \mathrm{id})(\rho) := \rho^{T_1} \geq 0$ is called* PPT (positive partial transpose) criterion*. Although the partial transpose depends on the choice of the local basis, the PPT criterion does not: if we take the*

[4] Admittedly this proposition has mainly and at most pedagogical value. Once examples of entangled PPT states are constructed it looses a lof of its charm...

partial transposition with respect to a different (local) basis, labeled by the superscript $^{\tilde{T}_1}$*, we get*

$$\rho^{\tilde{T}_1} = (U\otimes \mathbb{1})\big[(U^\dagger\otimes\mathbb{1})\rho(U\otimes\mathbb{1})\big]^{T_1}(U^\dagger\otimes\mathbb{1}) \tag{3.16}$$
$$= \big[(UU^T)\otimes\mathbb{1}\big]\rho^{T_1}\big[(UU^T)^\dagger\otimes\mathbb{1}\big]. \tag{3.17}$$

Eqs.(3.16-3.17) show that the eigenvalues of the partial transpose are basis independent, and so is the positivity of ρ^{T_1}*. Moreover,* $\rho^{T_1}\geq 0$ *is equivalent to* $\rho^{T_2}\geq 0$ *since the two operators only differ by a global transposition. We will see in ... that the PPT criterion is necessary and sufficient for separability in* $\mathbb{C}^2\otimes\mathbb{C}^2$ *and* $\mathbb{C}^2\otimes\mathbb{C}^3$*.*

Reduction criterion. *The set of maps* $T_n(X) := \mathrm{tr}\,[X]\,\mathbb{1} - X/n$*,* $n\in\mathbb{N}$ *in Eq.(3.11) is n-positive (and self-dual, i.e.,* $T_n^* = T_n$*). As a consequence*

$$n\rho_1\otimes\mathbb{1}\geq\rho \quad \textit{and} \quad n\mathbb{1}\otimes\rho_2\geq\rho, \tag{3.18}$$

(with ρ_k *the k'th reduced density matrix of* ρ*) is necessary for a bipartite state* ρ *to have Schmidt number n. For* $n=1$ *this is called* reduction criterion*. As* T_n *is a decomposable positive map, the reduction criterion is generally weaker than the PPT criterion for entanglement detection. For* $\mathbb{C}^2\otimes\mathbb{C}^2$ *the two criteria are equivalent since* $T_1(X) = \big(\sigma_y X\sigma_y\big)^T$ *for any* $X\in\mathcal{M}_2(\mathbb{C})$*. An entanglement witness corresponding to* T_n *is* $W_n = \mathbb{1}/d - |\Omega\rangle\langle\Omega|/n$*. That is, a state* ρ *with Schmidt number n has to satisfy* $\langle\Omega|\rho|\Omega\rangle\leq n/d$ *which is a weak version of Lemma 3.1.*

Breuer-Hall map $T_{BH}:\mathcal{M}_d\to\mathcal{M}_d$ *is positive and defined as*

$$T_{BH}(X) = \mathrm{tr}\,[X]\,\mathbb{1} - X - UX^TU^\dagger, \quad \textit{for any } U \textit{ with } U^T = -U \textit{ and } U^\dagger U\leq\mathbb{1}. \tag{3.19}$$

Positivity is easily seen by applying T_{BH} *to a pure state and noting that due to the anti-symmetry of U we have* $\langle\psi|U|\bar{\psi}\rangle = 0$ *which implies that the subtracted terms are orthogonal. For d even there are anti-symmetric unitaries* $U^\dagger = U^{-1} = -\bar{U}$ *(e.g. embeddings of the Pauli matrix* σ_y*), whereas for d odd these do not exist (which can be seen from the fact that* $\det(U) = \det(U^T) = \det(-U)$ *would be equal to* $-\det(U)$ *in this case; see Thms. 3.1,3.2). If U is an anti-symmetric unitary* T_{BH} *is not decomposable.*

Choi-type maps. *Let* $D\in\mathcal{B}(\mathcal{M}_d)$ *map a matrix X onto a diagonal matrix* $D(X)$ *with the same diagonal entries, and let* U_{k0} *be a cyclic shift as defined in Eq.(2.24). Then for all* $n\in\mathbb{N}$ *with* $1\leq n\leq d-2$ *the map* $T_C\in\mathcal{B}(\mathcal{M}_d)$ *defined by*

$$T_C(X) := (d-n)D(X) - X + \sum_{k=1}^{n} D\big(U_{k0}XU_{k0}^\dagger\big), \tag{3.20}$$

is positive and indecomposable. For $n=0$ *it is completely positive and for* $n=d-1$ *it is proportional to the map* T_n *appearing in the context of the reduction criterion.*

Unextendable product bases *(UPBs) are sets* $S = \{|\alpha_j\rangle\otimes|\beta_j\rangle\}$ *of normalized and mutually orthogonal product vectors in* $C^d\otimes C^{d'}$ *(*$d'\geq d$*) such that there is no product vector, which is orthogonal to every element of S, and* $|S| < dd'$*.*[5] *That is, the projection* $P_S := \sum_{j=1}^{|S|}|\alpha_j\rangle\langle\alpha_j|\otimes|\beta_j\rangle\langle\beta_j|$ *has a kernel which contains no product vector. An example of such a UPB in dimension* $3\otimes 3$ *can be constructed from five*

[5] Note that this is only possible if $\{|\alpha_j\rangle\}$ (resp. $\{|\beta_j\rangle\}$) is not a set of mutually orthogonal vectors.

real vectors forming the apex of a regular pentagonal pyramid, where the height h is chosen such that nonadjacent apex vectors are orthogonal. These vectors are

$$v_j = N\left(\cos\frac{2\pi j}{5}, \sin\frac{2\pi j}{5}, h\right), \quad j = 0, \ldots, 4 \;, \tag{3.21}$$

with $N = 2/\sqrt{5+\sqrt{5}}$ and $h = \frac{1}{2}\sqrt{1+\sqrt{5}}$. The UPB is then given by $|\alpha_j\rangle \otimes |\beta_j\rangle = |v_j\rangle \otimes |v_{2j \; mod 5}\rangle$. Since any subset of three vectors on either side spans the full space, there cannot be a product vector orthogonal to all these states. Based on this or any other UPB, one can easily construct an entangled PPT state:

$$\rho_S \propto (\mathbb{1} - P_S). \tag{3.22}$$

Since by construction ρ_S has no product vector in its range it has to be entangled. Moreover, $\rho^{T_1} \geq 0$, since $\{|\bar{\alpha}_j\rangle \otimes |\beta_j\rangle\}$ is again a set of mutually orthogonal and normalized vectors. From here we can construct an entanglement witness

$$W_S = P_S - \epsilon|\psi\rangle\langle\psi|, \tag{3.23}$$

where ψ is any maximally entangled state of dimension d fulfilling $\langle\psi|\rho_S|\psi\rangle > 0$ and $\epsilon := d \inf_{\{\phi_i\}} \langle\phi_1\otimes\phi_2|P|\phi_1\otimes\phi_2\rangle/||\phi_1\otimes\phi_2||^2$. So by construction $\mathrm{tr}\,[W_S\rho_S] < 0$ whereas $\mathrm{tr}\,[W_S\rho] \geq 0$ for all separable states. By the Choi-Jamiolkowski correspondence W_S thus defines a positive map which is not decomposable.

Entanglement distillation ... TO BE WRITTEN ...
LOCC & separable superoperators
PPT implies not distillable
reduction criterion, distillability (maybe majorization)

3.3 Transposition and time reversal

Due to the lack of complete positivity, the transposition (as well as the partial transposition) does not correspond to a physically implementable operation. However, it can physically be interpreted as "time reversal" [**?**] and appears as such quite often as a (global) symmetry, in particular in particle physics [**?**]. Although we will not make explicit use of this interpretation or symmetry in the following, we will briefly describe this relation.

One way to see that the transposition or complex conjugation (which is the same on Hermitian operators) admits an interpretation as time reversal, is looking at the *Wigner function* of a continuous variable system in Schrödinger representation.[6] If we choose $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i ||\psi_i||^2 = 1$ to be a decomposition of a density operator ρ and define an integral kernel $\mathcal{D}(x;y) = \sum_i \lambda_i \overline{\psi}_i(x)\psi_i(y)$ in the Schrödinger representation, then we can write the Wigner function as

$$\mathcal{W}(x,p) = \pi^{-n} \int \mathcal{D}(q-x; q+x)\; e^{-i2x\cdot p}\; d^n x \;, \tag{3.24}$$

[6] Note that we leave the case $\dim \mathcal{H} < \infty$ within this section. The underlying Hilbert space is now the space $\mathcal{L}^2$ of square integrable functions.

where n is the number of canonical degrees of freedom, i.e., $x, p, q \in \mathbb{R}^n$. From Eq.(3.24) we see, that the complex conjugation or transposition $\mathcal{D}(x; y) \mapsto \mathcal{D}(y; x)$ is equivalent to a reversal of momenta $p \mapsto -p$. Due to the basis dependency of the transposition this is, however, no longer true if we choose another representation. Hence we cannot identify the transposition or complex conjugation with time reversal in general.

Time is a parameter in quantum mechanics. Therefore there cannot be a general transformation in Hilbert space which directly affects this parameter and acts as $t \mapsto -t$. In this sense the term "time reversal" is misleading. It characterizes a reversal of momenta including angular momenta and spins, rather than a reversal of the arrow of time. The defining ansatz for time reversal is thus an operator $\mathcal{T}$ which acts on Hilbert space in a way that position and momentum operators transform as

$$\mathcal{T}Q\mathcal{T}^{-1} = Q \quad \text{and} \quad \mathcal{T}P\mathcal{T}^{-1} = -P, \tag{3.25}$$

and for consistency also $\mathcal{T}L\mathcal{T}^{-1} = -L$ and $\mathcal{T}S\mathcal{T}^{-1} = -S$ for angular momenta and spins. Left and right multiplication of the canonical commutation relations $[Q, P] = i\mathbb{1}$ leads with Eq.(3.25) to $\mathcal{T}i\mathcal{T}^{-1} = -i\ \mathbb{1}$. Thus $\mathcal{T}$ is an "anti-linear" operator, and, as it is supposed to be norm preserving, it is "anti-unitary".[7]

The simplest example of an anti-unitary operator is the complex conjugation Γ (which is on Hermitian matrices equal to the transposition). In fact, in Schrödinger representation we may identify the time reversal operator $\mathcal{T}$ with complex conjugation, since the latter leaves the position operator unchanged and maps $-i\frac{d}{dx} \overset{\Gamma}{\mapsto} i\frac{d}{dx}$ and therefore changes the sign of momentum and angular momentum operators. As the time reversal operator should not depend on the representation (resp. basis), $\mathcal{T}$ cannot be identified with Γ in general. However, every anti-unitary operator can be written as the product of a unitary operator with complex conjugation.

Let us now turn to spin angular momenta. Consider the standard representation, in which S_x and S_z are real and S_y is imaginary. In this basis we have

$$\Gamma S_x \Gamma^{-1} = S_x, \quad \Gamma S_y \Gamma^{-1} = -S_y, \quad \Gamma S_z \Gamma^{-1} = S_z, \tag{3.26}$$

hence we cannot identify $\mathcal{T}$ with complex conjugation. However, we may write $\mathcal{T} = \Gamma V$, where V has to invert the signs of S_x and S_z, while leaving S_y unchanged. Moreover, V is, as a product of two anti-unitaries, a linear unitary operator, and it is supposed to act only on the spin degrees of freedom. These requirements are satisfied by $V = \exp -i\pi S_y$, which rotates the spin through the angle π about the y axis. Hence, we have with respect to the considered representation[8]

$$\mathcal{T} = \Gamma e^{-i\pi S_y}, \tag{3.27}$$

[7]An anti-linear operator A satisfies $Az = \bar{z}A$ for any complex number z. The product of two anti-linear operators is thus again a linear operator. If the inverse A^{-1} exists and $\forall \Psi \in \mathcal{H} : ||A\Psi|| = ||\Psi||$, then we say that A is anti-unitary. We have then $\langle A\Phi|A\Psi\rangle = \langle \Psi|\Phi\rangle$.

[8]Note that time reversal squared is not the identity. The operator $\mathcal{T}^2$ in Eq.(3.27) has eigenvalues -1 for states with total spin $n/2$, with n odd.

and applying the time reversal to a density matrix describing a finite dimensional spin system gives then $\mathcal{T}^{-1}\rho\mathcal{T} = V^\dagger \rho^T V$.

Let us now consider (finite dimensional) composite systems, where $\rho \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. The global time reversal operator is then $\mathcal{T} \otimes \mathcal{T}$, which is a well defined anti-unitary operator again. Unfortunately, there exists no operator on Hilbert space, which corresponds to a "partial time reversal" of only one of the two subsystems. In fact, the action of a tensor product of the form $\mathrm{id} \otimes \mathcal{T}$ (linear$\otimes$anti-linear) on a vector $\Psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is not well defined. However, since for product vectors $(\mathrm{id} \otimes \mathcal{T}^{-1})(|\psi\rangle \otimes |\phi\rangle)$ gives only rise to a phase ambiguity, the action on the respective projectors

$$\left(\mathrm{id} \otimes \mathcal{T}^{-1}\right)\left(|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|\right)\left(\mathrm{id} \otimes \mathcal{T}\right) \tag{3.28}$$

becomes well defined without any ambiguity, and Eq.(3.28) describes a linear operator again. Since the real linear hull of projectors onto product states equals the set of all Hermitian operators, we can in this way define the "partial time reversal" as a map on density matrices. Due to Eq.(3.27) this is equivalent to the partial transposition up to a local unitary transformation. Note that the latter is irrelevant for the detection of entanglement since it preserves the eigenvalues.

The time reversal symmetry of a dynamics, i.e., the invariance of a Hamiltonian under time reversal, does not lead to a conserved quantity (like parity for space inversion). However, it sometimes leads to a degeneracy of the eigenvalues:

Suppose $[H, \mathcal{T}] = 0$ for some anti-unitary $\mathcal{T}$ and Hermitian H. Writing $\mathcal{T} = \Gamma V$ with V unitary this becomes equivalent to $HV^\dagger = V^\dagger H^T$. Moreover, $\mathcal{T}^2 = -\mathbb{1}$ is equivalent to $V^T = -V$ being anti-symmetric.

Theorem 3.1 (Kramer's theorem) *If $[H, \mathcal{T}] = 0$ and $\mathcal{T}^2 = -1$ for a Hermitian H and anti-unitary $\mathcal{T}$, then each eigenvalue of H is at least two-fold degenerate.*

Proof Let $H|\psi\rangle = \lambda|\psi\rangle$. Then $HV^\dagger|\bar\psi\rangle = V^\dagger H^T|\bar\psi\rangle = V^\dagger \bar H|\bar\psi\rangle = \lambda V^\dagger|\bar\psi\rangle$, so $V^\dagger|\bar\psi\rangle$ gives rise to the same eigenvalue. This is degenerate since the two vectors are orthogonal:

$$\langle\psi|V^\dagger|\bar\psi\rangle = -\langle\psi|\bar V|\bar\psi\rangle = -\overline{\langle\bar\psi|V|\psi\rangle} \tag{3.29}$$

$$= -\langle\psi|V^\dagger|\bar\psi\rangle. \tag{3.30}$$

□

Note that anti-symmetric unitaries only exist in even dimensions.[9] In fact, the restriction on the required symmetry in the theorem can be relaxed as the same proof shows:

Theorem 3.2 (Kramer's theorem II) *Let H be Hermitian and $HA = AH^T$ for some anti-symmetric $A \neq 0$. Then each eigenvalue of H is at least two-fold degenerate.*

[9] In odd diemnsions we get that $\det(V) = \det(V^T) = -\det(V)$ which contradicts the fact that the determinant of a unitary has to be a phase.

3.4 From Hilbert's 17th problem to indecomposable maps

The following pages will somehow be a detour following a historical path. Rather than writing down examples of indecomposable positive maps (as we did within example 3.1) we will see how they arose from a problem Hilbert thought about in 1888. For an overview of Hilbert's 17th problem see [?].

Hilbert's problem: A sum of squares (sos) of real polynomials is obviously positive semi-definite. The converse question, whether a real homogeneous psd polynomial of even degree has a sos decomposition, was posed and answered by David Hilbert in 1888. Let $H_h(\mathbb{R}^n)$ be the set of homogeneous polynomials of even degree h in n real variables, and

$$P_h(\mathbb{R}^n) = \{p \in H_h(\mathbb{R}^n) \big| \forall x \in \mathbb{R}^n : p(x) \geq 0\}, \tag{3.31}$$

$$\Sigma_h(\mathbb{R}^n) = \Big\{p \in P_h(\mathbb{R}^n) \big| \exists \{h_k \in H_{h/2}(\mathbb{R}^n)\} : p = \sum_k h_k^2\Big\}, \tag{3.32}$$

the set of psd polynomials in H and the subset of polynomials having an sos decomposition, respectively.[10] It is not difficult to see that $P = \Sigma$ if $n = 2$ or $h = 2$. In fact, the latter is nothing but the spectral decomposition of psd matrices. In 1888 Hilbert showed at first, that every $p \in P_4(\mathbb{R}^3)$ can be written as the sum of squares of three quadratic forms, i.e., $P_4(\mathbb{R}^3) = \Sigma_4(\mathbb{R}^3)$. Moreover, he proved, albeit in a non-constructive way, that $(h, n) = (2, n), (h, 2), (4, 3)$ are in fact the only cases where $P = \Sigma$. In 1900 Hilbert posed a generalization of these questions as his "17th problem" at the International Congress of Mathematics in Paris[11]: Does every homogeneous psd form admit a decomposition into a sum of squares of rational functions? This was answered in the affirmative by Artin in 1927 using the Artin-Schreier theory of real closed fields [?] .[12]

Choi's example: Concerning Hilbert's original 1888 work it took nearly 80 years for explicit polynomials $p \in P \backslash \Sigma$ to appear in the literature. One of these examples was provided by Choi, while investigating one of the numerous "sum of squares" variants of Hilbert's problem. Choi showed in 1975 that there are psd biquadratic forms that cannot be expressed as the sum of squares of bilinear forms. More precisely, let

$$F(x, y) = \sum_{i,j,k,l=1}^{n} F_{ijkl} x_i x_j y_k y_l \ , \quad \forall x, y : F(x, y) \geq 0 \tag{3.33}$$

[10]Note that P and Σ are both convex cones. Moreover, they are closed, i.e., if $p_n \to p$ coefficientwise, and each p_n is in P resp. Σ, then so is p.

[11]At this congress Hilbert outlined 23 major mathematical problems. Whereas some are broad, such as the axiomatization of physics (6th problem), others were very specific and could be solved quickly afterwards.

[12]Artin's proof holds for forms with coefficients from a field with unique order, i.e., in particular for $\mathbb{R}$.

be a real psd biquadratic form in $x, y \in \mathbb{R}^n$. For $n = 3$ Choi found a counterexample to the (wrongly proven[13]) conjecture that there is always a decomposition into a sum of squares of bilinear forms, i.e.

$$F(x,y) \stackrel{?}{=} \sum_\alpha \big(f^{(\alpha)}(x,y)\big)^2 = \sum_\alpha \Big(\sum_{ij=1}^n f_{ij}^{(\alpha)} x_i y_j\Big)^2. \tag{3.34}$$

Choi's counterexample, for which he gave very elementary proofs, is

$$F(x,y) = 2\sum_{i=1}^3 x_i^2(y_i^2 + y_{i+1}^2) - \Big(\sum_{j=1}^3 x_j y_j\Big)^2, \tag{3.35}$$

with $y_{3+1} \equiv y_1$. Note that $F \in P_4(\mathbb{R}^6)$ and that every quadratic form appearing in an sos decomposition of F has to be bilinear. Hence, Eq.(3.35) is an element of $P\backslash\Sigma$ for $(h,n) = (4,6)$.[14]

For the mere beauty of the example let us mention an element of $P\backslash\Sigma$ for $(h,n) = (4,5)$:

$$\sum_{i=1}^5 \prod_{j\neq i}(x_i - x_j). \tag{3.36}$$

Relation to positive maps: In our context the main significance of biquadratic forms lies in their relation to linear maps on symmetric matrices. This relation is in fact the real analogue of the correspondence between entanglement witnesses and positive maps expressed in Eq.(3.13). Let $\mathcal{S}_n$ be the set of all real symmetric $n\times n$ matrices, and let $\Phi : \mathcal{S}_n \to \mathcal{S}_n$ be a positive linear map on $\mathcal{S}_n$, i.e. $\forall x \in \mathbb{R}^n : \Phi\big(|x\rangle\langle x|\big) \geq 0$. Then Φ corresponds to a psd biquadratic form F and vice versa via[15]

$$F(x,y) = \langle y|\Phi\big(|x\rangle\langle x|\big)|y\rangle. \tag{3.37}$$

Particular instances of such maps are *congruence maps* of the form $\Phi(A) = K^T A K$, which in turn correspond to psd biquadratic forms $F(x,y) = f(x,y)^2$, where $f(x,y) = \sum_{ij} K_{ij} x_i y_j$ is a bilinear form. The existence of a psd biquadratic form, which does not admit an sos decomposition, thus implies that the set of positive maps $\Phi : \mathcal{S}_n \to \mathcal{S}_n$ properly contains the convex hull of all congruence maps (for $n \geq 3$). The latter is, however, the real analogue of the set of (complex) decomposable positive maps. Let us now see, whether the map

$$\Phi(S) = 2\big[\mathrm{tr}\,[S]\,\mathbb{1} - \mathrm{diag}(S_{33}, S_{11}, S_{22})\big] - S, \tag{3.38}$$

[13] For the case $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ with $n = 2$ or $m = 2$ Calderon [?] proved that there is always a sos decomposition of the form in Eq.(3.34). In [?] Koga claimed that this is also true for arbitrary n, m. Finding the flaw in Koga's proof, was apparently Choi's motivation for [?].

[14] By choosing dependent x and y, Choi also specified $F(x,y)$, such that it yielded elements of $P\backslash\Sigma$ for $(h,n) = (4,4)$ and $(6,3)$ respectively.

[15] $F \geq 0$ implies $\Phi \geq 0$ since any symmetric matrix $S \in \mathcal{S}_n$ is psd if $\langle y|S|y\rangle \geq 0$ for every real vector $y \in \mathbb{R}^n$.

which corresponds to the psd biquadratic form in Eq.(3.35) (see also example 3.1), represents an admissible complex non-decomposable positive map as well. Let Λ be a complex extension of Φ defined as in Eq.(3.38) but on arbitrary complex 3×3 matrices. Obviously, Λ maps Hermitian matrices onto Hermitian matrices. Moreover, if Λ is a decomposable positive map, it acts on symmetric matrices as $\Lambda(S) = \sum_j K_j^* S K_j$ for some complex matrices $\{K_j\}$. If we decompose the latter with respect to their real and imaginary parts $K_j = R_j + iI_j$ we get

$$\Lambda(S) = \sum_j R_j^T S R_j + I_j^T S I_j \stackrel{!}{=} \Phi(S), \tag{3.39}$$

contradicting the fact that Φ in Eq.(3.38) has no such decomposition into congruence maps. Hence, Λ is either not decomposable or not a positive map at all. The latter can, however, be excluded, since we have for every vector $z \in \mathbb{C}^3$ with $z_j = x_j \varphi_j$, $x_j = |z_j|$:

$$\Lambda\big(|z\rangle\langle z|\big) = D^* \Phi\big(|x\rangle\langle x|\big) D \ \geq \ 0 \, , \quad D = \mathrm{diag}(\varphi_1, \varphi_2, \varphi_3). \tag{3.40}$$

By Prop.(3.5) Choi's map thus proves, albeit in a non-constructive way, the existence of PPT entangled states in $3 \otimes 3$.

The question whether an entanglement witness W is decomposable or not, is equivalent to asking whether the biquadratic Hermitian psd form

$$W(\Phi, \Psi) = \langle \Phi \otimes \Psi | W | \Phi \otimes \Psi \rangle = \sum_{ijkl} W_{ijkl} \bar{\Phi}_i \bar{\Psi}_j \Phi_k \Psi_l \tag{3.41}$$

has an sos decomposition. It is therefore the complex analog of Eq.(3.34). Apart from examples there is, however, not much known about forms which do not admit such a decomposition.

Every real unextendable product basis (as the one in Eq.(3.21)) can be used to construct a real-valued non-decomposable entanglement witness. These witnesses thus correspond to biquadratic psd forms, which do not admit an sos decomposition and are therefore elements of $P \backslash \Sigma$.

3.5 Preservation of cones and balls

Proposition 3.6 (Automorphisms and rank preserving maps) *Consider a linear map $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$. The following are equivalent:*

1. *T maps the cone of positive semidefinite matrices onto itself.*
2. *T is positive and preserves the rank of Hermitian matrices.*
3. *There is an invertible $Y \in \mathcal{M}_d(\mathbb{C})$ such that T is either of the form $T(X) = YXY^\dagger$ or $T(X) = YX^TY^\dagger$.*

PROOF $1 \to 2$: Note first that T is bijective and T^{-1} is again positive, linear and maps the positive semidefinite cone onto itself. Consider an arbitrary positive semidefinite $P \in \mathcal{M}_d(\mathbb{C})$ and the cone defined by

$$\mathcal{C}(P) := \{A \in \mathcal{M}_d(\mathbb{C}) \mid \exists c > 0 : P \geq cA \geq 0\}. \tag{3.42}$$

Note that the subspace spanned by $\mathcal{C}(P)$ has dimension $\mathrm{rank}(P)^2$ and that, by the properties of T, we have $T(\mathcal{C}(P)) = \mathcal{C}(T(P))$. Since T (as a bijection) preserves the dimension of subspaces, this implies that $\mathrm{rank}(P) = \mathrm{rank}(T(P))$. In order to extend this to Hermitian matrices let us decompose an arbitrary $H = H^\dagger$ into $H = P_+ - P_-$ where $P_\pm$ are positive semidefinite and orthogonal. Then $\mathrm{rank}(T(H)) \leq \mathrm{rank}(T(P_+)) + \mathrm{rank}(T(P_-)) = \mathrm{rank}(H)$. Applying the same to $T^{-1}(H)$ instead of H gives the converse inequality.

$2 \to 3$: Define $\tilde{T}(X) := T(\mathbb{1})^{-1/2} T(X) T(\mathbb{1})^{-1/2}$. Since $T(\mathbb{1})$ has full rank, $\tilde{T}$ inherits the properties of T, i.e., it is positive and rank preserving and in addition unital. Now recall that $\lambda \in \mathrm{spec}(H)$ iff $H - \lambda\mathbb{1}$ has reduced rank. Using that $\tilde{T}(H - \lambda\mathbb{1}) = \tilde{T}(H) - \lambda\mathbb{1}$ together with the fact that $\tilde{T}$ preserves the rank for Hermitian H we get that it actually preserves the spectrum. Then Corr.1.1 completes this part of the proof by setting $Y = U(T(\mathbb{1}))^{-1/2}$. $3 \to 1$ should be obvious. □

From the definition of decomposable positive maps in Eq.(3.2) together with the Kraus decomposition in Thm.2.1 it is clear that every extreme point within the convex set of decomposable positive maps is of the form $X \mapsto YXY^\dagger$ or $X \mapsto YX^TY^\dagger$ (with Y not necessarily invertible). That the converse is true as well, i.e., that all these maps are extreme point has been proven in ... .

Proposition 3.7 (Confining Lorentz cones) *Consider a Hermitian matrix $A \in \mathcal{M}_d(\mathbb{C})$. Then*

1. $A \geq 0 \Rightarrow \mathrm{tr}[A]^2 \geq \mathrm{tr}[A^2]$. *Conversely,*
2. $A \geq 0 \Leftarrow \mathrm{tr}[A]^2 \geq (d-1)\mathrm{tr}[A^2]$.

PROOF 1. becomes obvious when writing it out in terms of the eigenvalues $\{a_i \in \mathbb{R}\}$: $(\sum_i a_i)^2 \geq \sum_i a_i^2$ is certainly true for $a_i \geq 0$ as the l.h.s. contains all terms of the r.h.s. plus some extra positive terms.

In order to prove the second implication (by contradiction) assume that A has a negative part such that it admits a non-trivial decomposition of the form $A = A_+ - A_-$ with $A_\pm \geq 0$ and $\mathrm{tr}[A_+ A_-] = 0$. Denoting by $d_+ < d$ the rank of A_+ we can bound

$$\mathrm{tr}[A] \leq \mathrm{tr}[A_+] = ||A_+||_1 \leq \sqrt{d_+}||A_+||_2 \tag{3.43}$$

$$\leq \sqrt{d-1}||A_+||_2 < \sqrt{d-1}||A||_2, \tag{3.44}$$

where the first inequality is a simple consequence of Hölder's inequality[16] and the last one uses $||A||_2^2 = ||A_+||_2^2 + ||A_-||_2^2$ and $A_- \neq 0$. □

[16] For every unitarily invariant norm we have $||CB|| \leq ||\,|C|^p||^{1/p}||\,|B|^q||^{1/q}$ for all $p > 1$ and $1/p + 1/q = 1$. Eq.(3.43) follows from applying this to the trace norm with $p = 2$, $C = A_+$ and B the projection onto the range of A_+.

Lemma 3.2 (Making positive maps trace preserving) *Let* $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ *be a positive map for which* $T^*(\mathbb{1}) > 0$. *There exists an invertible* $X \in \mathcal{M}_d(\mathbb{C})$ *such that* $T'(\rho) := T(X\rho X^\dagger)$ *is a trace-preserving positive map.*

Proof Let $\tau := (T \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$ be the Choi-Jamiolkowski operator corresponding to T and τ' the one for T'. By Prop.2.1 the map T' is trace preserving if the partial trace τ'_B equals $\mathbb{1}/d$. Since (again by Prop.2.1) $\tau'_B = X^T \tau_B \bar{X} = X^T T^*(\mathbb{1})^T \bar{X}/d$ this can be achieved by $X := \sqrt{T^*(\mathbb{1})/d}$. □

Hildebrand: ... maps preserving the Lorentz cone ...

Gorini, Sudarshan: ... maps preserving the sphere + Kossakowski and Chruscinski ...

3.6 Complexity issues

3.7 Literature

Chapter 4

Convex Structure

4.1 Convex optimization and Lagrange duality

Admittedly, a section about optimization techniques is a bit off topic. However, convex optimization and, in particular, *semidefinite programming*[1] turned out to be an extremely useful tool in the context of quantum information theory, so we will at least sketch the basic framework here.

The utility of these methods is threefold: (i) they provide powerful algorithms to solve problems numerically. In contrast to many other numerical techniques these algorithms are certifiable. That is, they return a result together with a (usually negligeable) error bound which is essentially enabled by duality theory. (ii) Duality theory allows to convert numerical results into analytic proofs and often enough gives some analytic insight (if not a solution) without exploiting any silicon brain. (iii) Having expressed a problem in terms of e.g. a semidefinite program may be considered as a solution of the problem in very much the same way as we consider a reduction to an eigenvalue problem to be a solution. Moreover, it enables a classification of the problems complexity in the sense of complexity theory since interior point methods provide a general polynomial-time solution.

Conic programs Consider finite-dimensional real Hilbert spaces $\mathcal{V}$ and $\mathcal{V}'$. These may be bare $\mathbb{R}^d$'s but could also be spaces of Hermitian matrices or spaces of polynomials of a certain degree. Suppose $\mathcal{K} \subset \mathcal{V}$ is a closed convex cone which is pointed ($\mathcal{K} \cap -\mathcal{K} = 0$) and has a non-empty interior. For any linear map $T : \mathcal{V} \to \mathcal{V}'$ and $c \in \mathcal{V}$, $b \in \mathcal{V}'$ the optimization problem

$$\inf_{x\in\mathcal{K}} \left\{ \langle c|x\rangle \;\middle|\; T(x) = b \right\} =: C_p \tag{4.1}$$

[1] Depending on what community one comes from it needs some time to get used to the terminology: conic, semidefinite or linear *programs* are just optimization problems of a certain type.

is called a *convex conic program*. Assigned to this is a *dual* problem[2] which reads

$$\sup_{y\in\mathcal{V}'} \big\{ \langle b|y\rangle \;\big|\; c - T^*(y) \in \mathcal{K}^* \big\} =: C_d, \tag{4.2}$$

where T^* is the dual map (i.e., $\langle T^*(y)|x\rangle = \langle y|T(x)\rangle$) and $\mathcal{K}^*$ is the dual cone defined by $\mathcal{K}^* := \{x \in \mathcal{V} | \forall z \in \mathcal{K} : \langle z|x\rangle \geq 0\}$.

It holds in general that $C_p \geq C_d$ – this is called *weak duality*. *Strong duality*, meaning $C_p = C_d$, holds if $T(x) = b$ is fulfilled for some x which is in the interior of $\mathcal{K}$. In this case the problem in Eq.(4.1) is called *strictly feasible* and the supremum C_d is attained in the dual. Similarly, strong duality holds if the dual problem in Eq.(4.2) is strictly feasible (i.e., $c - T^*(y)$ is in the interior of $\mathcal{K}^*$ for some y). In this case the infimum C_p is attained in Eq.(4.1).

For many conic programs there exist polynomial-time algorithms. Roughly speaking, the requirements for that are that membership in the cone can be decided efficiently (like for the cone of positive semidefinite matrices) and that bounds on the norm of candidate solutions can be given.

Semidefinite programs are special instances of convex conic programs. Here T is a linear map between spaces of Hermitian matrices and $\mathcal{K} = \mathcal{K}^*$ is the cone of positive semidefinite matrices. Traditionally, semidefinite programs are phrased in a slightly different way than in Eqs.(4.1,4.2). The weak duality inequality is usually stated as

$$\inf_{X\geq 0} \big\{ \mathrm{tr}\,[F_0 X] \;\big|\; \mathrm{tr}\,[F_i X] = b_i \big\} \geq \sup_{y\in\mathbb{R}^n} \left\{ \langle b|y\rangle \;\middle|\; F_0 \geq \sum_{i=1}^n y_i F_i \right\}, \tag{4.3}$$

where $b \in \mathbb{R}^n$, $F_0, F_1, \ldots, F_n$ as well as X are Hermitian matrices and "$\geq$" for matrices denotes the standard matrix ordering induced by the cone of positive semidefinite matrices. Strong duality, meaning equality in Eq.(4.3), holds if at least one of the two optimization problems is strictly feasible. That is, if there is either an $X > 0$ with $\mathrm{tr}\,[F_i X] = b_i$ or a y such that $F_0 > \sum_i y_i F_i$. As mentioned in the previous paragraph, if one of the problems is strictly feasible, then the extremum in the dual problem is attained.

If the extrema in Eq.(4.3) are both attained and equal (i.e., in particular strong duality holds) then a simple condition called *complementary slackness* relates optimal solutions X' and y':

$$\big[F_0 - \sum_i y_i' F_i\big] X' = 0. \tag{4.4}$$

[2]The relation between the *primal problem* in (4.1) and the dual problem (4.2) is more symmetric than it appears at first sight. In particular, the dual problem is a convex conic program in its own right. To see the symmetric structure note that $T(x) = b$ defines an affine set which can be parametrized by $\tilde{y} \in \mathbb{R}^n$ via $x = e_0 + \sum_{i=1}^n \tilde{y}_i e_i$ with suitable $e_i \in \mathcal{V}$. Defining further $\tilde{b}_i := \langle c|e_i\rangle$ we can rewrite the primal problem without equality constraints in a form similar to that of Eq.(4.2) up to a constant offset (since $\langle \tilde{b}|\tilde{y}\rangle = \langle c|x\rangle - \langle c|e_0\rangle$) and a sign which has to be spent to turn the inf into a sup.

That is, X' and $\left[F_0 - \sum_i y_i' F_i\right]$ have to have orthogonal supports. Moreover, any y' is then optimal iff there exists an $X' \geq 0$ such that Eq.(4.4) holds and the constraints $\mathrm{tr}\,[F_i X'] = b_i$ and $F_0 \geq \sum_i y_i' F_i$ are satisfied.

4.2 Literature

Convex optimization is described in detail in the textbook [**?**]. [**?**] is a good source for more on semidefinite programs and more on conic programs, in particular about complexity related questions, can be found in [**?**].

Chapter 5

Operator Inequalities

5.1 Operator ordering

The set of Hermitian operators is equipped with a natural partial order, i.e., a consistent way of saying that one operator is larger than another one or that two operators are actually incomparable. In fact, we will come across various partial orders, but spend most time on what is sometimes called *Löwner's partial order* which is based on positive (semi-)definite differences.

First recall the basics about positive matrices. A matrix $A \in \mathcal{M}_d(C)$ is called *positive semi-definite* if $\langle\psi|A|\psi\rangle \geq 0$ for all $\psi \in \mathbb{C}^d$ and *positive definite* or *strictly positive* if the inequality is strict (i.e., '> 0' holds). Being lazy we will occasionally just write 'positive' when meaning 'positive semi-definite'. This is equivalent to saying that A is Hermitian and has (strictly) positive eigenvalues and we will write $A \geq 0$ ($A > 0$). Recall that $A \geq 0$ is in turn equivalent to $A = B^\dagger B$ for some B and that $A \geq 0 \Rightarrow X^\dagger A X \geq 0$. An order relation for Hermitian matrices A, B is given by

$$A \geq B \quad \Leftrightarrow \quad A - B \geq 0. \tag{5.1}$$

This defines a partial order in the sense that two Hermitian matrices may be incomparable. If, for instance a Hermitian A has positive and negative eigenvalues, then neither $A \geq 0$ nor $A \leq 0$ holds. This is reminiscent to the ordering of events in Minkowski space, and for $\mathcal{M}_2(\mathbb{C})$ it is actually equivalent (see example 5.1).

Note that by the defining equation (5.1) a linear map $T : \mathcal{M}_d \rightarrow \mathcal{M}_{d'}$ is order preserving iff it is positive.

The above operator ordering leads to a peculiar structure: it makes the set of Hermitian matrices an *anti-lattice*. A partially ordered set is called a *lattice* if for every two elements, say A and B, there is an *infimum* (a largest lower bound, typically denoted $A \wedge B$) and a *supremum* (least upper bound, $A \vee B$). The set of Hermitian matrices in $\mathcal{M}_d$ ($d > 1$) is not a lattice (unlike the set of Hermitian projections; see example 5.2). It has the property that if two elements happen to have a supremum or infimum, than this coincides with one of the elements—this structure is called an 'anti-lattice'.

Before we come to the main operator inequalities involving quantum channels (or more generally positive maps) we collect some very useful basic results:

Theorem 5.1 (Douglas' Theorem) [1] *Let $A, B \in \mathcal{M}_d(\mathbb{C})$. Then the following are equivalent:*

1. $\mathrm{range}(A) \subseteq \mathrm{range}(B)$,
2. $AA^\dagger \leq \mu BB^\dagger$ *for some* $\mu \geq 0$,
3. $A = BC$ *for some* $C \in \mathcal{M}_d(\mathbb{C})$.

Moreover, if this is valid, then there is a unique C for which (i) $\ker(A) \subseteq \ker(C)$, *(ii)* $\mathrm{range}(C) \subseteq \mathrm{range}(B^\dagger)$ *and (iii)* $||C||_\infty^2 = \min\{\mu | AA^\dagger \leq \mu BB^\dagger\}$.

Theorem 5.2 (Block matrices and Schur complements) *Let $A \in \mathcal{M}_d(\mathbb{C})$ and $B \in \mathcal{M}_{d'}(\mathbb{C})$ be two positive semi-definite matrices and $C \in \mathcal{M}_{d,d'}(\mathbb{C})$. Then the following are equivalent:*

1. $\begin{pmatrix} A & C \\ C^\dagger & B \end{pmatrix} \geq 0$,
2. $\ker(B) \subseteq \ker(C)$ *and* $A \geq CB^{-1}C^\dagger$, *where B^{-1} is the pseudo-inverse (i.e., inverse on the support) when necessary,*
3. $\ker(B) \subseteq \ker(C)$ *and* $||A^{-1/2}CB^{-1/2}||_\infty \leq 1$.

Example 5.1 (Ordering in Minkowski space) *Every Hermitian 2×2 matrix can be expanded as $M = \sum_{i=0}^3 x_i \sigma_i$ with $x \in \mathbb{R}^4$. Since $\det(M) = x_0^2 - x_1^2 - x_2^2 - x_3^2 = \langle x, \eta x\rangle$ with $\eta := \mathrm{diag}(1,-1,-1,-1)$ we can identify x with the coordinates of an event in* Minkowski space, *so that the determinant provides the Minkowski metric. As discussed in example 2.2 Lorentz transformation then take on the form $M \mapsto XMX^\dagger$ with $X \in SL(2,\mathbb{C})$.*

The ordering of events in Minkowski space, i.e., the causal structure imposed by the forward and backward light cones, then becomes the operator ordering on the level of the matrix representation. Recall that in Minkowski space two elements x and x' are time-like *(i.e., potentially causally related) if $\langle x - x', \eta(x - x')\rangle$ is positive, and* space-like *(i.e., incomparable) if it is negative. The ordering between two time-like related events is then given by $(x - x')_0$, i.e., the sign of the time-component and x is in the forward/backward light cone of x' if this component is positive/negative.*

On the level of the matrix representation $\det(M - M')$ *tells us whether the matrices are comparable in matrix order and $M - M' \geq 0$ if and only iff* $\det(M - M') \geq 0$ *and* $\mathrm{tr}[M - M'] \geq 0$. *In other words a Hermitian 2×2 matrix is positive iff its determinant and trace are.*

In Minkowski space the absence of the lattice property means that for two space-like separated events there is no 'latest' event in the intersection of their backward light cones.

Example 5.2 (Lattice of projections) ...

[1] which is typically referred to as 'Douglas' Lemma'.

5.2 Schwarz inequalities

Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a completely positive, trace preserving map and T^* its dual (description in the Heisenberg picture). Then

$$\forall A \in \mathcal{M}_{d'} : \quad T^*(A^\dagger)T^*(A) \leq T^*(A^\dagger A), \tag{5.2}$$

is an operator version of the *Schwarz inequality*. Inequality (5.2) can easily be proven by employing Stinespring's representation $T^*(A) = V(A \otimes \mathbb{1})V^\dagger$ (with V an isometry, see Thm. 2.2) so that it can be written

$$V(A^\dagger \otimes \mathbb{1})\underbrace{V^\dagger V}_{\leq \mathbb{1}}(A \otimes \mathbb{1})V^\dagger \leq V(A^\dagger A \otimes \mathbb{1})V^\dagger. \tag{5.3}$$

The following relaxes the assumptions on the map T and characterizes the cases of equality:

Theorem 5.3 (Operator Schwarz inequality) *Let $T^* : \mathcal{M}_{d'} \to \mathcal{M}_d$ be a positive linear map. If T is 2-positive, then for all $A, B \in \mathcal{M}_{d'}$:*

$$T^*(A^\dagger B)T^*(B^\dagger B)^{-1}T^*(B^\dagger A) \leq T^*(A^\dagger A), \tag{5.4}$$

where the inverse is taken on the range.

Proof Define a rectangular matrix $C := (A, B)$ so that $C^\dagger C \in \mathcal{M}_{2d'}$ is positive semidefinite. Then, due to 2-positivity, $(\mathrm{id}_2 \otimes T)(C^\dagger C) \geq 0$ which by Thm.5.2 implies Eq.(5.4). □

Theorem 5.4 (Equality in the operator Schwarz inequality) *Let $T^* : \mathcal{M}_{d'} \to \mathcal{M}_d$ be a positive linear map such that for a given $B \in \mathcal{M}_{d'}$ the Schwarz inequality in Eq.(5.4) holds for all $A \in \mathcal{M}_{d'}$. Denote by $\mathcal{A}_B \subseteq \mathcal{M}_{d'}$ the set of A's for which equality is attained in Eq.(5.4) for the given B. Then for all $A \in \mathcal{A}_B$ we have*

$$T^*(A^\dagger B)T^*(B^\dagger B)^{-1}T^*(B^\dagger X) = T^*(A^\dagger X), \quad \forall X \in \mathcal{M}_{d'}. \tag{5.5}$$

Proof Assume that $A \in \mathcal{A}_B$. Applying Eq.(5.4) (where we put everything to the r.h.s.) to B and $\tilde{A} = tA + X$ with $t \in \mathbb{R}$, we obtain

$$\begin{aligned} 0 \quad \leq \quad & t\Big[T^*(A^\dagger X) + T^*(X^\dagger A) - T^*(A^\dagger B)T^*(B^\dagger B)^{-1}T^*(B^\dagger X) \\ & \qquad - T^*(X^\dagger B)T^*(B^\dagger B)^{-1}T^*(B^\dagger A)\Big] \\ & + T^*(X^\dagger X) - T^*(X^\dagger B)T^*(B^\dagger B)^{-1}T^*(B^\dagger X), \end{aligned} \tag{5.6}$$

as the quadratic order in t vanishes due to $A \in \mathcal{A}_B$. In order for this inequality to hold for all $t \in \mathbb{R}$ the term linear in t has to vanish. By applying this argument to both X and iX and taking the sum we obtain Eq.(5.5). □

Generalizations of the above inequalities to the case of block matrices can be easily obtained by setting $A = \sum_{i=1}^n A_i \otimes \langle i| \in \mathcal{M}_{d',nd'}$ and $T = \tilde{T} \otimes \mathrm{id}_n$ where $\tilde{T}$ has then to be $n+1$-positive.

Positivity alone is not sufficient for a linear map to satisfy the Schwarz inequality (5.2); the simplest counterexample being matrix transposition with $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. However, there are notable exceptions:

Proposition 5.1 (Schwarz inequality for commutative domains) *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a positive linear map which is trace non-increasing (i.e., $T^*(\mathbb{1}) \leq \mathbb{1}$). Then the Schwarz-inequality (5.2) holds for all normal operators $A \in \mathcal{M}_{d'}$.*[2]

Proof Take an arbitrary normal A. Then $[A, A^\dagger] = [A, A^\dagger A] = 0$ so that by Prop.1.6 T^* is completely positive when restricted to the domain $\mathcal{D} := \mathrm{span}\{A, A^\dagger, A^\dagger A, \mathbb{1}\}$. Following Prop.1.7 we can extend this to a completely positive map on $\mathcal{M}_{d'}$ since $\mathcal{D}$ is an operator space. The assertion then follows from the fact that (i) the extension coincides with T^* on $\mathcal{D}$ and (ii) the Schwarz inequality holds for completely positive maps. □

Prop.5.1 is the basis of the following generalizations. For the first one recall that an operator $A \in \mathcal{B}(\mathcal{H})$ is called *subnormal* if there is a Hilbert space $\mathcal{H}' = \mathcal{H} \oplus \mathcal{H}^\perp$ and some B, C so that $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \in \mathcal{B}(\mathcal{H}')$ is a normal operator. An alternative characterization of subnormal operators is that for all $n \in \mathbb{N}$ it holds that $\sum_{k,l=0}^n |l\rangle\langle k| \otimes A^{\dagger k} A^l \geq 0$ with $A^0 := \mathbb{1}$. An operator is called *quasinormal* iff $[A^\dagger A, A] = 0$ and the following relation between the mentioned sets holds:

$$\text{normal} \subset \text{quasinormal} \subset \text{subnormal}.$$

Theorem 5.5 (Schwarz inequality for subnormal operators) *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a positive linear map which is trace non-increasing (i.e., $T^*(\mathbb{1}) \leq \mathbb{1}$) and $A \in \mathcal{M}_{d'}$ a subnormal operator. Then*

$$T^*(A^\dagger)T^*(A) \leq T^*(A^\dagger A) \text{ and } T^*(A)T^*(A^\dagger) \leq T^*(A^\dagger A). \tag{5.7}$$

Proof Denote by $N \in \mathcal{B}(\mathcal{H}')$ the normal extension of A and by $\theta : \mathcal{B}(\mathcal{H}') \to \mathcal{B}(\mathcal{H})$ the unital and completely positive linear map which maps $\theta(N) = A$. That is, θ returns the North-West block of its argument. Then, by Prop.5.1

$$\tilde{T}(N^\dagger N) \geq \tilde{T}(N)\tilde{T}(N^\dagger), \tilde{T}(N^\dagger)\tilde{T}(N),$$

holds for the concatenated map $\tilde{T} := T^*\theta$ (and the second inequality follows from the first by $N \to N^\dagger$ and $NN^\dagger = N^\dagger N$). Writing things out in terms of T and A yields the claimed inequalities. □

Another possibility of generalizing the Schwarz inequality to beyond 2-positive maps is replacing the $A^\dagger A$ term by something which is larger and commutes with A:

[2] A is *normal* if it commutes with its adjoint. So in particular Hermitian and unitary operators are normal.

Theorem 5.6 (Schwarz inequality for commuting dominant operators) *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a positive linear map which is trace non-increasing (i.e., $T^*(\mathbb{1}) \leq \mathbb{1}$) and $A \in \mathcal{M}_{d'}$. Then for every positive $D \in \mathcal{M}_{d'}$ which satisfies $[D, A] = 0$ and $D \geq A^\dagger A$ it holds that*

$$T^*(A^\dagger)T^*(A) \leq T^*(D) \text{ and } T^*(A)T^*(A^\dagger) \leq T^*(D). \tag{5.8}$$

Proof First not that $D \geq AA^\dagger$. In order to see this, assume for the moment that D is invertible. Then $D \geq A^\dagger A$ implies that $\mathbb{1} \geq X^\dagger X$ with $X := AD^{-1/2}$. This in turn implies $\mathbb{1} \geq XX^\dagger$ from which $D \geq AA^\dagger$ follows by using $[D, A] = 0$. This generalizes to non-invertible D's if we first replace $D \to D + \epsilon\mathbb{1}$ in the argument and then let $\epsilon \to 0$.

Now construct the block matrix

$$N := \begin{pmatrix} A & \sqrt{D - AA^\dagger} \\ \sqrt{D - A^\dagger A} & -A^\dagger \end{pmatrix}.$$

Direct calculation (and using that $A\sqrt{D - A^\dagger A} = \sqrt{D - AA^\dagger}A$) shows that N is a normal operator. Analogous to the proof of Thm.5.5 we can now introduce a unital and completely positive linear map $\theta : N \to A$ and show the desired inequalities by using Prop.5.1 in order to get $\tilde{T}(N^\dagger N) \geq \tilde{T}(N^\dagger)\tilde{T}(N), \tilde{T}(N)\tilde{T}(N^\dagger)$ for the map $\tilde{T} := T^*\theta$. Using that $\theta(N^\dagger N) = D$ and $\theta(N) = A$ concludes the proof. □

One might wonder whether a unital positive map T^* which satisfies the Schwarz inequality in Eq.(5.2) is automatically 2-positive. The following example shows that this is not the case. Since the set of unital maps which satisfy Eq.(5.2) is therefore larger than the subset of 2-positive maps but smaller than the set of positive maps it may deserve an own name: *Schwarz maps* is the most commonly used; $\frac{3}{2}$-*positive maps* a not too serious alternative.

Example 5.3 (Schwarz maps which are not 2-positive) *Consider the map $T : \mathcal{M}_2 \to \mathcal{M}_2$ defined via*

$$T^*(A) := \frac{1}{2}\Big(A^T + \frac{1}{2}\mathrm{tr}\,[A]\,\mathbb{1}\Big). \tag{5.9}$$

As a convex combination of two positive maps it is positive. However, it fails to be 2-positive (which in this case means completely positive) since the Choi-Jamiolkowski operator of T^* is $(\mathcal{F} + \frac{1}{2}\mathbb{1})/4$ which has one eigenvalue $-1/8$. Nevertheless, the map satisfies the Schwarz inequality in Eq.(5.2). In order to see this define $F(A) := T^*(A^\dagger A) - T^*(A^\dagger)T^*(A)$ and note that $F(A) = F(A + \lambda\mathbb{1})$ for all $\lambda \in \mathbb{C}$ and $A \in \mathcal{M}_2$. So for proving that $F(A) \geq 0$ for all A we can w.l.o.g. assume that $\mathrm{tr}\,[A] = 0$. In this case

$$\begin{aligned} F(A) &= \frac{1}{4}\Big(2(A^\dagger A)^T + \mathrm{tr}\left[A^\dagger A\right]\mathbb{1} - \bar{A}A^T\Big) \\ &\geq \frac{1}{2}(A^\dagger A)^T \ \geq 0, \end{aligned} \tag{5.10}$$

since $\mathrm{tr}\left[A^\dagger A\right]\mathbb{1} = \mathrm{tr}\left[\bar{A}A^T\right]\mathbb{1} \geq \bar{A}A^T$.

By Thm.5.3 every 2-positive linear map T which is trace non-increasing (i.e., $T^*(\mathbb{1}) \leq \mathbb{1}$) satisfies the Schwarz inequality (5.2). The cases of equality in this inequality lead us to the sets

$$\mathcal{A}^R := \{A \in \mathcal{M}_{d'} \,|\, T^*(XA) = T^*(X)T^*(A),\ \forall X \in \mathcal{M}_{d'}\}, \tag{5.11}$$
$$\mathcal{A}^L := \{A \in \mathcal{M}_{d'} \,|\, T^*(AX) = T^*(A)T^*(X),\ \forall X \in \mathcal{M}_{d'}\}, \tag{5.12}$$

called the *right/left multiplicative domain* of T^*, and $\mathcal{A} := \mathcal{A}^R \cap \mathcal{A}^L$ its *multiplicative domain*.

Note that these three sets form each an algebra (i.e., they are in particular closed under multiplication) and that T^* acts as a homomorphism on these algebras. Since the multiplicative domain $\mathcal{A}$ is in addition closed under taking the adjoint, it is a $*$-algebra and T^* a $*$-homomorphism when restricted to $\mathcal{A}$.

Theorem 5.7 (Multiplicative domains) *Let T be a linear map which fulfills the Schwarz inequality (5.2). Then*

$$\{A \in \mathcal{M}_{d'} \,|\, T^*(A^\dagger A) = T^*(A^\dagger)T^*(A)\} = \mathcal{A}^R, \tag{5.13}$$
$$\{A \in \mathcal{M}_{d'} \,|\, T^*(AA^\dagger) = T^*(A)T^*(A^\dagger)\} = \mathcal{A}^L. \tag{5.14}$$

The proof of this statement is analogous to that of Thm.5.4. The fact that the set of operators achieving equality in the Schwarz inequality forms an algebra has remarkable consequences when it comes to the discussion of fixed points of quantum channels.

5.3 Operator convexity and monotonicity

Functional calculus gives a meaning to scalar functions evaluated on operators. Suppose a matrix $A \in \mathcal{M}_d(\mathbb{C})$ can be diagonalized as $A = UDU^{-1}$ and a given function f is defined on a domain including the eigenvalues of A (i.e., the entries of the diagonal matrix D). Then a consistent way of defining f applied to A is

$$f(A) = U\ f(D)\ U^{-1} \quad \text{where} \quad [f(D)]_{ii} = f(D_{ii}). \tag{5.15}$$

As diagonalizable matrices are dense, there is at most one continuous way of extending the definition of $f(A)$ to all $A \in \mathcal{M}_d(\mathbb{C})$. The core of this is the observation that for diagonalizable matrices Eq.(5.15) implies that $f(A) = g(A)$ for every polynomial g which coincides with f on the spectrum of A. In general one can show that replacing f by a such a polynomial g gives a well-defined meaning to $f(A)$ if the derivatives of f and g coincide up to order $j-1$ at every eigenvalue of A with Jordan-block of size j.

As we are only considering Hermitian matrices in this chapter we can, however, use Eq.(5.15) and take $U^{-1} = U^\dagger$ unitary. The above definition implies that $f(VAV^\dagger) = Vf(A)V^\dagger$ for every unitary V. Moreover, $f(A \oplus B) = f(A) \oplus f(B)$ and in particular $f(A \otimes \mathbb{1}) = f(A) \otimes \mathbb{1}$. Be aware, that in general $f(A \otimes B) \neq f(A) \otimes f(B)$ and $f(VAV^\dagger) \neq Vf(A)V^\dagger$ if V is merely an isometry.

Having defined functions on operators and Löwner's operator ordering in mind we may now ask how properties of functions like monotonicity or convexity translate to the world of matrices and operators.

Operator monotonicity and operator convexity . Recall that a real-valued function f is called *convex* on an interval $I \subseteq \mathbb{R}$ if for all $\lambda \in [0,1]$ and all $a, b \in I$ we have

$$f\big(\lambda a + (1-\lambda)b\big) \leq \lambda f(a) + (1-\lambda) f(b), \tag{5.16}$$

i.e., the second derivative (if it exists) should be positive on I. For continuous functions this is equivalent to *mid-point convexity*, which is the same statement with $\lambda = 1/2$. f is called *concave* if $-f$ is convex, and it is called *monotone* on I if for all $a, b \in I$

$$a \geq b \quad \Rightarrow \quad f(a) \geq f(b). \tag{5.17}$$

If Eqs.(5.16,5.17) hold for all Hermitian matrices $A, B \in \mathcal{M}_d(\mathbb{C})$ whose spectrum is contained in I, then we say that f is *matrix convex* or *matrix monotone* of order d on I. Functions which are convex/monotone on scalars need not be so on matrices (see example 5.4). In fact, the set of matrix convex/monotone functions becomes smaller and smaller when we increase d. Functions which remain matrix convex/monotone for all $d \in \mathbb{N}$ are called *operator convex* or *operator monotone*, respectively.

Note that the sets of operator monotone functions and operator convex functions (on a given interval) are both convex cones which are closed under pointwise limits. So, characterizing all such functions essentially boils down to identifying the extreme rays of these sets and then allowing for positive linear combinations. The two following theorems do this for the positive half-line (for other intervals analogous results hold).

Theorem 5.8 (Operator convexity) *A function $f : [0, \infty) \to \mathbb{R}$ is operator convex iff it can be written as*

$$f(x) = a + bx + cx^2 + \int_0^\infty \frac{yx^2}{y+x}\, d\mu(y), \tag{5.18}$$

where $a, b \in \mathbb{R}$, $c \geq 0$ and μ is a positive finite measure.

Theorem 5.9 (Löwner's theorem—operator monotonicity) *For a continuous function $f : (0, \infty) \to \mathbb{R}$ the following statements are equivalent:*

1. *f is operator monotone,*
2. *f has an analytic continuation to the entire domain $\mathrm{Im} z \neq 0$ such that it maps the upper half-plane $\{z \in \mathbb{C} | \mathrm{Im} z > 0\}$ into itself,*
3. *f has an integral representation of the form*

$$f(x) = a + bx + \int_{-\infty}^0 \frac{1+xy}{y-x} d\mu(y), \tag{5.19}$$

where $a \in \mathbb{R}$, $\beta \geq 0$ and μ is a finite positive measure.

In particular the relation between operator monotone functions and so-called *Pick functions* (holomorphic functions with positive imaginary part) is a very powerful result. It enables us to 'see' right away that certain functions (like the exponential function) are not operator monotone. The integral representations of operator convex and operator monotone functions seem somehow related and, in fact, there are close relations between these two classes. For instance, a continuous function $f : [0,\infty) \to [0,\infty)$ is operator monotone iff it is operator concave.

Example 5.4 (Operator monotone/convex functions) *The following lists some functions (in the variable $x \in \mathbb{R}$) and their properties regarding operator convexity/monotonicity on specified intervals $I \subseteq \mathbb{R}$:*

operator monotone	*operator convex*
∘ $a+bx$ *on* $\mathbb{R}$ *for* $a \in \mathbb{R}$, $b \geq 0$	∘ $a+bx+cx^2$ *on* $\mathbb{R}$ *for* $a,b \in \mathbb{R}$, $c \geq 0$
∘ x^p *on* $[0,\infty)$ *for* $p \in [0,1]$	∘ x^p *on* $(0,\infty)$ *for* $p \in [1,2]$
∘ $-x^p$ *on* $(0,\infty)$ *for* $p \in [-1,0]$	∘ x^p *on* $(0,\infty)$ *for* $p \in [-1,0]$
∘ $\log x$ *on* $(0,\infty)$	∘ $-\log x$ *on* $(0,\infty)$
∘ $(x\log x)/(x-1)$ *on* $(0,\infty)$	∘ $x \log x$ *on* $(0,\infty)$
∘ $\frac{x}{y+x}$ *on* $(0,\infty)$ *for* $y \geq 0$	∘ $\frac{x^2}{y+x}$ *on* $[0,\infty)$ *for* $y \geq 0$
∘ $\tan x$ *on* $(-\pi/2,\pi/2)$	∘ $f:(0,\infty)\to(0,\infty)$ *if* $x \mapsto 1/f(x)$ *is op. monotone*
not operator monotone	*not operator convex*
∘ x^p *on* $(0,\infty)$ *for* $p \notin [0,1]$	∘ x^p *on* $(0,\infty)$ *for* $p \notin [-1,0]\cup[1,2]$
∘ e^x *on any non-trivial* I	∘ e^x *on any non-trivial* I
∘ $\max(0,x)$ *on any* $I \supset \{0\}$	∘ $\lvert x\rvert$ *on any* $I \supset \{0\}$

For 'simple' functions which are not operator monotone/convex there are similarly 'simple' counterexamples. Take for instance

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{5.20}$$

When applied to them we see that $f(x) = x^2$ is not operator monotone on $[0,\infty)$ and that $f(x) = x^3$ fails to be operator convex on $[0,\infty)$.

We will now connect first operator convexity and then operator monotonicity to positive maps. Recall that positive maps are exactly the ones which preserve the order, i.e., if $A \geq B$, then $T(A) \geq T(B)$ for positive T. Conversely, if this holds for all A,B, then T has to be positive.

A typical requirement in the context addressed below is that the positive map satisfies $T(\mathbb{1}) \leq \mathbb{1}$. This implies that T is contractive in the following sense: consider a Hermitian A with $\mathrm{spec}(A) \subseteq [a,b]$, where $[a,b]$ is an interval containing zero. Then $T(\mathbb{1}) \leq \mathbb{1}$ implies that also $\mathrm{spec}(T(A)) \subseteq [a,b]$ since

$$a\mathbb{1} \leq T(a\mathbb{1}) \leq T(A) \leq T(b\mathbb{1}) \leq b\mathbb{1}. \tag{5.21}$$

Theorem 5.10 (Operator convexity from projection inequality) *Consider a real-valued function f defined on some interval $I \subseteq \mathbb{R}$ with $0 \in I$. If for all*

$d \in \mathbb{N}$, all Hermitian projections $P \in \mathcal{M}_d(\mathbb{C})$ and all Hermitian $A \in \mathcal{M}_d(\mathbb{C})$ with $\text{spec}(A) \subset I$

$$f(PAP) \leq Pf(A)P, \qquad \textit{then} \tag{5.22}$$

1. *$f(0) \leq 0$,*

2. *f is operator convex on I,*

3. *for all positive linear maps $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ with $T(\mathbb{1}) \leq \mathbb{1}$ and all Hermitian A with spectrum in I:*

$$f\big(T(A)\big) \leq T\big(f(A)\big). \tag{5.23}$$

PROOF 1. follows from Eq.(5.22) already by setting $A = P^\perp := \mathbb{1} - P$. In order to show 3. consider first a completely positive and unital map T. Using Stinespring's theorem (Thm.2.2) and expressing the isometry as a unitary times a projection we can write

$$T(A) \oplus 0_{dr-d'} = PU^\dagger(A \otimes \mathbb{1}_r)UP. \tag{5.24}$$

When applying f to both sides of this inequality we can use $f(U^\dagger \cdot U) = U^\dagger f(\cdot)U$, and $f(A \otimes \mathbb{1}) = f(A) \otimes \mathbb{1}$ so that the Eq.(5.23) indeed follows from (5.22) for unital and completely positive T. Now let us successively get rid of the additional constraints. Assume T is not unital. Then we can define a unital map $T' : \mathcal{M}_d(\mathbb{C}) \oplus \mathbb{C} \to \mathcal{M}_{d'}(\mathbb{C})$ by $T' : A \oplus b \mapsto T(A) + b(\mathbb{1} - T(\mathbb{1}))$ (and off-diagonals are mapped to zero). T' is completely positive iff T is completely positive and $T(\mathbb{1}) \leq \mathbb{1}$. Applying Eq.(5.23) to T' then yields the same inequality for T, so unitality can be replaced by $T(\mathbb{1}) \leq \mathbb{1}$. If T is merely positive we can use that A and $f(A)$ are elements of an abelian C^*-algebra (the one generated by A and $\mathbb{1}$) on which T is completely positive (Prop.1.6) with a completely positive extension on $\mathcal{M}_d$ (Prop.1.7). This completes the proof of 3.

Finally, 2. follows by applying Eq.(5.23) to the map $T : A \oplus B \to \lambda A + (1-\lambda)B$, $\lambda \in [0, 1]$. □

This theorem becomes yet more interesting as the converse also holds:

Theorem 5.11 (Projection inequality from operator convexity) *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be an operator convex function on an interval $I \ni 0$. Then for all Hermitian projections $P \in \mathcal{M}_d(\mathbb{C})$ (with $P^\perp := \mathbb{1} - P$) and all Hermitian $A \in \mathcal{M}_d(\mathbb{C})$ with* $\text{spec}(A) \subset I$*:*

$$f(PAP) \leq Pf(A)P + P^\perp f(0)P^\perp. \tag{5.25}$$

Consequently, if $f(0) \leq 0$, then for all positive maps T with $T(\mathbb{1}) \leq \mathbb{1}$:

$$f\big(T(A)\big) \leq T\big(f(A)\big). \tag{5.26}$$

Proof Define a unitary $U := P - P^\perp$. Then

$$
\begin{aligned}
P\, f(PAP)\, P &= P\, f(PAP + P^\perp A P^\perp)\, P &(5.27)\\
&= P\, f\big(A/2 + UAU^\dagger/2\big)\, P &(5.28)\\
&\leq P\Big(f(A)/2 + U f(A) U^\dagger/2\Big) P = P f(A) P, &(5.29)
\end{aligned}
$$

where the inequality is the assumed operator convexity. Eq.(5.25) then follows from $f(PAP) = Pf(PAP)P + P^\perp f(0) P^\perp$ and Eq.(5.26) is just Eq.(5.23) since Thm.5.10 applies if $f(0) \leq 0$. □

The following shows that we can drop the restriction $f(0) \leq 0$ if T is normalized, i.e., unital:

Theorem 5.12 (Operator convexity and unital positive maps) *Let $f : \mathbb{R} \to \mathbb{R}$ be operator convex on $[0, \infty)$ and $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ any positive unital map. Then for all $A \geq 0$:*

$$f\big(T(A)\big) \leq T\big(f(A)\big). \tag{5.30}$$

Proof By the integral representation in Thm.5.8 it suffices to prove Eq.(5.30) for $f(x) = x^2$ and $f(x) = x^2/(y+x)$, $y > 0$. The first case is nothing but the Schwarz inequality in Eq.(5.2). The second case can be simplified by rewriting it as $f(x) = x - y + y^2/(x+y)$. Inserting this into Eq.(5.30) and exploiting linearity and unitality of T we see that proving Eq.(5.30) boils down to showing that

$$T(B)^{-1} \leq T(B^{-1}), \tag{5.31}$$

for all $B > 0$. In order to show Eq.(5.31) note that by introducing the spectral decomposition $B = \sum_i b_i P_i$ (with spectral projections P_i) and using that $\mathbb{1} = T(\mathbb{1}) = \sum_i T(P_i)$ we get

$$\begin{pmatrix} T(B^{-1}) & \mathbb{1} \\ \mathbb{1} & T(B) \end{pmatrix} = \sum_i \begin{pmatrix} b_i^{-1} & 1 \\ 1 & b_i \end{pmatrix} \otimes T(P_i) \geq 0. \tag{5.32}$$

From here Eq.(5.31) follows by invoking the condition for the positivity of block matrices in Thm.5.2. □

By applying this to $f(x) = x \log x$, which is operator convex on the positive half-line, we obtain:

Corollary 5.1 (Entropic operator inequality) *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a positive unital map. Then for all $\rho \geq 0$:*

$$T(\rho) \log T(\rho) \leq T(\rho \log \rho). \tag{5.33}$$

Problem 10 *Provide a generalization of the inequality for positive maps with $T(\leq \mathbb{1})$ in the form of Eq.(5.30) for arbitrary operator convex f (not necessarily satisfying $f(0) \leq 0$).*

Let us now turn from convexity to monotonicity:

Theorem 5.13 (Operator monotonicity and positive maps) *Let f be an operator monotone function on an interval $I = [0, a]$ with $f(0) \geq 0$ and $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ any positive map for which $T(\mathbb{1}) \leq \mathbb{1}$. Then for all $A = A^\dagger$ with* spec$(A) \in I$*:*

$$T\big(f(A)\big) \leq f\big(T(A)\big). \tag{5.34}$$

Proof Let us first prove the statement for $T(A) = X^\dagger A X$ with $||X||_\infty \leq 1$. To this end define a unitary block matrix

$$U := \begin{pmatrix} X & \sqrt{\mathbb{1} - XX^\dagger} \\ \sqrt{\mathbb{1} - X^\dagger X} & -X^\dagger \end{pmatrix}, \tag{5.35}$$

and note that the upper left block of $U^\dagger(A \oplus 0)U$ is $X^\dagger A X$. Hence by Thm.5.2 we can find for every $\epsilon > 0$ a $\delta > 0$ such that $U^\dagger(A \oplus 0)U \leq (X^\dagger A X + \epsilon\mathbb{1}) \oplus (\delta\mathbb{1})$. Applying f to this inequality, using its monotonicity and that $f(U^\dagger \cdot U) = U^\dagger f(\cdot) U$ we get

$$X^\dagger f(A) X + \sqrt{\mathbb{1} - XX^\dagger} f(0) \sqrt{\mathbb{1} - XX^\dagger} \leq f\big(X^\dagger A X + \epsilon\mathbb{1}\big) \tag{5.36}$$

by comparing the upper left blocks on both sides. The asserted inequality (5.34) follows then from continuity of operator monotone functions by taking $\epsilon \to 0$ together with $f(0) \geq 0$.

This can be extended to all completely positive maps with $T(\mathbb{1}) \leq \mathbb{1}$ by Stinespring's theorem. Following Thms.2.2,2.3 we can represent every such map in the form $T(A) = X^\dagger(A \otimes \mathbb{1})X$, where $X^\dagger X = T(\mathbb{1}) \leq \mathbb{1}$. By invoking $f(A \otimes \mathbb{1}) = f(A) \otimes \mathbb{1}$ we can thus generalize the inequality to contractive completely positive maps. Finally, if T is merely positive with $T(\mathbb{1}) \leq \mathbb{1}$, then it is completely positive on the abelian algebra generated by $\mathbb{1}$ and A (Prop.1.6) and admits a completely positive extension on $\mathcal{M}_d$ (Prop.1.7). □

Applying this to some of the examples in example 5.4 leads to:

Corollary 5.2 *Let $T : \mathcal{M}_d \to \mathcal{M}_{d'}$ be a positive map with $T(\mathbb{1}) \leq \mathbb{1}$. Then*

1. $T(A) \leq T\big(A^p\big)^{1/p}$ *for all $p \geq 1$ and $A \geq 0$,*
2. $T(A) \geq T\big(A^p\big)^{1/p}$ *for all $A > 0$ and*[3] $p \in [1/2, 1]$,
3. $T\big(\log(A)\big) \leq \log\big(T(A)\big)$ *for all $A > 0$.*

Problem 11 *Prove a counterpart of Thm.5.12 for matrix convex functions of finite order. In other words, find a proof which does not use the integral representation (or any other representation implying operator convexity).*

Let us finally prove a useful implication of the above theorems:

Corollary 5.3 *Let $\{T_i : \mathcal{M}_d \to \mathcal{M}_{d'}\}_{i=1,\dots,n}$ be a collection of positive linear maps such that $\sum_i T_i(\mathbb{1}) \leq \mathbb{1}$.*

[3] This follows from Thm.5.11 and operator convexity of $x \mapsto x^{1/p}$ for $p \in [1/2, 1]$.

1. *If f is operator monotone on an interval $I \ni 0$ with $f(0) \geq 0$, then for all Hermitian operators A_i with spectrum in I:*

$$\sum_{i=1}^{n} T_i\big(f(A_i)\big) \leq f\Big(\sum_{i=1}^{n} T_i(A_i)\Big). \tag{5.37}$$

2. *If f is operator convex on an interval $I \ni 0$ and $f(0) \leq 0$, then for all Hermitian operators A_i with spectrum in I:*

$$\sum_{i=1}^{n} T_i\big(f(A_i)\big) \geq f\Big(\sum_{i=1}^{n} T_i(A_i)\Big). \tag{5.38}$$

Proof Let us construct a new map $T : \mathcal{M}_d \otimes \mathcal{M}_n \rightarrow \mathcal{M}_{d'}$ via $T(A) := \sum_{i=1}^n T_i(V_i A V_i^\dagger)$ with isometries $V_i := \mathbb{1}_d \otimes \langle i|$. Similarly, define $A := \sum_i A_i \otimes |i\rangle\langle i|$. Then $T(A) = \sum_i T_i(A_i)$ and $f(A) = \sum_i f(A_i) \otimes |i\rangle\langle i|$. Since T is by construction such that $T(\mathbb{1}) \leq \mathbb{1}$ we can now apply Thm.5.11 and Thm.5.13 leading to assertions 2. and 1., respectively. □

5.4 Joint convexity

Joint convexity is a property of certain functions of several variables. A function $f : \mathcal{D} \subseteq \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ with domain $\mathcal{D}$ is called *jointly convex* in its arguments iff

$$f(a, b) \leq \lambda f(a_1, b_1) + (1 - \lambda) f(a_2, b_2) \tag{5.39}$$

holds for all a_i, b_i with $a = \lambda a_1 + (1-\lambda)a_2$, $b = \lambda b_1 + (1-\lambda)b_2$ and all $\lambda \in [0, 1]$. It is defined analogously for more than two arguments and f is called *jointly concave* iff $-f$ is jointly convex. Clearly, a jointly convex function is convex in each of its arguments separately (i.e., $f(a, b) \leq \lambda f(a_1, b) + (1 - \lambda) f(a_2, b)$); the converse is, however, not true in general.

In the following we will apply the concept of joint convexity to operators, i.e., we will essentially replace $\mathbb{R}$ with the space of Hermitian operators. Two related types of joint convexity results will be of interest: (i) joint operator convexity, where the ordering "$\leq$" in Eq.(5.39) refers to the partial order in the space of Hermitian operators (see Eq.(5.1)) and f is a mapping between operators, and (ii) joint convexity of functionals which map operators into $\mathbb{R}$ (in most cases the 'functional' will essentially be the trace).

Joint operator convexity We consider mappings of the form $F : \mathcal{D} \subseteq \mathcal{M}_{d_1} \times \cdots \times \mathcal{M}_{d_n} \rightarrow \mathcal{M}_d$ and say that they are *jointly operator convex* in the domain $\mathcal{D}$ iff

$$F(A) \leq \sum_i \lambda_i F\big(A^{(i)}\big), \tag{5.40}$$

for all $A, A^{(i)} \in \mathcal{D}$ which are related via $A = \sum_i \lambda_i A^{(i)}$ with λ being a vector of probabilities. The ordering '$\leq$' used in this context is the one between Hermitian operators which means in particular that F should have Hermitian image.

The following theorem encompasses most of the results about joint operator convexity. It provides a tool for constructing maps in several variables from maps in fewer variables while retaining the convexity properties. In particular, it allows us to construct mappings which are jointly operator concave/convex from simple operator concave/convex functions:

Theorem 5.14 (Joint operator convexity) *Let $g : \mathcal{D} \subseteq \mathcal{M}_{d_1} \times \cdots \times \mathcal{M}_{d_n} \to \mathcal{M}_d$ be a map on the direct product $\mathcal{D}$ of n positive operators, and similarly $h : \mathcal{D}' \subseteq \mathcal{M}_{d'_1} \times \cdots \times \mathcal{M}_{d'_m} \to \mathcal{M}_d$. Assume that g is jointly operator concave and positive and h is semi-definite. Let $I \ni 0$ be the positive/negative real half line depending on whether h is positive or negative semi-definite. For any function $f : I \to \mathbb{R}$ with $f(0) \leq 0$ define $F : \mathcal{D}' \times \mathcal{D} \to \mathcal{M}_d$ as*

$$F(L,R) := \sqrt{g(R)}\, f\Big(g(R)^{-1/2} h(L) g(R)^{-1/2}\Big)\sqrt{g(R)}. \tag{5.41}$$

We consider joint operator convexity of F in its $n+m$ arguments. F is jointly operator convex on positive operators for which g is invertible if at least one of the following holds:

1. *h is jointly operator concave and f is operator anti-monotone,*
2. *h is affine and f is operator convex.*

Proof Consider the convex decompositions $R = \sum_i \lambda_i R_i$ and $L = \sum_i \lambda_i L_i$ and define $K_i := \sqrt{\lambda_i g(R_i)} g(R)^{-1/2}$. This satisfies

$$\sum_i K_i^\dagger K_i = g(R)^{-1/2}\Big(\sum_i \lambda_i g(R_i)\Big) g(R)^{-1/2} \leq \mathbb{1}, \tag{5.42}$$

where the inequality comes from joint concavity of g. Assume that the requirements of the first assertion are fulfilled. Then if we insert the definitions:

$$\begin{aligned} F(L,R) &\leq \sqrt{g(R)}\, f\Big(\sum_i \lambda_i g(R)^{-1/2} h(L_i) g(R)^{-1/2}\Big)\sqrt{g(R)}, && (5.43)\\ &= \sqrt{g(R)}\, f\Big(\sum_i K_i^\dagger g(R_i)^{-1/2} h(L_i) g(R_i)^{-1/2} K_i\Big)\sqrt{g(R)}, && (5.44)\\ &\leq \sqrt{g(R)}\, \sum_i K_i^\dagger f\Big(g(R_i)^{-1/2} h(L_i) g(R_i)^{-1/2}\Big) K_i \sqrt{g(R)}, && (5.45)\\ &= \sum_i \lambda_i F(L_i, R_i), && (5.46) \end{aligned}$$

where the first inequality comes from concavity of h together with anti-monotonicity of f and the second inequality follows from the first part of Cor.5.3 by setting $T_i(\cdot) := K_i^\dagger \cdot K_i$ and noticing that Eq.(5.42) means $\sum_i T_i(\mathbb{1}) \leq \mathbb{1}$. For the second assertion of the theorem note that if h is affine we have equality in Eq.(5.43). The inequality in Eq.(5.45) follows then from the second part of Cor.5.3. □

Note that in the preceding theorem we may extend $\mathcal{D}'$ to all Hermitian operators and allow h to have Hermitian image if we set $I = \mathbb{R}$. Similarly, we can enlarge $\mathcal{D}$ as long as g has positive image. Applying Thm.5.14 to various functions now provides us with a zoo of jointly operator convex/concave mappings:

Corollary 5.4 *1. $\mathcal{M}_d \times \mathcal{M}_d \ni (L, R) \mapsto LR^{-\alpha}L$ is jointly operator convex for $\alpha \in [0, 1]$ on positive L and Hermitian R,*

2. $\mathcal{M}_d \times \mathcal{M}_d \ni (L, R) \mapsto R^{\beta/2}\big(R^{-\beta/2}L^\gamma R^{-\beta/2}\big)^\alpha R^{\beta/2}$ is jointly operator concave on positive operators for $\alpha, \beta, \gamma \in [0, 1]$,

3. $\mathcal{M}_d^m \times \mathcal{M}_d \ni (L, R) \mapsto R^{\beta/2}f\big(\sum_{i=1}^m R^{-\beta/2}L_i R^{-\beta/2}\big)R^{\beta/2}$, $\beta \in [0, 1]$ is jointly operator convex on positive operators if f is operator convex with $f(0) \leq 0$, e.g. $f(x) = -x^\alpha, f(x) = x^{1+\alpha}$ with $\alpha \in [0, 1]$ or $f(x) = x \log x$,

Proof Use Thm.5.14 and for 1. set $f(x) = x^2$, $h(x) = x$ and $g(x) = x^\alpha$. For 2. set $f(x) = -x^\alpha$, $g(x) = x^\beta$, $h(x) = x^\gamma$. For 3. set $h(L) = \sum_i L_i$ and $g(x) = x^\beta$. □

By replacing L, R in the above discussion by $L \to L \otimes \mathbb{1}$ and $R \to \mathbb{1} \otimes R$ and thereby making things commutative, we obtain similar statements for tensor product mappings:

Corollary 5.5 *1. $\mathcal{M}_d \times \mathcal{M}_d \ni (L, R) \mapsto L^x \otimes R^y$ is jointly operator concave on positive operators for $x, y \geq 0$ with $x + y \leq 1$.*

2. $\mathcal{M}_d^m \ni (A_1, \ldots, A_m) \mapsto \bigotimes A_i^{x_i}$ is jointly operator concave on positive operators for $x_i \geq 0$ with $\sum_i x_i \leq 1$.

3. $\mathcal{M}_d \times \mathcal{M}_d \ni (L, R) \mapsto L^{1+\alpha} \otimes R^{-\beta\alpha}$ is jointly operator convex on positive operators for $\alpha, \beta \in [0, 1]$.

Proof Statement 1. follows by inserting into the second assertion of Cor.5.4. Statement 2. follows by induction from the previous argument. That is, for $m = 3$ we continue with $h(A_1, A_2) = A_1^x \otimes A_2^x \otimes \mathbb{1}$, exploit that it is concave by the first statement and apply Thm.5.14 with $R = \mathbb{1} \otimes \mathbb{1} \otimes A_3$ and $f(x) = -x^\alpha$ and $g(x) = x^\beta$, $\alpha, \beta \in [0, 1]$, etc. Assertion 3. follows from point 3. of Cor.5.4 when applied to $f(x) = x^{1+\alpha}$. □

Jointly convex functionals In the following we will discuss joint convexity/concavity properties of functionals, i.e., maps from a set of operators into $\mathbb{R}$. In most cases the results follow from the previous ones on joint operator convexity by taking the trace or the expectation value w.r.t. some positive operator. One of the central results in this context is the following:

Theorem 5.15 (Ando-Lieb) *For every $K \in \mathcal{M}_d(\mathbb{C})$ the map*

$$\mathcal{M}_d(\mathbb{C}) \times \mathcal{M}_d(\mathbb{C}) \ni (A, B) \mapsto \operatorname{tr}\left[K^\dagger A^x K B^y\right] \tag{5.47}$$

- *is jointly concave on positive operators for $x, y \geq 0$ if $x + y \leq 1$, and*

- *jointly convex on positive operators for* $x \in [1,2]$ *if* $y \in [1-x, 0]$.

Proof The result is an immediate consequence of items 1. and 3. of Corr.5.5. In order to see this consider the expectation value of $L^x \otimes R^y$ w.r.t. $|\psi\rangle = (K \otimes \mathbb{1}) \sum_{i=1}^d |ii\rangle$. Using Eq.(1.12) this gives

$$\langle\psi|L^x \otimes R^y|\psi\rangle = \operatorname{tr}\left[K^\dagger L^x K (R^T)^y\right], \tag{5.48}$$

so that the result follows by setting $A = L$ and $B = R^T$. □

In fact, the parameter ranges for x and y in Thm.5.15 are the only ones (apart from interchanging $x \leftrightarrow y$) for which joint convexity/concavity holds.

Theorem 5.16 (Monotonicity under cp-maps) *Consider a functional* $F : \mathcal{D} \subseteq \mathcal{M}_d \times \cdots \times \mathcal{M}_d \to \mathbb{R}$ *which is defined for all dimensions* $d \in \mathbb{N}$. *Assume that* F *satisfies*

1. Joint convexity *in* $\mathcal{D}$,

2. Unitary invariance, *i.e., for all* $A \in \mathcal{D}$ *and all unitaries* $U \in \mathcal{M}_d(\mathbb{C})$ *it holds that* $F\big(UA_1U^\dagger, \ldots, UA_nU^\dagger\big) = F(A_1, \ldots, A_n)$,

3. Invariance under tensor products, *meaning that for all* $A \in \mathcal{D}$ *and all density operators* $\tau \in \mathcal{M}_{d'}(\mathbb{C})$ *we have* $F\big(A_1 \otimes \tau, \ldots, A_n \otimes \tau\big) = F(A_1, \ldots, A_n)$.

Then F *is monotone w.r.t. all completely positive trace-preserving maps* $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d''}(\mathbb{C})$ *in the sense that for all* $A \in \mathcal{D}$

$$F\big(T(A_1), \ldots, T(A_n)\big) \leq F(A_1, \ldots, A_n). \tag{5.49}$$

Proof Let us follow Thm.2.5 and represent T as $T(\rho) = \operatorname{tr}_E\big[U(\rho \otimes \tau)U^\dagger\big]$ where τ is a density matrix, U a unitary and tr_E the partial trace over an 'environmental' system of dimension m, say. If we take a unitary operator basis $\{V_i\}_{i=1,..,m^2}$ in $\mathcal{M}_m$ (orthonormal w.r.t. the Hilbert-Schmidt inner product–see Exp.2.1), we can write

$$T(\rho) \otimes \mathbb{1}_m/m = \frac{1}{m^2} \sum_{i=1}^{m^2} (\mathbb{1} \otimes V_i) U (\rho \otimes \tau) U^\dagger (\mathbb{1} \otimes V_i^\dagger). \tag{5.50}$$

Observing that $\mathbb{1}_m/m$ is again a density matrix and that the r.h.s. of Eq.(5.50) represents a convex combination of unitarily equivalent terms we can straightforwardly arrive at the desired Eq.(5.49) by applying joint convexity together with the invariance properties of F. □

5.5 Convexity and monotonicity under the trace

Trace inequalities

Theorem 5.17 (Convex functions and positive maps under the trace) *Let $\{T_i : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})\}$ be a collection of positive linear maps, $f : I \subseteq \mathbb{R} \to \mathbb{R}$ a function which is convex on I and $\{X_i \in \mathcal{M}_d(\mathbb{C})\}$ a collection of Hermitian operators with spectrum in I. Then*

$$\mathrm{tr}\left[Cf\Big(\sum_i T_i(X_i)\Big)\right] \leq \sum_i \mathrm{tr}\left[CT_i\big(f(X_i)\big)\right] \tag{5.51}$$

holds for all positive semi-definite $C \in \mathcal{M}_{d'}(\mathbb{C})$ which commute with $f\big(\sum_i T_i(X_i)\big)$ if either

1. $\sum_i T_i(\mathbb{1}) = \mathbb{1}$, *or*
2. $\sum_i T_i(\mathbb{1}) \leq \mathbb{1}$ *and* $f(0) \leq 0$.

Proof By linearity it is sufficient to consider $C = |\psi\rangle\langle\psi|$, where ψ is a normalized eigenvector of $f\big(\sum_i T_i(X_i)\big)$. Denote by $X_i = \sum_x \mu_{x,i}|x_i\rangle\langle x_i|$ the spectral decomposition of X_i. Then

$$\begin{aligned} \langle\psi|f\Big(\sum_i T_i(X_i)\Big)|\psi\rangle &= f\Big(\langle\psi|\sum_i T_i(X_i)|\psi\rangle\Big) && (5.52)\\ &= f\Big(\sum_{x,i} p_{x,i}\ \mu_{x,i}\Big), \quad p_{x,i} := \langle\psi|\sum_i T_i(|x_i\rangle\langle x_i|)|\psi\rangle \\ &\leq \sum_{x,i} p_{x,i} f(\mu_{x,i}) = \sum_i \langle\psi|T_i\big(f(X_i)\big)|\psi\rangle. && (5.53) \end{aligned}$$

Here the first equality follows from ψ being an eigenvector. To see this recall that applying a function to a Hermitian matrix means applying it to the eigenvalues (see Eq.(5.15) and the discussion of functional calculus). For the inequality in Eq.(5.53)we use convexity of f together with the fact that the $p_{x,i}$'s are positive (due to positivity of the T_i's) and sum up to one if $\sum_i T_i(\mathbb{1}) \leq \mathbb{1}$ holds with equality. If equality does not hold we can add another term which completes the probability distribution and for which the corresponding $\mu = 0$. In this case the inequality then follows from convexity together with $f(0) \leq 0$. □

The following is a simple but important consequence of the above Thm.5.17. It follows by setting $C = \mathbb{1}$ and $T_i = \lambda_i \mathrm{id}$ with λ being a probability vector:

Corollary 5.6 (Convex trace functions) *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be a function which is convex on I. Then the map $X \mapsto \mathrm{tr}\,[f(X)]$ is convex on the set of Hermitian operators with spectrum in I.*

In order to get a similar result for function which are monotone, but not necessarily operator monotone, it is convenient to recall how functions are differentiated under the trace. If $f : I \subseteq \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function with derivative f', then

$$\frac{\partial}{\partial_x}\mathrm{tr}\,[f(A + xB)]\Big|_{x=x_0} = \mathrm{tr}\,[B\ f'(A + x_0B)]\,, \tag{5.54}$$

for all Hermitian operators $A, B \in \mathcal{M}_d(\mathbb{C})$ for which the spectrum of $A + xB$ in a neighborhood of x_0 is contained in I.[4] With this at hand we obtain:

Proposition 5.2 (Monotone trace functions) *Let $f : I \to \mathbb{R}$ be a continuously differentiable and non-decreasing function on an open interval $I \subseteq \mathbb{R}$. Then for all Hermitian operators $A, B \in \mathcal{M}_d(C)$ with spectrum in I and $A \leq B$ we have that*

$$\mathrm{tr}\left[f(A)\right] \leq \mathrm{tr}\left[f(B)\right]. \tag{5.55}$$

Proof Consider the function $g(x) := \mathrm{tr}\left[f(A + x\Delta)\right]$ with $\Delta := (B - A) \geq 0$. Following Eq.(5.54) we have $g'(x_0) = \mathrm{tr}\left[\Delta f'(A + x_0\Delta)\right]$ which is non-negative for $x_0 \geq 0$ since it is the trace of the product of two positive semi-definite operators. Thus $g(0) \leq g(1)$. □

A slightly stronger result can be obtained from Weyl's monotonicity theorem which states that for any pair of Hermitian $d \times d$ matrices for which $A \leq B$, it holds that

$$\forall j: \quad \lambda_j^{\downarrow}(B) \geq \lambda_j^{\downarrow}(A) + \lambda_d^{\downarrow}(B - A), \tag{5.56}$$

where the $\lambda_j^{\downarrow}$'s are the decreasingly ordered eigenvalues. Since $\lambda_d^{\downarrow}(B - A) \geq 0$, one obtains immediately:

Proposition 5.3 *For any pair of Hermitian operators $A, B \in \mathcal{M}_d(\mathbb{C})$ there exists a unitary $U \in \mathcal{M}_d(\mathbb{C})$ so that for any $f : I \to \mathbb{R}$ which is non-decreasing on an interval I which includes the spectra of both A and B, we have*

$$f(A) \leq U f(B) U^{\dagger}. \tag{5.57}$$

adding unitaries

5.6 Operator means

[4] This is easily verified by looking at a polynomial approximation of f.

Chapter 6

Spectral properties

If a linear map has equal input and output space, like $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$, we can assign a *spectrum* to it. As outlined in Sec.1.6 the spectrum can be defined in purely algebraic terms as the set of complex numbers λ for which $(\lambda\,\mathrm{id} - T)$ is not invertible. Since we restrict ourselves to finite dimensions this is equivalent to the set of λ's for which there is an $X \in \mathcal{M}_d(\mathbb{C})$ such that

$$T(X) = \lambda X. \tag{6.1}$$

Although X is an operator, will refer to it as an *eigenvector* of T as $\mathcal{M}_d(\mathbb{C})$ is considered as a vector space in this context. Hence, the spectrum of T is the collection of eigenvalues of the $d^2 \times d^2$ matrix representation $\hat{T}$ introduced in Sec.2.3. Using this representation we see that the eigenvalues of T and T^* coincide due to $\hat{T^*} = \hat{T}^T$ when using a Hermitian basis of operators in Eq.(2.20).

Location of eigenvalues: For Hermiticity preserving maps (i.e., in particular for positive maps) the adjoint of Eq.(6.1) implies that eigenvalues are either real or they come in complex conjugate pairs.[1] Moreover, for every real λ there is at least one corresponding Hermitian eigenvector X.

Recall that the *spectral radius* is defined as $\varrho(T) := \sup\{|\lambda| \mid \lambda \in \mathrm{spec}(T)\}$, i.e., as the eigenvalue which is largest in magnitude.

Proposition 6.1 (Spectral radius of positive maps) *If T is a positive map on $\mathcal{M}_d(\mathbb{C})$, then its spectral radius satisfies*

$$\varrho(T) \leq ||T(\mathbb{1})||_\infty. \tag{6.2}$$

If in addition T is unital or trace-preserving, then there is an eigenvalue $\lambda = 1$. That is, in this case $\varrho(T) = 1$ so that all eigenvalues of T lie in the unit disc of the complex plane.

[1] More precisely, if we decompose X into Hermitian and anti-Hermitian parts and use the Hermiticity preservation together with linearity we obtain $T(X^\dagger) = \bar{\lambda} X^\dagger$.

For a more detailed investigation of the spectral radius see the subsequent section (in particular Thm. 6.5).

PROOF By the theorem of Russo and Dye we have $||T(X)||_\infty \leq ||T(\mathbb{1})||_\infty ||X||_\infty$. Therefore if $T(X) = \lambda X$, then

$$|\lambda| \; ||X||_\infty = ||T(X)||_\infty \leq ||T(\mathbb{1})||_\infty ||X||_\infty, \tag{6.3}$$

which implies Eq.(6.2). If T is unital then $T(\mathbb{1}) = \mathbb{1}$ provides an eigenvalue $\lambda = 1$. Similar holds if T is trace-preserving, since then T^* (which has the same spectrum) is unital. □

Another way to see that the spectral radius of every quantum channel (or trace-preserving positive map) is one would be to use that $\varrho(\hat{T}) = \lim_{n\to\infty} ||\hat{T}^n||^{1/n}$ (see Eq.(1.35)) together with the fact that T^n remains a quantum channel for every n and that quantum channels in finite dimensions have bounded norm (see ...).

To summarize, the eigenvalues of every quantum channel are restricted to lie on the unit disc, there is one eigenvalue one and the others are real or come in complex conjugate pairs. In fact, there is no further restriction (as in the case of 'classical' stochastic matrices): for every $d > 1$, every $\lambda : |\lambda| \leq 1$ can appear as an eigenvalue of a quantum channel on $\mathcal{M}_d$. Only the set of eigenvalues as a whole underlies further non-trivial constraints (see Sec.6.1).

Spectral decomposition: Recall some basic linear algebra. Like every other square matrix, $\hat{T} \in \mathcal{M}_{d^2}(\mathbb{C})$ admits a *Jordan decomposition* of the form

$$\hat{T} = X \left(\bigoplus_{k=1}^{K} J_k(\lambda_k) \right) X^{-1}, \quad J_k(\lambda) := \begin{pmatrix} \lambda & 1 & & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \in \mathcal{M}_{d_k}(\mathbb{C}), \tag{6.4}$$

where the J_k's are *Jordan blocks* of size d_k (with $\sum_k d_k = d^2$) and the number K of Jordan blocks equals the number of different non-zero eigenvectors. It is sometimes useful to subdivide each Jordan block into a projection[2]and a nilpotent part so that we get

$$\hat{T} = \sum_{k=1}^{K} \lambda_k P_k + N_k, \qquad N_k^{d_k} = 0, \; N_k P_k = P_k N_k = N_k, \tag{6.5}$$

$$P_k P_l = \delta_{kl} P_k, \; \mathrm{tr}\,[P_k] = d_k, \; \sum_k P_k = \mathbb{1}.$$

The number of Jordan blocks with eigenvalue λ is the *geometric multiplicity* of λ, while their joint dimension $\sum_{k:\lambda_k=\lambda} d_k$ is its *algebraic multiplicity*. If these two multiplicities are equal for every eigenvalue, then $\hat{T}$ is called *non-defective*. Non-defective matrices are dense (which is easily seen by perturbing the eigenvalues)

[2]This implies that in Eq.(6.5) $P_k^2 = P_k$ but not necessarily $P_k = P_k^\dagger$, i.e., P_k need not be an *orthogonal projection*.

and exactly those for which complete bases of eigenvectors exist. We can then write

$$\hat{T} = \sum_k \lambda_k |R_k\rangle\langle L_k|, \quad \langle L_k, R_l\rangle = \delta_{kl}, \tag{6.6}$$

so that the right eigenvectors $\{R_k\}$ and the left eigenvectors $\{L_k\}$ are biorthogonal.

Returning to the fact that T is a linear map on $\mathcal{M}_d(\mathbb{C})$ this means that in the non-defective case there are biorthogonal operator bases $\{R_k \in \mathcal{M}_d(\mathbb{C})\}$ and $\{L_k \in \mathcal{M}_d(\mathbb{C})\}$ such that

$$T(A) = \sum_k \lambda_k \mathrm{tr}\left[L_k^\dagger A\right] R_k, \quad \mathrm{tr}\left[L_k^\dagger R_l\right] = \delta_{kl}, \tag{6.7}$$

$$\text{so that} \quad \forall k: \; T(R_k) = \lambda_k R_k. \tag{6.8}$$

The Choi-Jamiolkowski operator $\tau = (T \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$ then takes the form

$$\tau = \frac{1}{d} \sum_k \lambda_k \bar{L}_k \otimes R_k. \tag{6.9}$$

If T is a trace-preserving map, then all right eigenvectors R_k corresponding to eigenvalues $\lambda_k \neq 1$ have to be traceless, $\mathrm{tr}\,[R_k] = 0$ since $\mathbb{1}$ is a left eigenvector.

Note that if T is a quantum channel, then $\hat{T}$ is generally neither Hermitian (unless $T = T^*$, see Prop.2.6) nor is it necessarily non-defective. However, the uniformly bounded norm of trace-preserving positive maps prevents eigenvalues of magnitude one (the so-called *peripheral spectrum*) from having non-trivial Jordan blocks (meaning $d_k > 1$):

Proposition 6.2 (Trivial Jordan blocks for peripheral spectrum) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving (or unital) positive liner map. If λ is an eigenvalue of T with $|\lambda| = 1$, then its geometric multiplicity equals its algebraic multiplicity, i.e., all Jordan blocks for λ are one-dimensional.*

Proof First note that for every trace-preserving (or unital) positive map $\mathrm{tr}\,[AT(B)]$ can be upper bounded in terms of the dimension d and the norms of A and B only. While we will discuss such bounds in greater detail later it suffices for now to note that we may write A and B as a linear combination of positive matrices and that for positive A, B we have $\mathrm{tr}\,[AT(B)] \leq ||A||_\infty ||B||_\infty \mathrm{tr}\,[\mathbb{1}T(\mathbb{1})] = d||A||_\infty ||B||_\infty$. The same bound has to hold for every power T^n, $n \in \mathbb{N}$. However, if we take the $n'th$ power of a Jordan block $J(\lambda)$, it is a Toeplitz matrix whose first row reads

$$[J(\lambda)^n]_{1j} = \lambda^{n-j+1} \begin{pmatrix} n \\ j-1 \end{pmatrix}, \quad j = 1, \ldots, \dim(J(\lambda)). \tag{6.10}$$

Hence, if $\dim(J(\lambda)) > 1$ and $|\lambda| = 1$ this would contain entries which grow unboundedly with n contradicting the uniform boundedness of T^n. □

We can now use the obtained results in order to show that some of the spectral projections appearing in Eq.(6.5) are (completely) positive in their own right. Define

$$\hat{T}_\infty := \sum_{k:\lambda_k=1} P_k, \tag{6.11}$$

$$\hat{T}_\phi := \sum_{k:|\lambda_k|=1} P_k, \tag{6.12}$$

$$\hat{T}_\varphi := \sum_{k:|\lambda_k|=1} \lambda_k P_k, \tag{6.13}$$

from the spectral decomposition in Eq.(6.5) and the corresponding maps $T_\infty, T_\phi, T_\varphi$ via Eq.(2.20).

Proposition 6.3 (Cesaro means) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a linear map which is trace-preserving and (completely) positive. Then $T_\infty, T_\phi, T_\varphi$ are trace preserving and (completely) positive as well. More precisely, we have that (i) there exists an increasing sequence $n_i \in \mathbb{N}$ such that $\lim_{i\to\infty} T^{n_i} = T_\phi$, (ii) $T_\varphi = TT_\phi$ and (iii)*

$$T_\infty = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n. \tag{6.14}$$

Proof Showing (i)-(iii) essentially proves the claim that the three maps are trace-preserving and (completely) positive. Throughout we will use that the spectral radius of T is one, that there is an eigenvalue one and that the peripheral spectrum of T only has one-dimensional Jordan blocks. For (i) we exploit the subsequent Lemma on simultaneous Diophantine approximations. For our purpose Lemma 6.1 implies that for every $\epsilon > 0$ there is an $n \in \mathbb{N}$ such that $|\lambda_k^n - 1| \leq \epsilon$ for all eigenvalues λ_k of magnitude one. This means that there is an increasing sequence $n_i \in \mathbb{N}$ such that T^{n_i} is a better and better approximation to T_ϕ. Since the set of (completely) positive, trace-preserving maps on $\mathcal{M}_d$ is compact, the limit point T_ϕ belongs to this set. (ii) is obvious from the spectral decomposition in Eq.(6.5) and (iii) is an immediate consequence of the geometric series $\sum_{n=1}^N \lambda^n = (\lambda - \lambda^{N+1})/(1-\lambda)$ for $\lambda \neq 1$. □

In a similar vein as T_∞ is expressed as a *Cesaro mean* in Eq.(6.14) we can write

$$T_\phi = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{k:|\lambda_k|=1} \big(\lambda_k T\big)^n. \tag{6.15}$$

The maps T_∞ and T_ϕ are projections onto the set of fixed points (see Chap.6.4) and the space spanned by the eigenvectors of the peripheral spectrum respectively. That is, $X \in \mathcal{M}_d$ is in the range of T_∞ iff $T(X) = X$ and it is in the range of T_ϕ iff for every $\epsilon > 0$ there is an $n \in \mathbb{N}$ such that $||T^n(X) - X|| \leq \epsilon$. In other words the image of T_ϕ is the set of operators for which recurrences arise.

For completeness (and later use) Dirichlet's theorem on simultaneous Diophantine approximations which was used in the above proof:

Lemma 6.1 (Dirichlet's theorem) *Let $x_1, \ldots, x_m$ be m real numbers and $q > 1$ any integer. Then there exist integers $n, p_1, \ldots, p_m$ such that*

$$1 \leq n \leq q^m \quad \textit{and} \quad |x_k n - p_k| \leq \frac{1}{q}, \quad \forall k. \tag{6.16}$$

Resolvents: The *resolvent* R of a linear operator T is an operator-valued function on $\mathbb{C}$ defined as

$$R(z) := (z\mathrm{id} - T)^{-1}. \tag{6.17}$$

The singularities of R are exactly the eigenvalues of T. Hence, R is *meromorphic* and defined on the set which is the complement of the spectrum of T, called *resolvent set*. For $|z| > \varrho(T)$, i.e., outside the spectral radius of T, we can express the resolvent as

$$R(z) = \frac{1}{z} \sum_{t=0} \frac{T^t}{z^t}. \tag{6.18}$$

If T^t describes the discrete time evolution of a quantum (or classical) system this series implies that the spectrum of T is reflected in a peculiar way in the time evolution of expectation values: consider any $\rho, A \in \mathcal{M}_d(\mathbb{C})$ and the sequence $\langle A(t) \rangle := \mathrm{tr}\left[\rho T^t(A)\right]$ for $t = 0, 1, 2, \ldots$. If A represents an observable and ρ is a density matrix evolving according to T, then $\langle A(t) \rangle$ is the expectation value at time t. The *z-transform* (or *discrete Laplace transform*) of the 'signal' $\langle A(t) \rangle$ is given by

$$\mathcal{L}(z) := \frac{1}{z} \sum_{t \in \mathbb{N}_0} \frac{\langle A(t) \rangle}{z^t}, \tag{6.19}$$

which converges for $|z| > \varrho(T)$ and can be defined in the interior of that disc by analytic continuation. Obviously, the poles of $\mathcal{L}$ are located at eigenvalues of T. Note, however, that depending on A and ρ not every eigenvalue of T will give rise to a pole of $\mathcal{L}$.

Using the resolvent, many of the results of analytic function theory carry over to the world of operators. For instance if $\Delta \subseteq \mathbb{C}$ is simply connected and enclosed by a smooth curve $\partial\Delta$, then

$$\sum_{k:\lambda_k \in \Delta} P_k = \frac{1}{2\pi i} \int_{\partial\Delta} R(z) dz, \quad \text{and} \tag{6.20}$$

$$f(T) = \frac{1}{2\pi i} \int_{\partial\Delta} f(z) R(z) dz, \quad \text{if } \lambda_k \in \Delta \; \forall k, \tag{6.21}$$

where $f : \mathbb{C} \to \mathbb{C}$ is any function holomorphic in Δ and the P_k's are the projections appearing in the spectral decomposition in Eq.(6.5).

6.1 Determinants

The determinant (=product of all eigenvalues) is a useful tool in particular when it comes to the composition of maps since

$$\det(T_1 T_2) = \det(T_1)\det(T_2). \tag{6.22}$$

In fact, the property of being multiplicative essentially characterizes the determinant: if a functional $F : \mathcal{M}_D(\mathbb{C}) \to \mathbb{C}$ satisfies $F(T_1T_2) = F(T_1)F(T_2)$ for all $T_i \in \mathcal{M}_D(\mathbb{C})$, then $F(T) = f\big(\det(T)\big)$ for some multiplicative function $f : \mathbb{C} \to \mathbb{C}$ [**?**].

Also note that Eq.(6.22) implies the simple relation $\prod_i s_i = |\det(T)|$ between the determinant and the singular values $\{s_i\}$ of a map.

The following specifies the possible range of determinants for positive and trace-preserving maps on $\mathcal{M}_d(C)$ and characterizes the extremal values:

Theorem 6.1 (Determinants) *Let $T : \mathcal{M}_d \to \mathcal{M}_d$ be a positive and trace preserving linear map.*

1. $\det T$ *is real and contained in the interval* $[-1, 1]$*,*
2. $|\det T| = 1$ *iff T is a either a unitary conjugation (i.e., $T(A) = UAU^\dagger$ for some unitary $U \in \mathcal{M}_d$) or unitarily equivalent to a matrix transposition (i.e., $T(A) = UA^TU^\dagger$ for some unitary $U \in \mathcal{M}_d$),*
3. *if T is a unitary conjugation then* $\det T = 1$ *and if* $\det T = -1$ *then T is a matrix transposition up to unitary equivalence. In both cases the converse holds iff $\lfloor \frac{d}{2} \rfloor$ is odd.*

Proof Since eigenvalues of positive maps either come in complex conjugate pairs or are real, their product $\det T$ is real, too. Due to the additional trace-preserving property the spectral radius is one (Prop.6.1) so that $\det T \in [-1, 1]$.

Now consider the case $\det T = \pm 1$ where all eigenvalues are phases. Following Dirichlet's Theorem (Lem.6.1) as in the first part of Prop.6.3, there is always a subsequence n_i such that the limit of powers $\lim_{i\to\infty} T^{n_i} =: T_\phi$ has eigenvalues which all converge to one which implies that $T_\phi = \mathrm{id}$. Hence, the inverse $T^{-1} = T_\phi T^{-1} = \lim_{i\to\infty} T^{n_i - 1}$ is a trace-preserving positive map as well.

Assume that the image of any pure state density matrix Ψ under T is mixed, i.e., $T(\Psi) = \lambda\rho_1 + (1-\lambda)\rho_2$ with $\rho_1 \neq \rho_2$, $\lambda \in (0, 1)$. Then by applying T^{-1} to this decomposition we would get a nontrivial convex decomposition for Ψ (due to positivity of T^{-1}) leading to a contradiction. Hence, T and its inverse map pure states onto pure states. Furthermore, they are unital, which can again be seen by contradiction. So assume $T(\mathbb{1}) \neq \mathbb{1}$. Then the smallest eigenvalue of $T(\mathbb{1})$ satisfies $\lambda_{min} < 1$ due to the trace preserving property. If we denote by $|\lambda\rangle$ a corresponding eigenvector, then $\mathbb{1} - \frac{\lambda_{min}+1}{2} T^{-1}(|\lambda\rangle\langle\lambda|)$ is a positive operator, but its image under T would no longer be positive. Therefore we must have $T(\mathbb{1}) = \mathbb{1}$.

Every unital positive trace preserving map is contractive with respect to the Hilbert-Schmidt norm. As this holds for both T and T^{-1} we have that $\forall A \in \mathcal{M}_d : ||T(A)||_2 = ||A||_2$, i.e., T acts unitarily on the Hilbert-Schmidt Hilbert space. In particular, it preserves the Hilbert Schmidt scalar product $\mathrm{tr}\left[T(A)T(B)^\dagger\right] = \mathrm{tr}\left[AB^\dagger\right]$. Applying this to pure states $A = |\phi\rangle\langle\phi|$ and $B = |\psi\rangle\langle\psi|$ shows that T gives raise to a mapping of the Hilbert space onto itself which preserves the value of $|\langle\phi|\psi\rangle|$. By Wigner's theorem (Thm. 1.1) this has to be either unitary or anti-unitary. If T is a unitary conjugation then $\det T = \det(U \otimes \bar{U}) = 1$ due to Eq.(2.21). Since every anti-unitary is unitarily equivalent to complex conjugation, we get that T is in this case a matrix transposition $T(A) = A^T$ (up to unitary equivalence). The determinant of the matrix transposition is easily seen in the Gell-Mann basis of $\mathcal{M}_d$ (see Hermitian operator bases in Sec.2.3). That is, we take basis elements F_α of the form $\sigma_x/\sqrt{2}, \sigma_y/\sqrt{2}$ for $\alpha = 1, \ldots d^2 - d$ and diagonal for $\alpha = d^2 - d + 1, \ldots, d^2$. In this basis matrix transposition is diagonal and has eigenvalues 1 and -1 where the latter appears with multiplicity $d(d-1)/2$. This means that matrix transposition has determinant minus one iff $d(d-1)/2$ is odd, which is equivalent to $\lfloor \frac{d}{2} \rfloor$ being odd. □

From this we get the following corollaries:

Corollary 6.1 (Monotonicity of the determinant) *Let T, T' be two positive and trace-preserving linear maps on $\mathcal{M}_d$. Then the determinant is decreasing in magnitude under composition, i.e., $|\det T| \geq |\det TT'|$ where equality holds iff T' is a unitary conjugation, a matrix transposition or $\det T = 0$.*

Corollary 6.2 (Positive invertible maps) *Let $T : \mathcal{M}_d \to \mathcal{M}_d$ be a trace-preserving and positive linear map. Then the inverse T^{-1} is positive, too iff T is a unitary conjugation or matrix transposition.*

Note that the inverse (if it exists) of a trace-preserving linear map is automatically trace-preserving. Hence, the above statement implies that the inverse of a quantum channel is again a quantum channel iff it describes a unitary time evolution.

One might wonder whether completely positive maps can have negative determinants. The following simple example answers this question in the affirmative. It is build up on the map $\rho \mapsto \rho^{T_c}$ which transposes the corners of $\rho \in \mathcal{M}_d$, i.e., $(\rho^{T_c})_{k,l}$ is $\rho_{l,k}$ for the entries $(k,l) = (1,d), (d,1)$ and remains $\rho_{k,l}$ otherwise. Note that for $d = 2$ this is the ordinary matrix transposition.

Example 6.1 (Negative determinant) *The map $T : \mathcal{M}_d \to \mathcal{M}_d$ defined by*

$$T(\rho) = \frac{\rho^{T_c} + \mathbb{1}\mathrm{tr}\left[\rho\right]}{1 + d} \tag{6.23}$$

is trace preserving, completely positive with Kraus rank $d^2 - 1$ and has determinant $\det T = -\big(d+1\big)^{1-d^2}$. For $d = 2$ the channel is entanglement breaking and can be written as

$$T(\rho) = \frac{1}{3}\sum_{j=1}^{6} |\bar{\xi}_j\rangle\langle\xi_j|\rho|\xi_j\rangle\langle\bar{\xi}_j|, \tag{6.24}$$

where the six ξ_j are the normalized eigenvectors of the three Pauli matrices. Moreover, the example given in Eq.(6.24) minimizes the determinant within the set of all quantum channels on $\mathcal{M}_2$.

Proof *A convenient matrix representation of the channel is given in the generalized Gell-Mann basis. Choose F_1 as the $\sigma_y/\sqrt{2}$ element corresponding to the corners and $F_2 = \mathbb{1}/\sqrt{d}$ the only traceless element. Then $\hat{T} = \mathrm{diag}[-1, 1+d, 1, \ldots, 1]/(d+1)$ leading to $\det T = -(d+1)^{1-d^2}$.*

For complete positivity we have to check positivity of the Jamiolkowski state τ. The corner transposition applied to a maximally entangled state leads to one negative eigenvalue $-1/d$. This is, however, exactly compensated by the second part of the map such that $\tau \geq 0$ with rank d^2-1.

The representation for $d=2$ is obtained from $\mathrm{tr}\,[AT(B)] = \mathrm{tr}\,\left[(A\otimes B^T)\tau\right]/d$ by noting that in this case τ is proportional to the projector onto the symmetric subspace which in turn can be written as $\frac{1}{2}\sum_j |\xi_j\rangle\langle\xi_j|^{\otimes 2}$ in agreement with the given Kraus representation of the channel.

In order to see that the determinant is minimized for $d=2$, recall from Sec.2.4 that any trace-preserving T can be conveniently represented in terms of the real 4×4 matrix $\hat{T}_{ij} := \mathrm{tr}\,[\sigma_i T(\sigma_j)]/2$ where the σ_is are identity and Pauli matrices:

$$\hat{T} = \begin{pmatrix} 1 & 0 \\ v & \Lambda \end{pmatrix}, \tag{6.25}$$

where $v \in \mathbb{R}^3$, Λ is a 3×3 matrix and $\det\Lambda = \det T$. To simplify matters we can diagonalize Λ by special orthogonal matrices $O_1\Lambda O_2 = diag\{\lambda_1, \lambda_2, \lambda_3\}$ corresponding to unitary operations before and after the channel (see Exp.1.1 for the relation between $SO(3)$ and $SU(2)$). Obviously, this does neither change the determinant, nor complete positivity. For the latter it is necessary that $\vec{\lambda}$ is contained in a tetrahedron spanned by the four corners of the unit cube with $\lambda_1\lambda_2\lambda_3 = 1$. Fortunately, all these points can indeed be reached by unital channels ($v=0$) for which this criterion becomes also sufficient for complete positivity. By symmetry we can restrict our attention to one octant and reduce the problem to maximizing $p_1p_2p_3$ over all probability vectors $\vec{p}$ yielding $\vec{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = (\lambda_1, -\lambda_2, \lambda_3)$. Hence, the minimal determinant is $-(\frac{1}{3})^3$ and the corresponding channel can easily be constructed from $\hat{T}$ as $T: \rho \mapsto \frac{1}{3}(\rho^T + \mathbb{1})$. □

The example of a completely positive map with negative determinant constructed above has at least Kraus rank 3 (for $d=2$). In fact, this is the minimal number of Kraus operators which allows for a negative determinant:

Proposition 6.4 (Positive determinant for small Kraus rank) *If $T: \mathcal{M}_d \to \mathcal{M}_d$ is a completely positive linear map with Kraus rank at most two, then* $\det T \geq 0$.

Proof If $A, B \in \mathcal{M}_d$ are Kraus operators then $\hat{T} = A\otimes\bar{A} + B\otimes\bar{B}$ is a matrix representation of T (see Eq.(2.21)). Assume for the moment that $\det A \neq 0$. Then due to multiplicativity of the determinant

$$\det T = \det(A\otimes\bar{A})\det(\mathbb{1} + A^{-1}B\otimes\bar{A}^{-1}\bar{B}) \geq 0. \tag{6.26}$$

In order to understand the inequality note that if $\{\lambda_j\}$ are the eigenvalues of $A^{-1}B$, then $\{1+\lambda_i\bar{\lambda}_j\}$ are those of $\mathbb{1} + A^{-1}B\otimes\bar{A}^{-1}\bar{B}$. That is, they all come in complex conjugate pairs so that their product is positive. The same argumentation holds, of course, for the $A \otimes \bar{A}$ term. The proof is completed by using that the set of maps with $\det A \neq 0$ is dense and that $\det T$ is continuous. □

Thm. 6.1 shows that if for a quantum channel $\det T = 1$, then it has to be a unitary or, expressed in terms of the purity of the Jamiolkowsky state, $\mathrm{tr}\left[\tau^2\right] = 1$. By continuity a channel close to a unitary, described by a τ with purity close to one, will still have determinant close to one. The following is a quantitative version of the converse: if the determinant is large, then the purity has to be large, too. We will see later that and how this implies closeness to a unitary for quantum channels.

Proposition 6.5 (Determinant and Choi-Jamiolkowski operator) *Let $T : \mathcal{M}_d \to \mathcal{M}_d$ be a linear map and τ the corresponding Choi-Jamiolkowski operator. Then*

$$|\det T| \leq \mathrm{tr}\left[\tau^\dagger \tau\right]^{d^2/2}. \tag{6.27}$$

Proof To relate the 'purity' to the determinant we exploit that by Eq.(2.22) $\hat{T} = d\tau^\Gamma$ is a matrix representation of the map T, so that

$$\mathrm{tr}\left[\tau^\dagger \tau\right] = \frac{1}{d^2}\mathrm{tr}\left[\hat{T}^\dagger \hat{T}\right] = \frac{1}{d^2}\sum_{i=1}^{d^2} s_i^2, \tag{6.28}$$

where the s_i are the singular values of $\hat{T}$ (and thus T). From this Eq.(6.27) is obtained via the geometric–arithmetic mean inequality together with the fact that $|\det T| = \prod_i s_i$. □

6.2 Irreducible maps and Perron-Frobenius theory

Perron-Frobenius theory, roughly speaking, deals with the dominant eigenvalues and eigenvectors of (element-wise) non-negative matrices (see Exp. ...), i.e., in particular with those of classical channels. In order to make the theory work the considered maps have to fulfill a generic condition like *irreducibility, primitivity* or *strict positivity* (in order of increasing restriction).

In this section we will discuss the generalization of Perron-Frobenius theory to the context of positive maps on $\mathcal{M}_d(\mathbb{C})$. The following theorem defines *irreducible* positive maps and shows that this can be done in various equivalent ways:

Theorem 6.2 (Irreducible positive maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive linear map. The following properties are equivalent:*

1. Irreducibility*: If $P \in \mathcal{M}_d(\mathbb{C})$ is a Hermitian projector such that $T(P\mathcal{M}_d P) \subseteq P\mathcal{M}_d P$, then $P \in \{0, \mathbb{1}\}$.*

2. *For every non-zero $A \geq 0$ we have* $(\mathrm{id} + T)^{d-1}(A) > 0$.

3. *For every non-zero $A \geq 0$ and every strictly positive $t \in \mathbb{R}$ we have* $\exp[tT](A) > 0$.

4. *For every orthogonal pair of non-zero, positive semi-definite $A, B \in \mathcal{M}_d(\mathbb{C})$ there is an integer $t \in \{1, \ldots, d-1\}$ such that* $\mathrm{tr}\,[BT^t(A)] > 0$.

Proof 1. $\rightarrow$ 2.: From $T(A) \geq 0$ we get an inclusion for the kernels $\ker(\mathrm{id} + T)(A) \subseteq \ker A$. Suppose equality holds in this inclusion, then $\mathrm{supp} T(A) \subseteq \mathrm{supp} A$. Therefore $T(P\mathcal{M}_d P) \subseteq P\mathcal{M}_d P$ if we take P the Hermitian projector onto the support space of A. By 1. (irreducibility) this can only be if $A > 0$. Application of $(\mathrm{id}+T)$ thus has to increase the rank until there is no kernel left, which happens after $d-1$ steps at the latest.

2. $\rightarrow$ 3.: Looking at the series expansion of both $(\mathrm{id} + T)^{d-1}(A)$ and $\exp[tT](A)$ in terms of powers of T we realize that (i) all terms are positive, and (ii) all terms appearing in the first expansion are also present in the second one. Therefore we can bound $\exp[tT](A) \geq c(\mathrm{id}+T)^{d-1}(A)$ with some non-zero positive constant c.

3. $\rightarrow$ 1.: Suppose T is reducible, i.e., there is a proper projection $P \neq 0, \mathbb{1}$ for which $T(P) \leq cP$ for some scalar $c > 0$. Then $\exp[tT](P) \leq \exp[tc]P$, so T does not fulfill 3..

4. $\rightarrow$ 1.: If T is reducible by a proper projection P, then $\mathrm{tr}\,[(\mathbb{1} - P)T^t(P)] = 0$ for all $t \in \mathbb{N}$.

2. $\rightarrow$ 4.: Expanding $\mathrm{tr}\left[B(\mathrm{id} + T)^{d-1}(A)\right] > 0$ and using $\mathrm{tr}\,[BA] = 0$ together with positivity of all terms in the expansion, we get that $\mathrm{tr}\,[BT^t(A)] > 0$ for at least one $t \leq d-1$. □

Note that T is irreducible iff the dual map T^* is irreducible. If a proper projection P 'reduces' T in the sense that $\mathrm{tr}\,[(\mathbb{1} - P)T(P)] = 0$, then $\mathbb{1} - P$ reduces T^*.

In order to relate *irreducibility* to spectral properties of a positive map T it is useful to consider the following functionals defined on the cone of positive semi-definite operators:

$$r(X) \;:=\; \sup\{\lambda \in \mathbb{R} \mid (T - \lambda \mathrm{id})(X) \geq 0\}, \tag{6.29}$$

$$\tilde{r}(X) \;:=\; \inf\{\lambda \in \mathbb{R} \mid (T - \lambda \mathrm{id})(X) \leq 0\}. \tag{6.30}$$

We are especially interested in the maxima $r := \sup_{X \geq 0} r(X)$ and $\tilde{r} := \sup_{X \geq 0} \tilde{r}(X)$ which obviously satisfy $r \geq \tilde{r}$ and actually coincide for irreducible maps:

Theorem 6.3 (Spectral radius of irreducible maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$ be an irreducible positive map. Then*

1. *For the quantities introduced below Eqs.(6.29,6.30)* $r = \tilde{r}$.

2. *r is a non-degenerate eigenvalue of T and the corresponding eigenvector is strictly positive, i.e.,* $T(X) = rX > 0$.

3. *If there is any* $\lambda > 0$ *which is an eigenvalue of* T *with positive egenvector, i.e.,* $T(Y) = \lambda Y \geq 0$, *then* $\lambda = r$.

4. r *is the spectral radius of* T.

Proof First note that all suprema/infima in Eqs.(6.29,6.30) and for defining $r, \tilde{r}$ are attained since the respective sets are compact or can be made so, e.g., by imposing $\text{tr}[X] = 1$.

We begin with showing that r is attained for a strictly positive X and that $T(X) = rX$. To this end consider any non-zero $X \geq 0$ for which $\lambda := r(X) > 0$. The identity

$$(T + \text{id})^{d-1}(T - \lambda\text{id})(X) = (T - \lambda\text{id})(T + \text{id})^{d-1}(X), \tag{6.31}$$

shows two things. First, the supremum $\sup_{X \geq 0} r(X)$ is attained for a strictly positive X, since by Thm.6.2 $(T + \text{id})^{d-1}(X) > 0$. Second, the X achieving $r(X) = r$ must satisfy $(T - r\text{id})(X) = 0$, since otherwise the expression in Eq.(6.31) would be positive definite so that a larger multiple of the identity map could be subtracted—contradicting the supposed maximality. As $r(X) = \tilde{r}(X)$ for every eigenvector $X \geq 0$, the just mentioned observation together with the inequality $r \geq \tilde{r}$ proves item 1. of the theorem.

To prove 2. we still need to show non-degeneracy. Suppose there would be an eigenvector X' corresponding to r which is not a multiple of X. By taking the Hermitian (or i times the anti-Hermitian) part we can assume $X' = X'^{\dagger}$. As $X > 0$ there is a non-zero $c \in \mathbb{R}$ such that $0 \leq X + cX'$ has a kernel. This is, however, in conflict with

$$0 < (T + \text{id})^{d-1}(X + cX') = (r + 1)^{d-1}(X + cX'), \tag{6.32}$$

so that X' must be a multiple of X. Hence, r is indeed non-degenerate.

For the proof of 3. assume that $Y \geq 0$ is an eigenvector of T corresponding to an eigenvalue $\lambda > 0$. Using that r is an eigenvalue of the dual map T^* with a strictly positive eigenvector $X' > 0$ we get

$$r\text{tr}[X'Y] = \text{tr}[T^*(X')Y] = \text{tr}[X'T(Y)] = \lambda\text{tr}[X'Y]. \tag{6.33}$$

Since $\text{tr}[X'Y] > 0$ this implies $r = \lambda$.

Finally, in order to prove 4. we define a positive map $T'(\cdot) := X^{-1/2}T(X^{1/2} \cdot X^{1/2})X^{-1/2}/r$ with $X > 0$ being the positive eigenvector corresponding to the eigenvalue r of T. Note that up to the factor $1/r$ the maps T and T' differ only by a similarity transformation, i.e., their spectra coincide up to that factor. In particular the spectral radii satisfy $\varrho(T') = \varrho(T)/r$. Since T' is a unital positive map we have $\varrho(T') = 1$ by Prop.6.1 and thus $\varrho(T) = r$. □

The proof of the last item of Thm.6.3 leads to the following simple observation:

Proposition 6.6 (Similarity transformations preserving irreducibility) *Let* $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$ *be an irreducible positive map. Then for arbitrary* *constants $c > 0$ and invertible matrices $C \in \mathcal{M}_d(\mathbb{C})$ the maps*

$$T'(\cdot) := c\, C^{-1} T(C \cdot C^\dagger) C^{-\dagger}, \tag{6.34}$$

are irreducible as well.

PROOF Assume that a projection Q would 'reduce' T'. Then the projection P onto the support of $CQC^\dagger$ would reduce T. By invertibility of C we have that P and Q have the same rank and thus T' can only be reducible if T is. □

The invariance of irreducibility under similarity transformation of the form in Eq.6.34 shows that irreducibility does not only imply spectral properties of a positive map, but that it is determined by such properties:

Theorem 6.4 (Irreducibility from spectral properties) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive map with spectral radius ϱ. The following are equivalent:*

1. *T is irreducible.*

2. *The spectral radius ϱ is a non-degenerate eigenvalue and the corresponding right and left eigenvectors are positive definite (i.e., $T(X) = \varrho X > 0$ and $T^*(Y) = \varrho Y > 0$).*

PROOF 1. $\to$ 2. is given by Thm.6.3 when applied to T and T^*.

For the converse 2. $\to$ 1. we apply Eq.(6.34) with $c = 1/\varrho$ and $C = Y^{-1/2}$. In this way T' becomes trace preserving with $T(X') = X' > 0$, $X' := \sqrt{Y} X \sqrt{Y}$. Moreover, the non-degeneracy translates to that of the eigenvalue 1 of T'. Now assume that T' (and thus T) would be reducible, i.e., there is a Hermitian projection $P \notin \{0, \mathbb{1}\}$ such that $T' : \mathcal{A} \to \mathcal{A}$ with $\mathcal{A} := P\mathcal{M}_d(\mathbb{C})P$. In this case T' would have a fixed point density operator $\rho \in \mathcal{A}$ for which $T'(\rho) = \rho \geq 0$ (see Sec.6.4). Hence, the eigenvalue 1 of T would have to be degenerate since ρ is not a multiple of X' because of $X' \notin \mathcal{A}$ (as $X' > 0$). □

If we consider the time evolution of a density operator of the form $\rho(t) := T^t(\rho)$ where T is a positive and trace-preserving map and $t \in \mathbb{N}$, then the previous theorem allows us to relate irreducibility to a unique time-average corresponding to a full rank stationary state:

Corollary 6.3 (Time-average and ergodicity) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive, trace-preserving linear map. Then the following are equivalent:*

1. *T is irreducible,*

2. *There is a unique state with density matrix $\sigma > 0$ such that for every density matrix $\rho \in \mathcal{M}_d(\mathbb{C})$ we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} T^t(\rho) = \sigma. \tag{6.35}$$

PROOF Note that Eq.(6.35) is nothing but $T_\infty(\rho) = \sigma$, where T_∞ is the projection onto the eigenspace of T with eigenvalue one as given in Eq.(6.14). The equivalence 1. $\leftrightarrow$ 2. thus follows from the equivalence in Thm.6.4. □

Irreducible positive maps are generic, i.e., in particular dense within the set of all positive maps. For instance $X \mapsto T(X) + \epsilon \mathbb{1}\mathrm{tr}\,[X]$ is irreducible for every $\epsilon > 0$ and positive T since it maps every positive input to a positive definite output. By continuity of eigenvalues and eigenvectors we can therefore extend the above observation about the spectral radius to arbitrary positive maps:

Theorem 6.5 (Spectral radius and positive eigenvectors) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive map with spectral radius ϱ. Then ϱ is an eigenvalue and there is a positive semi-definite $X \in \mathcal{M}_d(\mathbb{C})$ such that $T(X) = \varrho X$.*

Of course, in general the spectral radius might be a degenerate eigenvalue and the corresponding positive eigenvector may have a kernel. Even for irreducible maps there can be several eigenvalues with modulus equal to the spectral radius:

The peripheral spectrum is the set of eigenvalues whose magnitude is equal to the spectral radius, i.e., equal to one if we are considering unital positive maps (Prop.6.1). As we saw in Prop.6.2 the properties positivity and unitality together imply that all eigenvalues in the peripheral spectrum have trivial (i.e., one-dimensional) Jordan blocks. If we add irreducibility and the assumed validity of the Schwarz inequality in Eq.(5.2) we can say much more:

Theorem 6.6 (Peripheral spectrum of irreducible Schwarz maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be an irreducible positive unital map which fulfills the Schwarz inequality in Eq.(5.2). Denote by $S := \mathrm{spec}(T) \cap \exp(i\mathbb{R})$ the peripheral spectrum of T. Then*

1. *There is an integer $m \in \{1, \ldots, d^2\}$ such that $S = \{\exp(2\pi i k/m)\}_{k \in \mathbb{Z}_m}$.*
2. *All eigenvalues in S are non-degenerate.*
3. *There is a unitary U such that for all $k \in \mathbb{N}$: $T(U^k) = \gamma^k U^k$ where $\gamma := \exp(2\pi i/m)$.*
4. *U has spectral decomposition $U = \sum_{k \in \mathbb{Z}_m} \gamma^k P_k$ where the spectral projections P_k satisfy $T(P_{k+1}) = P_k$.*

PROOF Take any eigenvalue $\lambda \in S$ with corresponding eigenvector $U \in \mathcal{M}_d(\mathbb{C})$, i.e., $T(U) = \lambda U$. Then the Schwarz inequality reads $U^\dagger U \leq T(U^\dagger U)$. Since T is irreducible by assumption we can apply the tools in the proof of Thm.6.3, in particular that $r(U^\dagger U) = 1$, which implies that equality has to hold in the Schwarz inequality and that U can be rescaled to be a unitary. Moreover, by Thm.5.3 we have

$$\bar{\lambda} U^\dagger T(X) = T(U^\dagger X), \quad \forall X \in \mathcal{M}_d(\mathbb{C}). \tag{6.36}$$

Applying this to the case where X is another eigenvector of T w.r.t. an eigenvalue in S shows that S is closed under multiplication and thus a group. Since there are at most d^2 elements in S we must have $S = \exp(2\pi i\mathbb{Z}_m/m)$ for some integer $m \leq d^2$.

Now assume that U and V are both eigenvectors w.r.t. the same eigenvalue in S. Then $T(U^\dagger V) = U^\dagger V$ and, since the eigenvalue one of an irreducible unital positive map is non-degenerate (Prop.6.1 and Thm.6.3), we must have $U \propto V$. Hence, all eigenvalues in S are simple.

If U is the unitary eigenvector corresponding to the eigenvalue $\gamma := \exp(2\pi i/m)$, then the group property gives $T(U^k) = \gamma^k U^k$ for all $k \in \mathbb{N}$. In particular $T(U^m) = U^m$ which has to be equal to $\mathbb{1}$ by non-degeneracy of the eigenvalue one. Therefore we can write $U = \sum_{k\in\mathbb{Z}_m} \gamma^k P_k$ with orthogonal spectral projections P_k. This gives

$$\sum_{k\in\mathbb{Z}_m} \gamma^{k+1} P_k = \gamma U = T(U) = \sum_{k\in\mathbb{Z}_m} \gamma^k T(P_k). \tag{6.37}$$

Due to $T(U^k) = \gamma^k U^k$, T acts as an automorphism on the *-algebra generated by U. Projections are thus mapped onto projections and by comparing the left and right hand side in Eq.(6.37) we get $T(P_{k+1}) = P_k$ from the uniqueness of the spectral decomposition. □

The last item implies that $T^m(P_k) = P_k$, i.e., if $m > 1$, then T^m is reducible. Similarly, if $l, m > 1$, then T^l is reducible if l and m have a non-trivial integer factor in common.

In the proof of the above theorem Eq.(6.36) is remarkable and has consequences beyond the peripheral spectrum: if the peripheral spectrum is non-trivial ($m > 1$), then every eigenvalue of T is an element of a group of eigenvalues of the same magnitude but rotated in the complex plane:

Proposition 6.7 (Covariance and eigensystem of irreducible maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be an irreducible positive and unital linear map which fulfills the Schwarz inequality. Following Thm.6.6 denote by $\{\gamma^k\}_{k=1,\ldots,m}$ the peripheral spectrum of T and by $U = \bar{\gamma}T(U)$ the respective unitary eigenvector.*

1. *For every $X \in \mathcal{M}_d(\mathbb{C})$ we have $T(U^\dagger X U) = U^\dagger T(X) U$.*

2. *If X is an eigenvector with $T(X) = \mu X$, then*

$$T(U^\dagger X) = \bar{\gamma}\mu\, U^\dagger X, \qquad T(XU) = \gamma\mu\, XU. \tag{6.38}$$

Proof Both statements follow directly from Eq.(6.36) and its adjoint. □

Note that the statements 1. and 2. taken together imply that if μ is non-degenerate, then $U^\dagger X U = e^{i\varphi}X$. So either $[X, U] = 0$ or $e^{i\varphi}$ is a root of unity and $\mathrm{spec}(X) = e^{i\varphi}\mathrm{spec}(X)$.

6.3 Primitive maps

We will continue the discussion on Perron-Frobenius theory and direct the focus to *primitive maps*. A primitive map is an irreducible positive map with trivial peripheral spectrum. Similar to the irreducible case primitive maps can, however, be characterized by a number of equivalent properties each of which may be used as a definition:

Theorem 6.7 (Primitive maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive and trace-preserving linear map. Then the following statements are equivalent:*

1. *There exists an $n \in \mathbb{N}$ such that for every density matrix $\rho \in \mathcal{M}_d(\mathbb{C})$ we have $T^n(\rho) > 0$.*
2. *T^k is irreducible for every $k \in \mathbb{N}$ and the limit T_ϕ (see Prop.6.3) is irreducible as well.*
3. *For all density matrices ρ the limit $\lim_{k\to\infty} T^k(\rho)$ exists, is independent of ρ and given by a positive definite density matrix ρ_∞.*
4. *T has trivial peripheral spectrum (i.e., counting algebraic multiplicities, there is only one eigenvalue of magnitude one) and the corresponding eigenvector is positive definite.*

Proof 1. $\to$ 2. can be proven by contradiction: assume for instance $T_\phi = \lim_{i\to\infty} T^{n_i}$ is reducible. Then there is a density matrix ρ_0 which has a kernel and for which $\rho_0 = T_\phi(\rho_0) = \lim_i T^n\big(T^{n_i-n}(\rho)\big)$ which would imply that $T^n(\rho) \not> 0$ for some ρ. Similarly, if T^k and therefore also T^{km} is reducible for all $m \in \mathbb{N}$, then there cannot be any n for which 1. holds.

2. $\to$ 4.: If T_ϕ has a non-degenerate eigenvalue one, then, by definition of T_ϕ, the map T has trivial peripheral spectrum. Moreover, since the fixed point set of T is included in the one of T_ϕ which in turn contains only a single positive definite element, this fixed point has to be the one of T as well.

4. $\to$ 3.: Since T has only one eigenvalue of magnitude one with corresponding positive definite eigenvector $\rho_\infty = T(\rho_\infty)$ we have $\lim_{k\to\infty} T^k = T_\phi$ and $T_\phi(\rho) = \rho_\infty \mathrm{tr}\,[\rho]$ is a projection.

3. $\to$ 1.: First note that 3. implies that the eigenvalue of T which is second largest in magnitude, denote it by λ_2, satisfies $|\lambda_2| < 1$. Assume that $T^n(\rho)$ would have a kernel with eigenvector ψ. Then

$$\begin{aligned} \lambda_{\min}(\rho_\infty) &\leq \big|\langle\psi|T^n(\rho) - \rho_\infty|\psi\rangle\big| \;\leq\; ||T^n(\rho) - \rho_\infty|| && (6.39)\\ &= ||(T^n - T_\infty)(\rho - \rho_\infty)|| \;\leq\; \mu^n c\,||\rho - \rho_\infty||, && (6.40) \end{aligned}$$

where $\lambda_{\min}$ denotes the smallest eigenvalue, μ is such that $1 > \mu > |\lambda_2|$ and $||\cdot||$ is the operator norm. The constant $c > 0$ depends on T but is independent of n (see the discussion around Thm.8.23 for more details). Thus if n is sufficiently large, then Eqs.(6.39,6.40) would lead to a contradiction for every ρ. Hence, there is a finite n such that $T^n(\rho) > 0$ for all ρ.

□

If we interpret $T^t(\rho)$ for $t \in \mathbb{N}$ as discrete time-evolution, then primitive maps are exactly those for which the evolution eventually converges to a stationary state which is strictly positive (i.e., it may in some contexts be interpreted as a finite temperature Gibbs state) and independent of the initial state.

Note that if T in addition satisfies the Schwarz inequality as required by Thm.6.6, then we can replace point 2. in Thm.6.7 by the statement that T^k is irreducible for all $k < d^2$.

Moreover, if a map is completely positive, we can give three more equivalent characterizations of primitivity:

Theorem 6.8 (Completely positive primitive maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a completely positive and trace-preserving linear map with Kraus decomposition $T(\cdot) = \sum_i K_i \cdot K_i^\dagger$. Denote by $\mathcal{K}_m := \text{span}\{\prod_{k=1}^m K_{i_k}\}$ the complex linear span of all degree-m monomials of Kraus operators. Then the following are equivalent:*

1. *T is primitive.*
2. *There exists an $n \in \mathbb{N}$ such that for every non-zero $\psi \in \mathbb{C}^d$ and all $m \geq n$: $\mathcal{K}_m|\psi\rangle = \mathbb{C}^d$.*
3. *There exists a $q \in \mathbb{N}$ such that for all $m \geq q$: $\mathcal{K}_m = \mathcal{M}_d(\mathbb{C})$.*
4. *There exists a $q \in \mathbb{N}$ such that for all $m \geq q$: $(T^m \otimes \text{id}_d)(|\Omega\rangle\langle\Omega|) > 0$, where $\Omega \in \mathbb{C}^d \otimes \mathbb{C}^d$ denotes a maximally entangled state.*

The numbers q appearing in 3., 4. can be chosen equal and if q and n are chosen minimal, then $n \leq q$.

Proof First consider item 2. and observe that it obviously holds for all $m \geq n$ if it holds for some n since we can always incorporate $m - n$ Kraus operators in ψ. Similar holds for item 4. In order to see this, define $\tau_q := (T^q \otimes \text{id}_d)(|\Omega\rangle\langle\Omega|)$ and assume $\tau_q > 0$. Then there is in particular a constant $c > 0$ such that $\tau_q \geq c\tau_{q-1}$ and therefore

$$\tau_{q+1} = (T \otimes \text{id})(\tau_q) \geq c(T \otimes \text{id})(\tau_{q-1}) = c\tau_q > 0. \tag{6.41}$$

By induction the same holds then true for all integers larger than q.

1. $\leftrightarrow$ 2.: Let us introduce $K_I := \prod_{k=1}^n K_{i_k}$ with multi-index $I = (i_1, \ldots, i_n)$. If T is primitive, then by Thm.6.7 there exists an n such that $\sum_I K_I|\psi\rangle\langle\psi|K_I^\dagger$ has full rank for every ψ. Thus, for every ϕ there is an I so that $\langle\phi|K_I|\psi\rangle \neq 0$ and therefore $\text{span}\{K_I|\psi\rangle\} = \mathbb{C}^d$. Conversely, if the latter is true, then for every ϕ there will be a unitary U so that $|\phi\rangle \propto \sum_I U_{JI}K_I|\psi\rangle =: K'_J|\psi\rangle$. Since $\{K_I\}$ and $\{K'_J\}$ both provide a Kraus representation of T^n (see Thm.2.1) the full rank property is guaranteed. Following Thm.6.7 T then has to be primitive.

3. $\leftrightarrow$ 4.: Following the above lines and denoting by K_I the Kraus operators of T^q it is clear that $(T^q \otimes \text{id}_d)(|\Omega\rangle\langle\Omega|) > 0$ is equivalent to $\text{span}\{(K_I \otimes \mathbb{1}_d)|\Omega\rangle\} =$

$\mathbb{C}^d \otimes \mathbb{C}^d$. From here the equivalence between item 3. and 4. follows due to the fact that the relation $K \leftrightarrow (K \otimes \mathbb{1})|\Omega\rangle$ is a linear one-to-one mapping between $\mathcal{M}_d(\mathbb{C})$ and $\mathbb{C}^d \otimes \mathbb{C}^d$.

4. $\rightarrow$ 1.: If the Choi-Jamiolkowski matrix $\tau := (T^q \otimes \mathrm{id}_d)(|\Omega\rangle\langle\Omega|)$ has full rank, then so does

$$T^q\big(|\psi\rangle\langle\psi|\big) = d(\mathbb{1}_d \otimes \langle\bar{\psi}|)\,\tau\,(\mathbb{1}_d \otimes |\bar{\psi}\rangle) \quad \text{for all } \psi \in \mathbb{C}^d. \tag{6.42}$$

This also shows that $n = q$ is sufficient. So if we choose both n and q minimal, then $n \leq q$ in general.

1. $\rightarrow$ 4.: The proof for this implication is entirely parallel with Eqs.(6.39,6.40) where q now plays the role of n: just replace ρ by Ω, ρ_∞ by $\rho_\infty \otimes \mathbb{1}/d$ and T by $T \otimes \mathrm{id}$. Also the constant c now depends on $T \otimes \mathrm{id}$, but the argumentation remains the same. □

Note that all the above relations for both positive and completely positive maps hold as well for maps which are not trace-preserving[3] so that the notion of primitivity is well-defined for arbitrary positive maps.

Finally, we will show that the number q (and therefore also n) appearing in Thm.6.8 can be bounded by the dimension d of the system. These bounds can be seen as a quantum counterpart to Wielandt's inequality, which does the same for the classical case of primitive stochastic matrices. We need two preparatory Lemmas before the main result. As before $\mathcal{K}_n$ denotes the complex linear span of the Kraus operators of T^n. We will frequently use that every $K \in \mathcal{K}_n$ is, up to a scalar multiple, a possible Kraus operator of T^n. This follows from the freedom in the Kraus decomposition (Thm.2.1, point 4.).

Lemma 6.2 *Let T be a primitive quantum channel on $\mathcal{M}_d(\mathbb{C})$ with Kraus rank k. Then, there is an natural number $n \leq d^2 - k + 1$ and a $K \in \mathcal{K}_n$ such that* $\mathrm{tr}\,[K] \neq 0$.

PROOF Let us denote by R_n the complex linear span of all $\mathcal{K}_m$ with $m \leq n$. We have to show that: (*) for any $n \in \mathbb{N}$, if $\dim[R_n] < d^2$, then $\dim[R_{n+1}] > \dim[R_n]$. Since $\dim[R_1] = k$, by iteration we obtain that $R_{d^2-k+1} = \mathcal{M}_d(\mathbb{C})$. This implies that a linear combination of the elements of $\mathcal{K}_n$ with various $n \leq d^2 - k + 1$ must be equal to the identity, and thus at least one of the elements must have non-zero trace. To prove (*) we note that, by definition, $R_n \subseteq R_{n+1}$. If they would be equal, then $R_m = R_n$ for all $m > n$. Thus, equality can only occur when $\dim[R_n] = d^2$ since otherwise the map T would not be primitive.

Lemma 6.3 *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a primitive completely positive map such that at least one of its Kraus operators, say K_1, has a non-zero eigenvalue, i.e., $K_1|\varphi\rangle = \mu|\varphi\rangle$ with $\mu \neq 0$. Then: (a) $\mathcal{K}_{d-1}|\varphi\rangle = \mathbb{C}^d$. (b) If K_1 is not invertible, then for all $|\psi\rangle \in \mathbb{C}^d$, $|\varphi\rangle\langle\psi| \in \mathcal{K}_{d^2-d+1}$.*

[3] The limit $\lim_{k\to\infty} T^k$ has then to be normalized by the k'th power of the spectral radius of T.

Proof (a) We define S_n as the span of all $\mathcal{K}_m|\varphi\rangle$ with $m \leq n$ together with $|\varphi\rangle$. If $\dim[S_n] < d$, then $\dim[S_{n+1}] > \dim[S_n]$, since otherwise the map would not be primitive. Thus, $S_{d-1} = \mathbb{C}^d$. That is, for all $|\phi\rangle \in \mathbb{C}^d$, there exist matrices $K^{(n)} \in \mathcal{K}_{k_n}$, $k_n \leq d-1$ such that (with $K^{(0)} \propto \mathbb{1}$)

$$|\phi\rangle = \sum_{n=0}^{d-1} K^{(n)}|\varphi\rangle = \sum_{n=0}^{d-1} K^{(n)} \frac{K_1^{d-k_n}}{\mu^{d-k_n}}|\varphi\rangle, \tag{6.43}$$

and thus, $|\phi\rangle \in \mathcal{K}_{d-1}|\varphi\rangle$. (b) We write K_1 in the Jordan standard form and divide it into two blocks. The first one, of size $\tilde{d} \times \tilde{d}$, consists of all Jordan blocks corresponding to non-zero eigenvalues, whereas the second one contains all those corresponding to zero eigenvalues. We denote by P the (not necessarily Hermitian) projector onto the subspace where the first block is supported and by $r \leq d-\tilde{d}$ the size of the largest Jordan block corresponding to a zero eigenvalue. We have

$$K_1 P = P K_1, \quad K_1^r = K_1^r P. \tag{6.44}$$

We define $R_n := P\mathcal{K}_n$ and show that $R_{d\tilde{d}} = P\mathcal{M}_d(\mathbb{C})$. For all $n \in \mathbb{N}$, $\dim[R_{n+1}] \geq \dim[R_n]$. The reason is that for any set of linearly independent matrices $K_k^{(n)} \in R_n$, $K_1 K_k^{(n)} \in R_{n+1}$ are also linearly independent, given that K_1 is invertible on its range. By following the reasoning of [**?**, Appendix A] we get that, if $\dim[R_{n+1}] = \dim[R_n] =: d'$, then $\dim[R_m] = d'$ for all $m > n$, which is incompatible with T being primitive unless $d' = \tilde{d}d$. Thus, for any $|\psi\rangle \in \mathbb{C}^d$, there exists a $K \in \mathcal{K}_{\tilde{d}d}$ with $|\varphi\rangle\langle\psi| = PK = K_1^r PK/\mu^r = K_1^r K/\mu^r \in \mathcal{K}_{\tilde{d}d+r}$. By using that $\tilde{d} \leq d-r$ and that $r \geq 1$ (since K_1 is supposed to be not invertible) we get $\tilde{d}d + r \leq d^2 - d + 1$, which concludes the proof.

We have now the necessary tools to prove our main result.

Theorem 6.9 (Quantum Wielandt inequality) *Let T be a primitive quantum channel on $\mathcal{M}_d(\mathbb{C})$ with Kraus rank k. Then for the minimal q as it appears in Thm.6.8 we have that*

1. *in general $q \leq (d^2 - k + 1)d^2$,*
2. *if the span of Kraus operators $\mathcal{K}_1$ contains an invertible element, then $q \leq d^2 - k + 1$,*
3. *if the span of Kraus operators $\mathcal{K}_1$ contains an element with at least one non-zero eigenvalue, then $q \leq d^2$.*

Proof

2. If there is an invertible element, then it follows from [**?**, Appendix A, Proposition 2]that $\dim \mathcal{K}_{n+1} > \dim \mathcal{K}_n$ until the full matrix space $\mathcal{M}_d(\mathbb{C})$ is spanned and thus $q \leq d^2 - k + 1$.

1. Let us denote by $\{K_k^{(n)}\}$ the Kraus operators corresponding to T^n. According to Lemma 6.2, one of them, say $K_1^{(n)}$, has non-zero trace for some $n \leq d^2 - k + 1$ and therefore there exists $|\varphi\rangle$ such that $K_1^{(n)}|\varphi\rangle = \mu|\varphi\rangle$ with

$\mu \neq 0$. If $K_1^{(n)}$ is invertible, then 1. is implied by 2., so we can assume that it is not invertible. According to Lemma 6.3.(b), for all $|\psi\rangle, |\chi\rangle \in \mathbb{C}^d$ we have $|\varphi\rangle\langle\psi| \in \mathcal{K}_{(d^2-d+1)n}$; and according to Lemma 6.3.(a) $|\chi\rangle\langle\psi| \in \mathcal{K}_{nd^2}$. This implies that $\mathcal{K}_{nd^2} = \mathcal{M}_d(\mathbb{C})$ and hence the general bound 1. follows.

The argument which proves 3. is completely analogous just with $n = 1$ as we do, by assumption, not have to block in order to get a Kraus operator with non-zero eigenvalue. □

In the previous proof we use, for the general case, blocks of T^n for some $n \leq d^2 - k + 1$ just in order to get any Kraus operator with a non-zero eigenvalue. This rather clumsy step is not necessary if the dimension is small:

Corollary 6.4 *Let T be a primitive quantum channel on $\mathcal{M}_d(\mathbb{C})$ with $d = 2, 3$. Then there is always a Kraus operator with non-zero eigenvalue in $\mathcal{K}_1$ and thus $q \leq d^2$.*

Proof The fact that $\mathcal{K}_1$ has this property for $d = 2, 3$ stems from the classification of nilpotent subspaces[**?**]: assume that $\mathcal{K}_1$ would be a nilpotent subspace within the space of $d \times d$ matrices. Then for $d = 2$ its dimension would have to be one, so it could not arise from the Kraus operators of a quantum channel. Similarly, for $d = 3$ there are (up to similarity transformations) two types of nilpotent subspaces [**?**] of dimension $k > 1$: one of dimension $k = 3$, the space of upper-triangular matrices, whose structure does not allow the trace-preserving property, and one of dimension $k = 2$ which only leads to quantum channels having a (in modulus) degenerate largest eigenvalue. Hence, if $\mathcal{K}_1$ is generated by the Kraus operators of a primitive quantum channel, then it cannot be nilpotent if $d = 2, 3$. □

For $d \geq 4$ the set of Kraus operators can form a nilpotent subspace even for primitive quantum channels. However, exploiting the structure of such subspaces[**?**] the general bound in Thm.6.9 can still be improved.

Problem 12 (Optimal Wielandt bound) *Find the smallest integers n and q (as a function of the dimension d) for Thm.6.8.*

6.4 Fixed points

Theorem 6.10 (Brouwer's fixed point theorem) *Let T be a continuous map from a non-empty, compact, convex set $S \subset \mathbb{R}^n$ into itself, then there is an $x \in S$ such that $T(x) = x$.*

As the set of Hermitian matrices in $\mathcal{M}_d(\mathbb{C})$ is a real vector space and the density matrices therein form a compact, convex set, Brouwer's fixed point theorem immediately implies:

Theorem 6.11 (Stationary states) *Every continuous, trace-preserving, positive (not necessarily linear) map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ has at least one* stationary state. *That is, there is a density matrix $\rho \in \mathcal{M}_d(\mathbb{C})$ such that $T(\rho) = \rho$.*

In the following we are interested in linear maps, which are automatically continuous.

Proposition 6.8 (Positive fixed-points) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, positive, linear map and $X = T(X)$ a fixed point. If $X = \sum_{j=1}^4 c_j P_j$ is the decomposition of X into four positive operators $P_j \geq 0$ which arises from decomposing X first into Hermitian and anti-Hermitian parts and these further into orthogonal positive and negative parts, then*

$$T(P_j) = P_j \quad j = 1, \ldots, 4, \quad \text{are fixed points as well.} \tag{6.45}$$

PROOF For every Hermiticity preserving linear map we obviously have that the two Hermitian operators $X + X^\dagger$ and $i(X - X^\dagger)$ are fixed points if X is one. So let us assume that X is Hermitian and that $P_\pm \geq 0$ are its orthogonal positive an negative parts, i.e., $X = P_+ - P_-$ and $\operatorname{tr}[P_+ P_-] = 0$. The fixed point equation then reads $T(P_+) - T(P_-) = P_+ - P_-$ and we want to identify $T(P_\pm)$ with $P_\pm$. To this end, denote by Q the projection onto the support of P_+ and note that the fixed point equation together with the trace-preserving property of T implies that

$$\begin{aligned} \operatorname{tr}[P_+] &= \operatorname{tr}[Q(P_+ - P_-)] = \operatorname{tr}\left[Q\big(T(P_+ - P_-)\big)\right] & (6.46) \\ &\leq \operatorname{tr}[T(P_+)] = \operatorname{tr}[P_+]\,. & (6.47) \end{aligned}$$

Since equality has to hold for the inequality in Eq.(6.47), we have to have that $T(P_+)$ is supported on Q while $T(P_-)$ is orthogonal to it. Consequently, T preserves the orthogonality of $P_\pm$ and we have $T(P_\pm) = P_\pm$. □

For a map T on $\mathcal{M}_d(\mathbb{C})$ let us define the set of fixed points

$$\mathcal{F}_T := \big\{X \in \mathcal{M}_d(\mathbb{C}) \mid X = T(X)\big\}. \tag{6.48}$$

Note that if T is linear, then $\mathcal{F}_T$ is closed under linear combination, i.e., it is a vector space, and if T is Hermiticity preserving (in particular, if it is positive), then $\mathcal{F}$ is closed under Hermitian conjugation. The previous Prop.6.8 immediately implies that for trace-preserving quantum channels the set of fixed points is spanned by stationary density operators:

Corollary 6.5 (Linearly independent stationary states) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, positive, linear map. If the space $\mathcal{F}_T$ has dimension D, then there are D linearly independent density operators $\{\rho_\alpha \in \mathcal{M}_d(\mathbb{C})\}_{\alpha=1,\ldots,D}$ such that $T(\rho_\alpha) = \rho_\alpha$ and*

$$\mathcal{F}_T \;=\; \mathrm{span}\big\{\rho_1, \ldots, \rho_D\big\}. \tag{6.49}$$

Proposition 6.9 (Maximal support of fixed points) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, positive, linear map and T_∞ the projection onto its fixed point space as given in Eq.(6.14). Then every fixed point $X = T(X) \in \mathcal{M}_d(\mathbb{C})$ has support and range within the support space of the fixed point $T_\infty(\mathbb{1})$.*

Proof For positive X the claim follows from positivity of T_∞: since $0 \leq T_\infty(X) = X \leq ||X||_\infty T_\infty(\mathbb{1})$ the range (which equals the support) of X has to be contained in the one of $T_\infty(\mathbb{1})$ (See Douglas' theorem 5.1).

If X is not positive we can by Prop.6.8 decompose it into a linear combination of four positive fixed points $P_1, \ldots, P_4$. Using that $\mathrm{supp}(X)$ is in the linear span of $\bigcup_j \mathrm{supp}(P_j)$ together with the result for positive fixed points then shows $\mathrm{supp}(X) \subseteq \mathrm{supp}\big(T_\infty(\mathbb{1})\big)$. The same argument applies for the range of X. Alternatively, we may use that X is a fixed point iff $X^\dagger$ is, and that the range of X equals the support of $X^\dagger$. □

The following shows that the ranges of all stationary states play an exceptional role: every state which is supported within such a space will remain so after the action of the channel:

Proposition 6.10 (Stationary subspaces) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, positive, linear map, and let $Q \in \mathcal{M}_d(\mathbb{C})$ be the Hermitian projection onto the support of any stationary density operator $\sigma = T(\sigma) \in \mathcal{M}_d(\mathbb{C})$. Then $\mathrm{tr}\left[(\mathbb{1} - Q)T(Q)\right] = 0$ and for every density operator $\rho \in \mathcal{M}_d(\mathbb{C})$*

$$\rho \leq Q \quad \Rightarrow \quad T(\rho) \leq Q. \tag{6.50}$$

Proof By definition of Q we have that $Q \leq c\sigma \leq c'Q$ for some strictly positive numbers c and c'. Therefore $T(Q) \leq cT(\sigma) = c\sigma \leq c'Q$ which implies $\mathrm{tr}\left[(\mathbb{1} - Q)T(Q)\right] = 0$. Similarly, if $\rho \leq Q$, then $T(\rho) \leq c'Q$, i.e., $T(\rho)$ has support in the range of Q. Since $T(\rho)$ has eigenvalues at most one, this means that $T(\rho) \leq Q$. □

Projections onto stationary subspaces (in the sense of Eq.(6.50)) can be easily characterized in the Heisenberg picture, as well:

Proposition 6.11 (Stationary subspaces II) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, positive, linear map, and let $Q \in \mathcal{M}_d(\mathbb{C})$ be any Hermitian projection. Then the following statements are equivalent:*

1. *For every density matrix $\rho \in \mathcal{M}_d(C)$ with $\rho \leq Q$, we have that $T(\rho) \leq Q$ as well,*
2. *$T^*(Q) \geq Q$.*

PROOF 1. $\rightarrow$ 2.: define the orthogonal projection $P := (\mathbb{1} - Q)$. The assumption implies that $QT^*(P)Q = 0$ which in turn means that $PT^*(P)P = T^*(P)$. Using that $T^*(\mathbb{1}) = \mathbb{1}$, the latter can be rewritten as $T^*(Q) - PT^*(Q)P = Q$ from which $T^*(Q) \geq Q$ follows. 2. $\rightarrow$ 1. follows from the chain of inequalities

$$0 \leq \operatorname{tr}[T(\rho)P] = 1 - \operatorname{tr}[\rho T^*(Q)] \leq 1 - \operatorname{tr}[\rho Q] = 0.$$

□

A projection onto a stationary subspace is often called *sub-harmonic* and item 2. of the previous proposition is used as the defining property of this notion. Note that while the support projection of every stationary state is sub-harmonic (due to Props.6.10,6.11), the converse is not necessarily true: for instance $T^*(\mathbb{1}) \geq \mathbb{1}$ holds trivially for every trace-preserving T even if there is no stationary state of full rank.

If T is completely positive with Kraus decomposition $T(\cdot) = \sum_i K_i \cdot K_i^\dagger$, one can relate support spaces of fixed points to properties of the Kraus operators: suppose there is a stationary state with support $\mathcal{H} \in \mathbb{C}^d$. Then all Kraus operators have to preserve this subspace in the sense that $K_i\mathcal{H} \subseteq \mathcal{H}$. In particular, if $\rho = |\psi\rangle\langle\psi|$ is a pure state, then ψ has to be a simultaneous eigenvector of all Kraus operators.

We will now investigate the structure of the set of fixed points $\mathcal{F}$ of quantum channels. As a first step we exploit Props.6.9,6.10 in order to restrict the discussion to trace-preserving maps which have a full rank fixed point:

Lemma 6.4 (Restriction to full rank fixed points) *Let $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, (completely) positive, linear map, and $\mathcal{H} \subseteq \mathbb{C}^d$ the support space of $T_\infty(\mathbb{1})$. Then the linear map $\tilde{T} : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{H})$ defined via $\tilde{T}(X) := T(X \oplus 0)\big|_{\mathcal{H}}$ is trace-preserving and (completely) positive as well. Moreover, $\tilde{T}$ has a fixed point of full rank and*[4]

$$\mathcal{F}_T = \mathcal{F}_{\tilde{T}} \oplus 0. \tag{6.51}$$

PROOF Denote by $V : \mathcal{H} \rightarrow \mathbb{C}^d$ the isometry which embeds $\mathcal{H}$ into $\mathbb{C}^d$ so that $V^\dagger V = \mathbb{1}$ and $Q := VV^\dagger$ is the projection onto the range of $T_\infty(\mathbb{1})$. Then by definition $\tilde{T}(X) = V^\dagger T(VXV^\dagger)V$ so that $\tilde{T}$ inherits the property of being (completely) positive from T. In order to see that $\tilde{T}$ is trace-preserving note that $VXV^\dagger = X \oplus 0$ has support and range within the range of Q. Thus we can decompose it into four positive operators with the same property and apply Prop.6.10 which leads to

$$T(X \oplus 0) = \tilde{T}(X) \oplus 0, \tag{6.52}$$

so that the trace-preserving property is inherited from T as well. Finally, Eq.(6.52) also implies Eq.(6.51) since by Prop.6.9 all fixed points of T are of the form $X \oplus 0$. In particular, by definition of $\mathcal{H}$, there is one fixed point which has full rank on $\mathcal{H}$. □

[4]Here the '0' is supposed to act on the orthogonal complement of $\mathcal{H}$ in $\mathbb{C}^d$.

Before returning to trace-preserving maps we will discuss their duals–unital maps. This will provide an important prerequisite for the discussion of the trace-preserving case. We will also see that the structure of the fixed point set of a map simplifies a lot when the dual map has at least one fixed point of full rank. This is already visible in the dual analogue of Lem.6.4:

Lemma 6.5 *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, positive, linear map, and $\tilde{T}$ and V defined as in Lem.6.4. Let us further denote by $\mathcal{S}$ the set of linear unital maps which satisfy the Schwarz inequality in Eq.(5.2). Then*

$$A \in \mathcal{F}_{T^*} \quad \Rightarrow \quad V^\dagger A V \in \mathcal{F}_{\tilde{T}^*} \quad \textit{and} \tag{6.53}$$

$$X \in \mathcal{F}_{\tilde{T}^*} \quad \Leftrightarrow \quad V^\dagger T^*(VXV^\dagger)V = X. \tag{6.54}$$

$$T^* \in \mathcal{S} \quad \Rightarrow \quad \tilde{T}^* \in \mathcal{S}. \tag{6.55}$$

Proof Eq.(6.54) follows from inserting the definitions. For Eq.(6.53) note that $\tilde{T}^*(V^\dagger AV) = V^\dagger T^*(QAQ)V$ (using the notation of Lem.6.4 again). Taking the trace with an arbitrary $C \in \mathcal{B}(\mathcal{H})$ we obtain

$$\operatorname{tr}\left[C\tilde{T}^*(V^\dagger AV)\right] = \operatorname{tr}\left[T(VCV^\dagger)QAQ\right] \tag{6.56}$$

$$= \operatorname{tr}\left[T(VCV^\dagger)A\right] = \operatorname{tr}\left[CV^\dagger AV\right], \tag{6.57}$$

where the second equality follows from Prop.6.10 and the last step uses the assumption that A is a fixed point of T^*. Finally, the Schwarz inequality for $\tilde{T}^*$ follows from

$$V^\dagger T^*(VA^\dagger AV^\dagger)V \geq V^\dagger T^*(VA^\dagger V^\dagger)T^*(VAV^\dagger)V \tag{6.58}$$

$$\geq V^\dagger T^*(VA^\dagger V^\dagger)VV^\dagger T^*(VAV^\dagger)V. \tag{6.59}$$

Here the first inequality comes from first inserting $V^\dagger V = \mathbb{1}$ between $A^\dagger$ and A and then using the Schwarz inequality for T^* and the second inequality is due to $VV^\dagger \leq \mathbb{1}$. Written in terms of $\tilde{T}^*$ the above equations lead to the claimed $\tilde{T}^*(A^\dagger A) \geq \tilde{T}^*(A^\dagger)\tilde{T}^*(A)$. □

Due to the aforementioned issue we will first characterize the fixed point set of unital maps which have a dual with fixed point of full rank:

Theorem 6.12 (Structure of fixed points for unital Schwarz maps) *Let $T^* : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a unital map which (i) satisfies the Schwarz inequality Eq.(5.2) (e.g., a completely positive map) and (ii) is such that T has a full rank fixed point. Then the set of fixed points $\mathcal{F}_{T^*}$ is a *-algebra and has thus the form specified in Eq.(1.39) (without the zero-block).*

Proof Let $\rho \in \mathcal{M}_d(C)$ be a full rank fixed point of T. Note that by Prop.6.9 we can w.l.o.g. chose a positive definite $\rho > 0$. Denote by $A = T^*(A) \in \mathcal{M}_d(\mathbb{C})$ a fixed point of T^*. Then, using the fixed point properties and the fact that $A^\dagger$ is a fixed point of T^* as well we obtain

$$\operatorname{tr}\left[\big(T^*(A^\dagger A) - T^*(A^\dagger)T^*(A)\big)\rho\right] = 0. \tag{6.60}$$

Note that the expression in parentheses is positive semi-definite by the Schwarz inequality Eq.(5.2) and therefore, in fact, identical zero since ρ is assumed to be positive definite. Hence, we have equality in the Schwarz inequality so that we can apply Thm.5.4 which implies that if $A, X \in \mathcal{F}_{T^*}$, then $AX \in \mathcal{F}_{T^*}$ is a fixed point as well. Since $\mathcal{F}_{T^*}$ is also closed under linear combinations and Hermitian conjugation, it is a *-algebra. Evidently, it is unital since $T^*(\mathbb{1}) = \mathbb{1}$ holds by assumption. □

If T does not have a full rank fixed point, then $\mathcal{F}_{T^*}$ need not be an algebra. Together with Lem.6.5 we obtain a weaker structure in this case:

Corollary 6.6 *Let $T^* : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a unital linear map which satisfies the Schwarz inequality Eq.(5.2). Let $Q = Q^\dagger \in \mathcal{M}_d(\mathbb{C})$ be the projection onto the maximum rank fixed point of T, then the following set is a *-algebra:*

$$\big\{Y \in Q\mathcal{M}_d(\mathbb{C})Q \;\big|\; QT^*(Y)Q = Y\big\}. \tag{6.61}$$

Proof The stated set is exactly the one appearing in Eq.(6.54), i.e., the fixed point set of $\tilde{T}^*$. Since $\tilde{T}$ has, however, a fixed point of full rank by construction (see Lem.6.4) we can apply Thm.6.12 which proves the claim. □

The relation between fixed point sets of unital maps and *-algebras becomes more transparent for completely positive maps:

Theorem 6.13 (Fixed points and commutants) *Let $T^* : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a completely positive unital map with Kraus decomposition $T^*(\cdot) = \sum_i K_i^\dagger \cdot K_i$. If both $X, X^\dagger X \in \mathcal{F}_{T^*}$, then $[X, K_i] = 0$ for all i. Consequently, the largest *-subalgebra contained within $\mathcal{F}_{T^*}$ is given by the set*

$$\big\{X \in \mathcal{M}_d(\mathbb{C}) \;\big|\; \forall i : [X, K_i] = [X, K_i^\dagger] = 0\big\}. \tag{6.62}$$

Proof The assertion follows from the equation

$$\sum_i [X, K_i]^\dagger [X, K_i] = T^*(X^\dagger X) - X^\dagger T^*(X) - T^*(X)^\dagger X + X^\dagger X = 0$$

by exploiting the fixed point properties of X and $X^\dagger X$ together with the fact that the l.h.s. is a sum over positive terms which is zero iff all terms vanish individually. Hence every *-subalgebra of $\mathcal{F}_{T^*}$ is a subset of Eq.(6.62). Conversely, the set in Eq.(6.62) is a *-subalgebra within $\mathcal{F}_{T^*}$ on its own and thus the largest one. □

Finally, we discuss the structure of the fixed point set of trace-preserving maps:

Theorem 6.14 (Fixed points of trace-preserving maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a trace-preserving, positive, linear map for which T^* satisfies the Schwarz inequality in Eq.(5.2). Then there is a unitary $U \in \mathcal{M}_d(\mathbb{C})$ and a set of positive definite density matrices $\rho_k \in \mathcal{M}_{m_k}(\mathbb{C})$ such that the fixed point set of T is given by*

$$\mathcal{F}_T \;=\; U\left(0 \oplus \bigoplus_{k=1}^{K} \mathcal{M}_{d_k} \otimes \rho_k\right) U^\dagger, \tag{6.63}$$

for an appropriate decomposition of the Hilbert space $\mathbb{C}^d = \mathbb{C}^{d_0} \oplus \bigoplus_k \mathbb{C}^{d_k} \otimes \mathbb{C}^{m_k}$. $\mathcal{M}_{d_k}$ *stands for the full algebra of complex matrices on* $\mathbb{C}^{d_k}$.

PROOF We exploit that by Lem.6.4 $\mathcal{F}_T = 0 \oplus \mathcal{F}_{\tilde{T}}$ after an appropriate basis transformation which we can assign to a unitary U. Now $\tilde{T}$ has, by construction, a full rank fixed point and following Lem.6.4 it is trace-preserving and its dual satisfies the Schwarz inequality due to Lem.6.5. This implies by Thm.6.12 that $\tilde{T}_\infty^*$ projects onto a *-algebra. The general form of a positive map projecting onto a $^*-$algebra (i.e., a conditional expectation) is given in Eq.(1.40). Taking the dual of this map we arrive at $\tilde{T}_\infty$ projecting onto a set of the form in Eq.(6.63) (without the zero-block). Putting things together and using that we can always move the kernel of each ρ_k into the zero-block completes the proof. □

Note that Thm.6.14 implies that the fixed point $T_\infty(\mathbb{1})$ is (up to normalization) given by $U(0 \oplus \bigoplus_k \mathbb{1}_{m_k} \otimes \rho_k)U^\dagger$. Moreover, it shows that again the set of fixed points is closely related to a *-algebra (actually *is* a *-algebra w.r.t. a modified product):

Corollary 6.7 *Let* $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ *be a trace-preserving, positive, linear map for which* T^* *satisfies the Schwarz inequality in Eq.(5.2). Then for any density matrix* $\rho \in \mathcal{M}_d(\mathbb{C})$ *which is a maximum rank fixed point of* T *we have that the set* $\rho^{-1/2}\mathcal{F}_T\rho^{-1/2}$ *is a* **-algebra (with the inverse taken on the support of* ρ*).*

Another interesting consequence of Thm.6.14 is its implication according to Prop.6.10: the decomposition of the fixed point space into a direct sum in Eq.(6.63) implies the same decomposition into stationary subspaces. What is more, the $\mathcal{M}_{d_k}$'s give rise to subsystem which are invariant, i.e., on which T acts as the identity (up to the reversible rotation given by U).

Theorem 6.15 (Unique fixed points of full rank) *Let* $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ *be a trace-preserving quantum channel with Kraus decomposition* $T(\cdot) = \sum_i K_i \cdot K_i^\dagger$. *If for some* $n \in \mathbb{N}$ *we have that*

$$\text{span}\left\{\prod_{k=1}^{n} K_{i_k}\right\} \;=\; \mathcal{M}_d(\mathbb{C}), \tag{6.64}$$

i.e., homogeneous polynomials of the Kraus operators span the entire matrix algebra, then there exist a unique positive definite density matrix $\rho \in \mathcal{M}_d(\mathbb{C})$ *such that* $\mathcal{F}_T \propto \rho$.

PROOF Consider the map T^n, i.e., the n-fold concatenation of T. This has Kraus operators $R_I := \prod_{k=1}^n K_{i_k}$ with multi-index $I = (i_1, \ldots, i_n)$ and it satisfies $\mathcal{F}_T \subseteq \mathcal{F}_{T^n}$. Now for every $\psi \in \mathbb{C}^d$ we have that $T^n(|\psi\rangle\langle\psi|)$ has full rank since by assumption $\text{span}\{R_I|\psi\rangle\} = \mathbb{C}^d$. This implies that there cannot be any stationary density operator which has a kernel. Since we can always construct a rank deficient positive fixed point if the dimension of $\mathcal{F}_T$ is larger than one (see Cor.6.5), the assertion follows. □

Problem 13 (Fixed points of general positive maps) *Characterize the set of fixed points of unital, positive, linear maps which do not satisfy the Schwarz inequality.*

6.5 Cycles and recurrences

We will now have a closer look at the space corresponding to the peripheral spectrum (i.e., eigenvalues which are phases) of a positive linear map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ which is assumed to be either trace-preserving or unital. Denote the complex linear span of all the respective eigenspaces by

$$\mathcal{X}_T := \text{span}\big\{X \in \mathcal{M}_d(\mathbb{C})\big|\exists\varphi \in \mathbb{R} : T(X) = e^{i\varphi}X\big\}, \tag{6.65}$$

for which we will simply write $\mathcal{X}$ if the dependence on T is clear from the context. The questions which we address first are: what's the structure of $\mathcal{X}$? what's the structure of T restricted to $\mathcal{X}$? and what does this imply on the peripheral spectrum?

Note that evidently $\mathcal{X}_T = \mathcal{X}_{T_\varphi}$ if T_φ is constructed from T as defined in Eq.(6.13) by discarding all but those terms in the spectral decomposition which correspond to the peripheral spectrum. Recall that by Prop.6.3 if T is trace-preserving or unital, then T_φ will be (completely) positive if T is and that by Prop.6.2 there cannot any non-trivial Jordan-block associated to an eigenvalue of modulus one.

Proposition 6.12 (Asymptotic image) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive and trace-preserving linear map. Then $\mathcal{X}_T$ is the image of the projection T_ϕ, it is spanned by positive operators, and it is closed under the action of T. That is,*

1. $T_\phi\big(\mathcal{M}_d(\mathbb{C})\big) = \mathcal{X}_T$,
2. $\exists\{\rho_i \geq 0\} : \mathcal{X}_T = \text{span}\{\rho_i\}$,
3. $T\big(\mathcal{X}_T\big) = \mathcal{X}_T$.

Proof Note that by the proof of Prop.6.3 every $X \in \mathcal{X}_T$ is a fixed point of T_ϕ since there is an ascending sequence $n_i \in \mathbb{N}$ such that $\lim_{i\infty} T_\varphi^{n_i} = T_\phi$. Conversely, if $X = T_\phi(X)$ is such a fixed point, then it has to be a linear combination of eigenvectors of T which correspond to eigenvalues of modulus one. $\mathcal{X}_T$ thus coincides with the space of fixed points of T_ϕ which is in turn given by $T_\phi\big(\mathcal{M}_d(\mathbb{C})\big)$, proving 1.

2. follows from 1. together with Corr.6.5 and in order to arrive at 3. we just have to use the definition of $\mathcal{X}_T$ and express an element $X \in \mathcal{X}_T$ as a linear combination of peripheral eigenvectors of T. □

The following refines this Proposition by adding the assumption that the map under consideration satisfies the Schwarz inequality. Exploiting this fact one observes that $\mathcal{X}$ is a *-algebra (w.r.t. a modified product) and that T acts on it either by permuting blocks or by unitary conjugation on subsystems:

Theorem 6.16 (Structure of cycles) *Let $T:\mathcal{M}_d(\mathbb{C})\rightarrow\mathcal{M}_d(\mathbb{C})$ be a trace-preserving and positive linear map which fulfills the Schwarz inequality in Eq.(5.2).*

1. *There exists a decomposition of the Hilbert space $\mathbb{C}^d=\mathcal{H}_0\oplus\bigoplus_{k=1}^K\mathcal{H}_k$ into a direct sum of tensor products*[5] *$\mathcal{H}_k=\mathcal{H}_{k,1}\otimes\mathcal{H}_{k,2}$ and positive definite density matrices ρ_k acting on $\mathcal{H}_{k,2}$ such that*

$$\mathcal{X}_T = 0\oplus\bigoplus_{k=1}^K \mathcal{M}_{d_k}\otimes\rho_k, \tag{6.66}$$

where $\mathcal{M}_{d_k}$ is a full complex matrix algebra on $\mathcal{H}_{k,1}$ of dimension $d_k := \dim(\mathcal{H}_{k,1})$. That is, for every $X\in\mathcal{X}_T$ there are $x_k\in\mathcal{B}(\mathcal{H}_{k,1})$ such that

$$X = 0\oplus\bigoplus_k x_k\otimes\rho_k. \tag{6.67}$$

2. *There exist unitaries $U_k\in\mathcal{B}(\mathcal{H}_{k,1})$ and a permutation π, which permutes within subsets of $\{1,\ldots,K\}$ for which the corresponding $\mathcal{H}_k$'s have equal dimension, so that for every $X\in\mathcal{X}_T$ represented as in Eq.(6.67)*

$$T(X) = 0\oplus\bigoplus_{k=1}^K U_k x_{\pi(k)} U_k^\dagger\otimes\rho_k. \tag{6.68}$$

Proof The basic ingredients of the proof are those which already appeared in the proof of Thm.6.1 where we characterized maps with unit determinant. Exploiting Dirichlet's lemma on Diophantine approximations (Lem.6.1) we can find an ascending subsequence $n_i\in\mathbb{N}$ so that $\lim_{i\rightarrow\infty}T^{n_i}$ converges to a map I with eigenvalues which are either one or zero. Clearly, I is again a trace-preserving, positive linear map which satisfies the Schwarz inequality since all these properties are preserved under concatenation. Moreover, $\mathcal{X}_T=\mathcal{X}_I$ is the fixed-point space of I. So, statement 1. of the theorem follows from the structure of the space of fixed points provided in Thm.6.14.

2. The map $T^{-1}:=\lim_{i\rightarrow\infty}T^{n_i-1}$ is the inverse of T on $\mathcal{X}$ since by construction $T^{-1}T=I$. Hence, both T and T^{-1} map $\mathcal{X}\rightarrow\mathcal{X}$ in a bijective way. The crucial point here is that T^{-1} is a again a positive, trace-preserving map since it is constructed as a limit of such maps. To understand the consequences, consider any pure state in $\mathcal{X}$, i.e., a density operator $\sigma\in\mathcal{X}$ which has no non-trivial convex decomposition within $\mathcal{X}$. Then the image of σ under T has to be a pure state as well: assume this is not the case, i.e., $T(\sigma)=\sum_i\lambda_i\sigma_i$ is a non-trivial convex decomposition into states $\sigma_i\in\mathcal{X}$. Then applying T^{-1} to this equation leads to a contradiction since $\sigma=\sum_i\lambda_iT^{-1}(\sigma_i)$ is not pure. Consequently, both T and T^{-1} map pure states in $\mathcal{X}$ onto pure states. Note that a

[5] Note that we may have $\dim\mathcal{H}_0=0$ as well as $\dim\mathcal{H}_{k,i}=1$. Moreover, the direct sum in the decomposition does not necessarily correspond to a block structure in computational basis – just in *some* basis.

pure state can only have support in one of the K blocks, for instance $\sigma = x \otimes \rho_k$ where $x \in \mathcal{B}(\mathcal{H}_{k,1})$ is a rank one projection. Now we know that $T(\sigma) = x' \otimes \rho_{k'}$ for some k' and some rank one projection $x' \in \mathcal{B}(\mathcal{H}_{k',1})$. By continuity and the fact that T is a bijective linear map on $\mathcal{X}$, we have that within $\mathcal{X}$: (i) every element of block k is mapped to an element of the block k' (i.e., k' does not depend on x), and (ii) the blocks k and k' must have equal dimension $d_k = d_{k'}$. Therefore, there is a permutation π which permutes blocks of equal size so that $X \in \mathcal{X}$ is mapped to

$$T(X) = 0 \oplus \bigoplus_k T_k(x_{\pi(k)}) \oplus \rho_k,$$

with some linear maps $T_k : \mathcal{M}_{d_k} \to \mathcal{M}_{d_k}$. Since the latter have, together with their inverses, to be positive and trace-preserving they must be either matrix transpositions or unitary conjugations by Cor.6.2. Matrix transpositions are, however, ruled out by the requirement that T and thus each T_k is a Schwarz-map. □

6.6 Inverse eigenvalue problems

6.7 Literature

There are various extensions of Brouwer's fixed point theorem. *Schauder's theorem*, for instance, extends it to topological vector spaces, which may be of infinite dimension. That is, S becomes a convex, compact subset of a topological vector space. *Kakutani's theorem* is a generalization to upper semi-continuous set-valued functions (correspondences) T from a non-empty, compact, convex set $S \in \mathbb{R}^n$ into itself such that for all $x \in S$ the set $T(x)$ is non-empty and convex, too. Dealing with these and more refined fixed point theorems is part of *algebraic topology.*

Chapter 7

Semigroup Structure

7.1 Continuous one-parameter semigroups

In this section we discuss continuous time-evolution which is described by continuous one-parameter semigroups. We begin with a brief discussion of general properties of such semigroups and later, in subsection 7.1.2, consider those semigroups which are quantum channels.

7.1.1 Dynamical semigroups

For a set Σ of 'observables' or 'states', a family of maps $T_t : \Sigma \to \Sigma$, parameterized by $t \in \mathbb{R}_+$, is called *dynamical semigroup* if for all $t, s \in \mathbb{R}_+$

$$T_t T_s = T_{t+s} \quad \text{and} \quad T_0 = \text{id}. \tag{7.1}$$

Here 'dynamical' should remind us that we may think of T_t as time evolution for a time-interval $[0, t]$. The semigroup property $T_t T_s = T_{t+s}$ in turn means that the time evolution is both *Markovian* and *homogeneous*. That is, the time evolution operation neither depends on the history nor on the actual time.[1]

Continuity and differentiability In order to arrive at a reasonable theory we have to add some form of continuity to the algebraic definition in Eq.(7.1), meaning that T_t should depend continuously on t. At this point there are different choices - usually chosen by the nature of the problem rather than by us. Assuming that Σ is equipped with a norm, we say that $T_t \to T_{t_0}$ *converges strongly* when $t \to t_0$ if $||T_t(x) - T_{t_0}(x)|| \to 0$ for all $x \in \Sigma$ and we speak about *uniform convergence* or *convergence in norm* if $||T_{t_0} - T_t|| \to 0$ where

[1] Note that, from a nerdy mathematical point of view, the semigroup throughout this section is actually always the same, namely $(\mathbb{R}_+, +)$, i.e., the positive numbers. What changes is its representation.

the norm of maps is defined as $||T|| := \sup_{x\in\Sigma} ||T(x)||/||x||$.[2] Following the spirit of this exposition we will not make much use of this difference since our main focus lies on linear spaces Σ which are of finite dimension, so that these notions actually coincide. We will, however, occasionally comment on infinite dimensional analogues of the derived results and those will crucially depend on the form of continuity. As a rule of thumb the results derived for matrices tend to hold when we ask for uniform convergence and replace matrices by bounded operators.

From now on let us assume that Σ is a finite-dimensional vector space, say $\mathbb{C}^D$ for simplicity. For later purpose keep in mind that there might be additional structure, e.g. $\Sigma = \mathcal{M}_d(\mathbb{C})$ isomorphic to $\mathbb{C}^{d^2}$.

Not surprisingly, $T_t = e^{tL} = \sum_{k=0}^{\infty}(tL)^k/(k!)$ forms a dynamical semigroup for any $L \in \mathcal{M}_D(\mathbb{C})$ and it fulfills the differential equation

$$\frac{d}{dt}T_t = LT_t. \tag{7.2}$$

L is called the *generator* or *infinitesimal generator* of the semigroup.

Conversely, whenever Eq.(7.2) is fulfilled for a differentiable map $t \mapsto T_t \in \mathcal{M}_D(\mathbb{C})$ and $T_0 = \mathbb{1}$, then $T_t = e^{tL}$ with $L = \frac{d}{dt}T_t\big|_{t=0}$. Surprisingly, differentiability for a finite-dimensional dynamical semigroup is implied by continuity:

Proposition 7.1 (From continuous semigroups to differentiable groups) *Let $\{T_t \in \mathcal{M}_D(\mathbb{C})\}$ be a dynamical semigroup which is continuous in $t \in \mathbb{R}_+$. Then T_t is differentiable for $t \in \mathbb{R}_+$ and of the form $T_t = e^{tL}$ for some $L \in \mathcal{M}_D(C)$. Consequently, T_t can be embedded into a group by extending the range of t to $\mathbb{R}$ or $\mathbb{C}$.*

Proof Since $T_0 = \mathbb{1}$ and T_t is, by assumption, continuous in t,

$$M_\epsilon := \int_0^\epsilon T_s ds \tag{7.3}$$

will be invertible for sufficiently small $\epsilon > 0$. The idea is now to express T_t in terms of this integral expression and thereby showing that it is differentiable. To this end note that

$$T_t = M_\epsilon^{-1} M_\epsilon T_t = M_\epsilon^{-1} \int_0^\epsilon T_{s+t} ds \tag{7.4}$$

$$= M_\epsilon^{-1} \int_t^{t+\epsilon} T_s ds = M_\epsilon^{-1}\big(M_{t+\epsilon} - M_t\big). \tag{7.5}$$

Hence, T_t is differentiable and the exists a generator L so that $T_t = e^{tL}$ which evidently becomes a group if we extend the range of t to $\mathbb{R}$ or $\mathbb{C}$. □

This assertion remains true in infinite dimensions if we ask for uniform continuity. The generators are then bounded operators.

[2] There is also the notion of *weak convergence* but this coincides with strong convergence for dynamical semigroups. Needless to say, but nonetheless mentioned, uniform implies strong implies weak convergence in general.

Resolvents Recall from Eq.(6.17) that the *resolvent* of a matrix $L \in \mathcal{M}_D(\mathbb{C})$ is a matrix-valued function on the complex plane defined by

$$R(z) := (z\mathbb{1} - L)^{-1}. \tag{7.6}$$

The resolvent of a generator of a dynamical semigroup can be obtained via

$$R(z) = \int_0^\infty e^{-zs} T_s ds, \tag{7.7}$$

if $\mathrm{Re}(z) > \sup\{\mathrm{Re}(\lambda) | \lambda \in \mathrm{spec}(L)\}$. Conversely, if the resolvent of L is given we can obtain the dynamical semigroup via the two expressions

$$T_t = \frac{1}{2\pi i} \int_{\partial\Delta} e^{zt} R(z) dz \tag{7.8}$$

$$= \lim_{n\to\infty} (n/t)^n R(n/t)^n. \tag{7.9}$$

Note that in the Chauchy integral formula in Eq.(7.8) we have to choose $\Delta \supset \mathrm{spec}(L)$. Eq.(7.9) on the other hand is nothing but Euler's approximation $e^{tL} = \lim_{n\to\infty}(\mathbb{1} - (tL)/n)^{-n}$.

Some remarks on the above equations: it follows from Eq.(7.7) that if $z \in \mathbb{R}$ is sufficiently large, then $R(z)$ is an element of a convex cone (e.g., of positive or completely positive maps) if T_t is for all $t \in \mathbb{R}_+$. Conversely, if $R(t)$ is an element of a closed convex cone for large $t \in \mathbb{R}_+$ and this cone is closed under taking powers, then Eq.(7.9) implies that T_t is contained in this cone as well. Hence, positivity properties in which we will be interested further down are simultaneously reflected by the dynamical semigroup and the resolvent of its generator.

Another point worth mentioning is that if leave the realm of matrices for a moment and allow L to be unbounded, then the resolvent, where defined, remains bounded so that the above expressions enable us to construct dynamical semigroups from unbounded generators via bounded expressions. They are therefore a central tool in proving the *Hille-Yoshida theorem* which characterizes the generators of strongly continuous contraction semigroups. These generators are in general unbounded and merely defined on a dense subspace.

Perturbations and series expansions Consider two dynamical semigroups of matrices $T_t = e^{tL}$ and $T'_t = e^{tL'}$ with $\Delta := L' - L$ the difference of their generators. Here is a useful relation between the two:

Lemma 7.1 *Let $\{T_t \in \mathcal{M}_D(\mathbb{C})\}$ and $\{T'_t \in \mathcal{M}_D(\mathbb{C})\}$ be two dynamical semigroups and define $\Delta := \frac{d}{dt}(T'_t - T_t)\big|_{t=0}$ the difference of their generators. Then*

$$T'_t = T_t + \int_0^t T_{t-s} \Delta T'_s \, ds. \tag{7.10}$$

Proof Defining a function $f(s) := T_{t-s}T'_s$ and evaluating its derivative as $\frac{d}{ds}f(s) =: f'(s) = T_{t-s}(L' - L)T'_s$ we can express the difference of the two dynamical semigroups as

$$T'_t - T_t = f(t) - f(0) = \int_0^t f'(s)ds = \int_0^t T_{t-s}\Delta T'_s ds. \tag{7.11}$$

□

This simple relation has some remarkable consequences. One of them which is hard to overlook is an upper bound on the distance between the semigroups in terms of the distance of their generators:

Corollary 7.1 (Perturbation of generators) *Let $\{T_t \in \mathcal{M}_D(\mathbb{C})\}$ and $\{T'_t \in \mathcal{M}_D(\mathbb{C})\}$ be two dynamical semigroups and let Δ be the difference of their generators. Then for any norm and $t \in \mathbb{R}_+$*

$$||T'_t - T_t|| \leq t||\Delta|| \sup_{s,s' \in [0,t]} ||T_s||\, ||T'_{s'}||. \tag{7.12}$$

Note that for various applications which we may have in mind, this simplifies further as for instance the relevant norms of the dynamical semigroups on the right hand side in Eq.(7.12) are one (e.g., the cb-norm of quantum channels).

A second implication of the above Lemma is the *Dyson-Philips series* (which is just called *Dyson series* if both semigroups are unitary): by recursively inserting Eq.(7.10) into itself, we obtain

$$T'_t = \sum_{n=0}^{\infty} \tilde{T}_t^{(n)}, \quad \tilde{T}_t^{(n+1)} = \int_0^t T_{t-s}\Delta \tilde{T}_t^{(n)}\, ds \;\text{ with }\; \tilde{T}_t^{(0)} = T_t. \tag{7.13}$$

7.1.2 Quantum dynamical semigroups

After having analyzed general properties of 'unstructured' dynamical semigroups, we will now turn to dynamical semigroups of quantum channels, i.e., for any $t \in \mathbb{R}_+$ the map $T_t : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(C)$ of interest will be completely positive. A central question in this discussion is that about the structure of generators. It turns out that complete positivity of $T_t = e^{tL}$ is equivalent to a very similar property of the associated generator L, called *conditional complete positivity*:

Proposition 7.2 (Conditional complete positivity) *Let $L : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(C)$ be a linear map. Then the following properties are equivalent:*

1. *There is a completely positive map $\phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ and a matrix $\kappa \in \mathcal{M}_d(C)$ so that for every $\rho \in \mathcal{M}_d(\mathbb{C})$:*

$$L(\rho) = \phi(\rho) - \kappa\rho - \rho\kappa^\dagger. \tag{7.14}$$

2. *L is Hermiticity preserving and denoting by $|\Omega\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ an arbitrary maximally entangled state and by $P = \mathbb{1} - |\Omega\rangle\langle\Omega|$ the projection onto its orthogonal complement, we have*

$$P\big(L \otimes \mathrm{id}\big)\big(|\Omega\rangle\langle\Omega|\big)P \geq 0. \tag{7.15}$$

Proof Before proving the equivalence, let us make sure that the second property is well-defined in the sense that it doesn't depend on the choice of the maximally entangled state. To this end, recall that any two maximally entangled states Ω and Ω' are related via a local unitary $|\Omega'\rangle = (\mathbb{1} \otimes U)|\Omega\rangle$. Choosing a different maximally entangled state thus changes Eq.(7.15) by a unitary conjugation with $(\mathbb{1} \otimes U)$ which clearly doesn't change positivity.

1. $\rightarrow$ 2.: Inserting Eq.(7.14) into Eq.(7.15) and using that P annihilates $|\Omega\rangle$, gives $P\big(L \otimes \mathrm{id}\big)\big(|\Omega\rangle\langle\Omega|\big)P = P\big(\phi \otimes \mathrm{id}\big)\big(|\Omega\rangle\langle\Omega|\big)P$ which is positive by complete positivity of ϕ. Moreover, L is Hermiticity preserving since $L(\rho)^\dagger = L(\rho^\dagger)$.

2. $\rightarrow$ 1.: Since L is Hermiticity preserving, $\tau := (L \otimes \mathrm{id})\big(|\Omega\rangle\langle\Omega|\big)$ is Hermitian. Moreover, since $P\tau P \geq 0$ we can write $\tau = Q - |\psi\rangle\langle\Omega| - |\Omega\rangle\langle\psi|$ where $Q \geq 0$, when written in a basis containing Ω, has non-zero entries only in columns and rows orthogonal to Ω and ψ contains all the entries of τ in the column/row corresponding to Ω. By Prop.2.1 we can now define a completely positive map via $(\phi \otimes \mathrm{id})\big(|\Omega\rangle\langle\Omega|\big) := Q$ so that setting $(\kappa \otimes \mathbb{1})|\Omega\rangle := |\psi\rangle$ completes the proof. □

Using this equivalence together with the fact that by Prop.7.1 there exists always a generator, we can now derive the structure of the latter in cases where T_t is a semigroup of completely positive maps on $\mathcal{M}_d(\mathbb{C})$:

Proposition 7.3 (Completely positive dynamical semigroups) *Consider a family of linear maps $T_t : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$ for $t \in \mathbb{R}_+$. The following are equivalent:*

1. *T_t is a dynamical semigroup of completely positive maps which is continuous in t,*

2. *$T_t = e^{tL}$ for some conditional completely positive linear map $L : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$.*

Proof 1. $\rightarrow$ 2.: By Prop.7.1 there exists a generator $L : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$. Since T_t is, by assumption, completely positive for all $t \geq 0$ we may look at infinitesimal t for which this implies

$$0 \leq \big(e^{tL} \otimes \mathrm{id}\big)\big(|\Omega\rangle\langle\Omega|\big) = |\Omega\rangle\langle\Omega| + t(L \otimes \mathrm{id})\big(|\Omega\rangle\langle\Omega|\big) + o(t). \tag{7.16}$$

Exploiting that $o(t)/t \rightarrow 0$ as $t \rightarrow 0$ and projecting onto the orthogonal complement of Ω we obtain that L is Hermiticity preserving and fulfills Eq.(7.15).

2. $\rightarrow$ 1.: Clearly, e^{tL} is a continuous dynamical semigroup. In order to see complete positivity we use Eq.(7.14) and decompose the generator into two

parts $L = \phi + \phi_\kappa$ with $\phi_\kappa(\rho) := -\kappa\rho - \rho\kappa^\dagger$. From the Lie-Trotter formula we get

$$e^{tL} = \lim_{n\to\infty} \left(e^{t\phi/n} e^{t\phi_\kappa/n}\right)^n . \tag{7.17}$$

Since concatenations of completely positive maps are again completely positive, it is sufficient to show that $e^{t\phi/n}$ and $e^{t\phi_\kappa/n}$ are both completely positive. For $e^{t\phi/n}$ this follows from complete positivity of ϕ by Taylor expansion. For $e^{t\phi_\kappa/n}$ we invoke the matrix representation (see Eq.(2.21)) $\hat{\phi}_\kappa = -\kappa \otimes \mathbb{1} - \mathbb{1} \otimes \bar{\kappa}$ from which we get $e^{t\hat{\phi}_\kappa/n} = K \otimes \bar{K}$ with $K := e^{-t\kappa/n}$. Therefore $e^{t\phi_\kappa/n}(\rho) = K\rho K^\dagger$ is completely positive as well. □

Proposition 7.4 (Freedom in representation of generators) *Let $\phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a completely positive map with Kraus representation $\phi(\rho) = \sum_i L_i \rho L_i^\dagger$ and let $\kappa \in \mathcal{M}_d(\mathbb{C})$. Consider the generator of a continuous dynamical semigroup of completely positive maps which is represented by the pair (ϕ, κ) as in Eq.(7.14).*

1. *A matrix κ' and a completely positive map ϕ' with Kraus operators $\{L_i'\}$ represent the same generator if there are constant $c_i \in \mathbb{C}$ and $\lambda \in \mathbb{R}$ such that*

$$L_i' = L_i + c_i \mathbb{1}, \tag{7.18}$$
$$\kappa' = \kappa + \sum_i \bar{c}_i L_i + i\lambda\mathbb{1} + \frac{1}{2}\sum_i |c_i|^2 \mathbb{1}. \tag{7.19}$$

   In particular, there is always a choice in which all Kraus operators are trace-less.

2. *If (ϕ, κ) and (ϕ', κ') represent the same generator and the Kraus operators of ϕ as well as those of ϕ' are trace-less, then $\phi = \phi'$, $\kappa = \kappa' + i\lambda\mathbb{1}$ for some $\lambda \in \mathbb{R}$ and any two Kraus representations of ϕ and ϕ' are related via a unitary so that $L_i' = \sum_j U_{ij} L_j$ (where the smaller set of operators is padded by zeros).*

PROOF 1. That (ϕ, κ) and (ϕ', κ') lead to the same generator can be seen by direct inspection. 2. If we impose that ϕ has trace-less Kraus operators, then its Choi matrix has support and range on the orthogonal complement of the corresponding maximally entangled state Ω. The map $\rho \mapsto \kappa\rho + \rho\kappa^\dagger$, however, has a Choi matrix which (in an orthogonal basis containing Ω) has entries only in the row or column corresponding to Ω. Hence, the Choi matrix corresponding to the generator L has a unique decomposition into a part leading to ϕ and a part leading to κ. The statement that $\phi = \phi'$ and $\kappa = \kappa' + i\lambda\mathbb{1}$ then follows from the one-to-one correspondence between maps and their Choi matrices (see Prop.2.1) and the fact that the map $\rho \mapsto \kappa\rho + \rho\kappa^\dagger$ is invariant under adding imaginary multiples of the identity matrix to κ. Finally, the remaining unitary freedom in the Kraus decomposition follows from Thm.2.1. □

Theorem 7.1 (Generators for semigroups of quantum channels) *A linear map $L : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ is the generator of a continuous dynamical semigroup of trace-preserving, completely positive linear maps, iff it can be written in one (and thus any) of the following equivalent forms:*

$$\begin{aligned} L(\rho) &= \phi(\rho) - \kappa\rho - \rho\kappa^\dagger, \quad \textit{with} \quad \phi^*(\mathbb{1}) = \kappa + \kappa^\dagger && (7.20)\\ &= i[\rho, H] + \sum_j L_j \rho L_j^\dagger - \frac{1}{2}\{L_j^\dagger L_j, \rho\}_+ && (7.21)\\ &= i[\rho, H] + \frac{1}{2}\sum_j [L_j, \rho L_j^\dagger] + [L_j\rho, L_j^\dagger] && (7.22)\\ &= i[\rho, H] + \sum_{k,l=1}^{d^2-1} C_{l,k}\big([F_k, \rho F_l^\dagger] + [F_k\rho, F_l^\dagger]\big), && (7.23) \end{aligned}$$

where ϕ is a completely positive linear map, $H = H^\dagger$ Hermitian, $\{L_j \in \mathcal{M}_d(\mathbb{C})\}$ a set of matrices, $C \in \mathcal{M}_{d^2-1}(\mathbb{C})$ positive semi-definite and $F_1, \ldots, F_{d^2-1}$ a basis of the space of trace-less matrices in $\mathcal{M}_d(\mathbb{C})$.

Proof The first equation follows from Props.7.3,7.2 by imposing $\operatorname{tr}[L(\rho)] = \operatorname{tr}[\rho]$ for all ρ's. Eq.(7.21) is derived by identifying the L_j's with the Kraus operators of ϕ and

$$\kappa = iH + \frac{1}{2}\phi^*(\mathbb{1}). \qquad (7.24)$$

Eq.(7.22) is merely an occasionally convenient rewriting of Eq.(7.21). Eq.(7.23) follows from Eq.(7.23) by using first that the L_j's can be chosen trace-less by Prop.7.4 and then expanding them in terms of a basis as $L_j = \sum_{k=1}^{d^2-1} M_{j,k} F_k$. This leads to $C = M^\dagger M$ which is thus positive semi-definite. Conversely, if $C \geq 0$, then there is always an M such that $C = M^\dagger M$, which brings us back from Eq.(7.23) to Eq.(7.22). □

Some jargon and remarks are in order: H, i.e., the imaginary/anti-Hermitian part of κ, is the *Hamiltonian* which governs the *coherent* part of the evolution. Tee freedom $\kappa \to \kappa + i\lambda\mathbb{1}, \lambda \in \mathbb{R}$ is thus nothing but the irrelevance of a global energy shift. The map ϕ together with the real/Hermitian part of κ (which is determined by ϕ due to trace preservation) is the *incoherent* or *dissipative* part. As Prop.7.4 shows, the decomposition into a Hamiltonian and dissipative part is not unique but becomes so if we impose trace-less Kraus operators. In the context of quantum dynamical semigroups the Kraus operators of ϕ are often called *Lindblad operators* and the matrix C is often named *Kossakowski matrix*, in particular when the F's are chosen Hermitian and orthonormal ($\operatorname{tr}\left[F_k^\dagger F_l\right] = \delta_{kl}$) which is always possible.

Proposition 7.5 (Irreducibility implies primitivity) *Let $T_t = e^{Lt} : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ with $t \in \mathbb{R}_+$ be a dynamical semigroup of trace-preserving, completely positive linear maps. Then the following statements are equivalent:*

1. *There is a* $t_0 > 0$ *such that* T_{t_0} *is irreducible.*

2. T_t *is irreducible for all* $t > 0$.

3. T_t *is primitive for all* $t > 0$.

4. *There is a* $\rho_\infty > 0$ *such that for all density matrices* $\rho \in \mathcal{M}_d(\mathbb{C})$ *we have* $\lim_{t\to\infty} T_t(\rho) = \rho_\infty$.

5. L *has a one-dimensional kernel which consists out of multiples of a positive definite density operator* $\rho_\infty > 0$.

Proof Recall from Thm.6.4 that irreducibility is equivalent to having a non-degenerate eigenvalue one whose corresponding 'eigenvector', call it ρ_∞, is positive definite. Moreover, by Thm.6.6 all eigenvalues of an irreducible map which have magnitude one, denote them by $\{\lambda_1, \ldots, \lambda_m\}$ are roots of unity. Thus if T_{t_0} is irreducible, then $T_{(\pi t_0)}$ will be irreducible, too, since $e^{L\pi t}$ has still a non-degenerate eigenvalue one with corresponding eigenvector ρ_0. However, $\{\lambda_1^\pi, \ldots, \lambda_m^\pi\}$ can only remain roots of unity if $m = 1$, i.e., if there is a trivial peripheral spectrum in the first place. So T_t is irreducible for all $t > 0$ and thus primitive. This and the equivalence with point 4. follows from Thm.6.7 which shows that primitive channels can either be characterized as irreducible channels with trivial peripheral spectrum, or as channels for which the evolution eventually converges to a unique stationary state of full rank.

In order to see that 5. $\rightarrow$ 1. assume there would be imaginary eigenvalues of L. We could then choose a t_0 such that e^{Lt_0} still has a unique fixed point and since this has to be $\rho_\infty > 0$, e^{Lt_0} is irreducible. Conversely, in order to see 1. $\rightarrow$ 5 note that the kernel of L is contained in the fixed point space of any e^{Lt}. So if the kernel of L would be of higher dimension or contain a density matrix which is not full rank, then no e^{Lt} could be irreducible for any t. □

Proposition 7.6 (Reducible quantum dynamical semigroups) *Let* $T_t = e^{Lt} : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$ *with* $t \in \mathbb{R}_+$ *be a dynamical semigroup of trace-preserving, completely positive linear maps with generator* L *represented as in Thm.7.1. Then the following are equivalent:*

1. *There is a density matrix* ρ_0 *with non-trivial kernel and such that for all* $t \geq 0$: $\rho_0 = T_t(\rho_0)$.

2. *There is a density matrix* ρ_0 *with non-trivial kernel and such that* $L(\rho_0) = 0$.

3. *There is a Hermitian projector* $P \notin \{0, \mathbb{1}\}$ *such that for all* $t \geq 0$: $T_t\big(P\mathcal{M}_d(\mathbb{C})P\big) \subseteq P\mathcal{M}_d(\mathbb{C})P$.

4. *There is a basis in which all Kraus operators* L_j *and* κ *are block-upper triangular, i.e., they satisfy* $(\mathbb{1} - P)L_jP = (\mathbb{1} - P)\kappa P = 0$ *for some Hermitian projector* $P \notin \{0, \mathbb{1}\}$.

Proof The equivalence 1. $\leftrightarrow$ 2. can be verified by differentiation and Taylor expansion of the exponential respectively. 1. $\rightarrow$ 3. can be shown by taking P the projector onto the support of ρ_0. Using that $0 \leq Q \leq \rho_0$ implies $0 \leq T_t(Q) \leq \rho_0$ and therefore $T_t(Q) \subseteq P\mathcal{M}_d(\mathbb{C})P$ we can obtain 3. by exploiting that any element of $P\mathcal{M}_d(\mathbb{C})P$ is a linear combination of such Q's, i.e., positive semi-definite operators whose support space is contained in the one of ρ_0.

For 3. $\rightarrow$ 4. we use that $T_t(P)(\mathbb{1} - P) = 0$ which, by differentiation, implies

$$L(P)(\mathbb{1} - P) = 0. \tag{7.25}$$

We first multiply this equation from the left with $(\mathbb{1} - P)$ and represent L as in Thm.7.1. Using $P = P^2$ and abbreviating $X_j := (\mathbb{1} - P)L_jP$ this leads to $\sum_j X_j X_j^\dagger = 0$ and thus $X_j = 0$ for every j. Inserting this back into Eq.(7.25) then yields $(\mathbb{1} - P)\kappa P = 0$. This condition is equivalent to κ being block upper-triangular in a basis in which $P = (\mathbb{1} \oplus 0)$.

Conversely, 4. $\rightarrow$ 3. follows by noticing that the block upper-triangular structure of the L_j's and κ implies that $L\big(P\mathcal{M}_d(\mathbb{C})P\big) \subseteq P\mathcal{M}_d(\mathbb{C})P$. By Taylor expansion of the exponential e^{Lt} then preserves this subspace as well. Finally, 3. $\rightarrow$ 2. is verified as follows: choose a t_0 such that the fixed point space of e^{Lt_0} equals the kernel of L. This is the case whenever no purely imaginary eigenvalue of L is an integer multiple of $2\pi/t_0$. Since T_{t_0} by assumption is reducible, there is a rank deficient fixed point density matrix $\rho_0 = T_{t_0}(\rho_0)$ and therefore $L(\rho_0) = 0$. $\square$

In principle, Prop.7.6 gives necessary and sufficient conditions for reducibility and thus, by negation, for irreducibility. However, the validity of the conditions may not be easily decidable, so that the following necessary (but not sufficient) conditions may be useful:

Corollary 7.2 (Necessary conditions for relaxation) *Let $T_t = e^{Lt} : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$ with $t \in \mathbb{R}_+$ be a dynamical semigroup of trace-preserving, completely positive linear maps with generator L represented as in Thm.7.1. There is a positive definite density matrix ρ_∞ such that for every density matrix ρ we have $\lim_{t\rightarrow\infty} T_t(\rho) = \rho_\infty$ if one of the following statements is true:*

1. *The algebra generated by the set of Lindblad operators $\{L_j\}$ and κ (i.e., the space of all polynomials thereof) is the entire matrix algebra $\mathcal{M}_d(\mathbb{C})$.*

2. *The linear space spanned by the set of Lindblad operators is Hermitian*[3] *and its commutant contains only multiples of the identity.*

3. *The Kossakowski matrix has rank* $\mathrm{rank}(C) > d^2 - d$.

Proof Following Prop.7.5 we have to show that neither of three conditions is compatible with reducibility as discussed in Prop.7.6. So assume that T_t is reducible for one and thus any $t > 0$. Then by Prop.7.6, there is a basis in which the Lindblad operators $\{L_j\}$ and κ are block upper-triangular. Since this

[3]This means that for every $X = \sum_j x_j L_j$ there are $y_j \in \mathbb{C}$ such that $X^\dagger = \sum_j y_j L_j$.

property is preserved under multiplication and addition, this proves 1.. Regarding 2., if the linear space spanned by the L_j's is Hermitian, then reducibility would imply a block-diagonal structure (i.e., $L_j = PL_jP + (\mathbb{1} - P)L_j(\mathbb{1} - P)$) which is incompatible with the assumed trivial commutant. Finally, for 3. first note that the rank of $C \in \mathcal{M}_{d^2-1}(\mathbb{C})$ is independent of the basis we choose for the space of trace-less operators. If the evolution is reducible, then the block upper-triangular structure implies that the linear space $\mathcal{L}$ spanned by Lindblad operators has dimension at most $(d^2 - d)$. Hence, using a basis $\{F_k\}$ of the space of trace-less operators such that $F_1, \ldots, F_{d^2-d}$ is a basis for $\mathcal{L}$ then shows that $\text{rank}(C) \leq d^2 - d$. □

Note that the inequality in 3. in Corr.7.2 is optimal in the sense that there are reducible cases with $\text{rank}(C) = d^2 - d$.

The kernel of the Liouvillian L leads to a space of fixed points of the map e^{tL}. Hence, the analysis of the space of fixed points of quantum channels can be used to learn something about the structure of $kern(L)$:

Theorem 7.2 (Kernel of the Liouvillian) *Let $L : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a generator of a dynamical semigroup of trace-preserving, completely positive linear maps as presented in Thm.7.1 with Lindblad operators $\{L_i\}$ and Hamiltonian H. Then*

$$\{H, L_i, L_i^\dagger\}' \subseteq \{X \in \mathcal{M}_d(\mathbb{C}) | L^*(X) = 0\}. \tag{7.26}$$

If there is a positive definite element in the kernel of L, i.e., $\rho > 0, L(\rho) = 0$, then the converse inclusion is also true, i.e., the commutant equals the kernel.

Proof The inclusion in Eq.(7.26) follows directly from the characterization of generators in Thm.7.1. In order to arrive at the converse inclusion note that for some $t \in \mathbb{R}_+$ we have $L^*(X) = 0 \Leftrightarrow e^{tL^*}(X) = X$. Thm. 6.12 then implies that the set $\{X | L^*(X) = 0\}$ forms a *-algebra, in particular $L^*(A) = 0 \Rightarrow L^*(A^\dagger A) = 0$ and therefore

$$\sum_i [A, L_i]^\dagger [A, L_i] = \phi^*(A^\dagger A) + A^\dagger \phi^*(\mathbb{1})A - \phi^*(A^\dagger)A - A^\dagger \phi^*(A) = 0. \tag{7.27}$$

Here the last equality used that $L^*(A) = 0 \Rightarrow \phi^*(A) = A\kappa + \kappa^\dagger A$ (and the same for $A^\dagger A$ in place of A) together with $\phi^*(\mathbb{1}) = \kappa + \kappa^\dagger$.

Since the l.h.s. in Eq.(7.27) is a sum of positive terms, each of them has to be zero. So $[A, L_i] = 0$ and since we deal with a *-algebra, also $[A^\dagger, L_i] = [L_i^\dagger, A]^\dagger = 0$. Finally, we can exploit this in order to obtain

$$\kappa A = \phi^*(\mathbb{1})A - \kappa^\dagger A = \phi^*(A) - \kappa^\dagger A = A\kappa,$$

which implies $[H, A] = 0$. □

7.2 Literature

Suzuki [?] gives a detailed investigation of the speed of convergence of the Lie-Trotter formula and proves in particular that for all bounded operators A, B on any Banach space:

$$||e^{A+B} - \left(e^{A/n}e^{B/n}\right)^n|| \leq \frac{2}{n}\left(||A|| + ||B||\right)^2 e^{\frac{n+2}{n}(||A||+||B||)}. \tag{7.28}$$

Chapter 8

Measures for distances and mixedness

Theorem 8.1 (Russo-Dye) *...*

8.1 Norms

A *matrix norm* is a vector norm on the space of matrices $\mathcal{M}_{n,m}(\mathbb{C})$. That is, it is a positive functional which is convex (i.e., it satisfies the triangle inequality $||A+B|| \leq ||A||+||B||$ for all A, B) and such that $||aA|| = |a|\ ||A||$ for all $a \in \mathbb{C}$ and $(||A|| = 0 \Rightarrow A = 0)$. For square matrices some norms have the additional property of being *sub-multiplicative*, meaning that $||AB|| \leq ||A||\ ||B||$ holds. Sometimes this property is included in the definition of matrix norms. Equipped with a sub-multiplicative norm, $\mathcal{M}_d(\mathbb{C})$ becomes a *Banach algebra* (see Sec.1.6).

An important class of sub-multiplicative matrix norms on $\mathcal{M}_d(\mathbb{C})$ are those which are *unitarily invariant* in the sense that $||UAV|| = ||A||$ for all unitaries U and V. Since every element $A \in \mathcal{M}_d(\mathbb{C})$ has a singular value decomposition $A = U\text{diag}(s_1, \ldots, s_n)V$, a unitarily invariant norm depends solely on the *singular values* $\{s_i(A)\}$.[1] The two mostly used classes of unitarily invariant norms are:

- *Schatten p-norms:* $||A||_p := \left(\sum_{i=1}^d [s_i(A)]^p\right)^{1/p}$ for any $p \geq 1$,
- *Ky Fan k-norms:* $||A||_{(k)} := \sum_{i=1}^k s_i(A)$ where $k = 1, \ldots, d$ and the singular values are decreasingly ordered.

Both classes have a monotonicity property in their parameter. Evidently, $||A||_{(k)} \geq$

[1]In fact, unitarily invariant norms on $\mathcal{M}_d$ are in one-to-one correspondence with norms $f : \mathbb{R}^d \to \mathbb{R}^+$ which are (i) permutational invariant and (ii) depend only on absolute values. Such a vector norm f is called *symmetric gauge function* and it give rise to a unitarily invariant norm on $\mathcal{M}_d \ni A$ via $||A|| = f[(s_1(A), \ldots, s_d(A))]$.

$||A||_{(k')}$ for $k \geq k'$, but also

$$||A||_1 \geq ||A||_p \geq ||A||_{p'} \geq ||A||_\infty, \tag{8.1}$$

for all $1 \leq p \leq p'$. Moreover, $p < p'$ leads to a strict inequality iff A has rank at least two.[2] The norm $||A||_\infty = ||A||_{(1)}$ is called the *operator norm* and it can be defined alternatively as $||A||_\infty = \sup_{\psi,\phi} |\langle\psi|A|\phi\rangle|$ with supremum over all unit vectors $\phi, \psi \in \mathbb{C}^d$. With the operator norm $\mathcal{M}_d(\mathbb{C})$ becomes a C^*-algebra as the C^*-equality $||A^\dagger A|| = ||A||^2$ is satisfied (see Sec.1.6). This property uniquely identifies the operator norm. The norm $||A||_1 = ||A||_{(d)} = \mathrm{tr}\,[|A|]$ is called the *trace norm*. Trace norm and operator norm are the largest/smallest norms among all unitarily invariant norms in the sense that $||A||_\infty \leq ||A||/||\rho|| \leq ||A||_1$, where $\rho = |1\rangle\langle 1|$ takes care of normalization. The norm $||A||_2$ is called *Hilbert-Schmidt* norm or *Frobenius norm*. It can be regarded as arising from the Hilbert-Schmidt scalar product (see Sec.2.3) and it coincides with the l_2-norm. The latter is defined within the class of l_p norms $||A||_{l_p} := \big(\sum_{kl} |A_{kl}|^p\big)^{1/p}$ for $p \geq 1$.[3] They are bounded by the Schatten p-norms (which one may call L_p-norms in this context) via $||A||_{l_p} \geq ||A||_p$ for $p \in [1,2]$ and with inequality in the opposite direction for $p \geq 2$. Apart from l_2 the l_p-norms are not unitarily invariant.

Besides being sub-multiplicative unitarily invariant norms exhibit some very useful properties which are collected in the following:

Theorem 8.2 (Unitarily invariant norms) *Let $||\cdot||$ be any unitarily invariant norm on $\mathcal{M}_d(\mathbb{C})$. Then the following holds for $A, B \in \mathcal{M}_d(\mathbb{C})$:*

1. Hölder's inequality *for $p \geq 1$ and $p^{-1} + q^{-q} = 1$:*

$$||AB|| \leq ||\,|A|^p||^{1/p}\;||\,|B|^q||^{1/q}. \tag{8.2}$$

2. Lidskii's inequalities: *If A and B are Hermitian and $E^\downarrow(A)$, $E^\uparrow(A)$ denote the diagonal matrices with decreasingly/increasingly ordered eigenvalues of A respectively, then*[4]

$$||A - B|| \left\{ \begin{array}{l} \leq ||E^\downarrow(A) - E^\uparrow(B)|| \\ \geq ||E^\downarrow(A) - E^\downarrow(B)|| \end{array} \right. . \tag{8.3}$$

3. *If AB is normal, then $||AB|| \leq ||BA||$,*

4. *For every non-negative concave function f on $[0,\infty)$ and all $A, B \geq 0$:*

$$||f(A) + f(B)|| \geq ||f(A+B)||. \tag{8.4}$$

The inequality is reversed if f is non-negative and convex on $[0,\infty)$ with $f(0) = 0$.

---

[2] In infinite dimensions the analogue of Eq.(8.1), in particular $||A||_1 \geq ||A||_2 \geq ||A||_\infty$ gives rise to the inclusion *trace class operators $\subset$ Hilbert-Schmidt class operators $\subset$ bounded operators.*

[3] Note that in the definition of the l_p-norms as well as for the Schatten p-norms $p \geq 1$ is crucial for convexity.

[4] If one drops the restriction on A and B to be Hermitian and replaces eigenvalues by singular values in the diagonal matrices, then still the lower bound holds true.

5. Ky Fan dominance*: if we have* $||A||_{(k)} \geq ||B||_{(k)}$ *for all Ky-Fan norms* $(k = 1, \ldots, d)$*, then* $||A|| \geq ||B||$ *for every unitarily invariant norm.*

6. Araki-Lieb-Thirring inequalities: *If* $A, B \geq 0$ *and* $p \geq 1$*, then*

$$||(BAB)^p|| \leq ||B^p A^p B^p||, \qquad ||\,|BA|^p\,|| \leq ||B^p A^p||, \tag{8.5}$$

   and the reversed inequalities hold for $p \in (0, 1]$.

Applying Hölder's inequality to the trace norm and using that $|\mathrm{tr}\left[A^\dagger B\right]| \leq \mathrm{tr}\left[|A^\dagger B|\right]$ leads to a useful Cauchy-Schwarz type version of the inequality, namely:

$$\left|\mathrm{tr}\left[A^\dagger B\right]\right| \leq ||A||_p ||B||_q. \tag{8.6}$$

For $p = 2$ this is indeed the Chauchy-Schwarz inequality for the Hilbert-Schmidt inner product. Another simple application of Hölder's inequality is a relation between different p-norms opposite to the direction of Eq.(8.1): setting $B = \mathbb{1}$ in Eq.(8.2) and applying it to the p-norm we arrive at

$$||A||_p \leq d^{\frac{1}{p} - \frac{1}{p'}}\ ||A||_{p'}, \qquad \text{for } p' \geq p \geq 1. \tag{8.7}$$

In a similar vein we can derive the following moment inequality which generalizes the well known variance inequality $\langle X^2 \rangle \geq \langle X \rangle^2$ known for Hermitian operators:

Proposition 8.1 (Moment inequalities) *Consider* $1 \leq p \leq p'$ *and* $\rho, X \in \mathcal{M}_d(\mathbb{C})$ *where* ρ *is a density operator and* $X \geq 0$*. Then*

$$\mathrm{tr}\left[\rho X^p\right]^{p'} \leq \mathrm{tr}\left[\rho X^{p'}\right]^p. \tag{8.8}$$

Proof This follows from Hölder's inequality $||AB||_1 \leq ||A||_{\tilde{p}} ||B||_{\tilde{q}}$ by setting $\tilde{q} = p'/p, A = \rho^{1/\tilde{p}}$ and $B = \rho^{\frac{p}{2p'}} X^p \rho^{\frac{p}{2p'}}$. □

Variational characterization of norms: In the same spirit as the extremal eigenvalue of a Hermitian matrix can be obtained variationally, we can characterize the Schatten p-norms and the Ky-Fan norms as the solution of optimization problems. For the latter it holds for instance that

$$||A||_{(k)} = \sup\left\{|\mathrm{tr}\left[A^\dagger B\right]| \;\middle|\; B^\dagger B \leq \mathbb{1},\ \mathrm{rank}(B) \leq k\right\}. \tag{8.9}$$

Inequality (8.6) is the basis for an analogous variational characterization of the Schatten p-norms:

Theorem 8.3 (Variational ways to p-norms) *Let* $A, P \in \mathcal{M}_d(\mathbb{C})$ *with* $P \geq 0$ *and* $p^{-1} + q^{-1} = 1$ *with* $p \geq 1$*. Then*

$$\begin{aligned}
||A||_p &= \sup\left\{|\mathrm{tr}\left[A^\dagger B\right]| \;\middle|\; ||B||_q = 1\right\}, &(8.10)\\
||A||_1 &= \sup\left\{|\mathrm{tr}\left[A^\dagger U\right]| \;\middle|\; UU^\dagger = \mathbb{1}\right\}, &(8.11)\\
||A||_\infty &= \sup\left\{|\langle\psi|A|\phi\rangle| \;\middle|\; ||\psi|| = ||\phi|| = 1\right\}, &(8.12)\\
||P||_p^p &= \frac{1}{p}\sup_{X \geq 0}\left[\mathrm{tr}\left[PX\right] - \frac{1}{q}||X||_q^q\right]. &(8.13)
\end{aligned}$$

Proof In order to arrive at Eq.(8.10) first note that the supremum is an upper bound by Eq.(8.6). The supremum is attained if we set $B = UP^{p-1}\mathrm{tr}\left[P^p\right]^{-1/q}$ constructed from the polar decomposition of $A = UP$. Eqs.(8.11,8.12) are special cases which arise from (8.10) due to the fact that the set of operators with $||B||_q \leq 1$ is convex and, since we maximize a convex functional, the maximum is attained at an extreme point. The extreme points are the set of unitaries for $q = 1$ and normalized rank-one operators for $q = \infty$, respectively.
... ... □

Eq.(8.10) is the reason why p-norms and q-norms are called *dual* to each other if $p^{-1} + q^{-1} = 1$. In fact, for every norm $||\cdot||$ on $\mathcal{M}_{n,m}$ we can define a *dual norm* $||\cdot||^D$ w.r.t. the Hilbert Schmidt inner product via

$$||A||^D := \sup\{|\mathrm{tr}\left[A^\dagger B\right]| \;\big|\; B \in \mathcal{M}_{n,m}(\mathbb{C}),\ ||B|| = 1\}. \tag{8.14}$$

The dual norm is unitarily invariant iff $||\cdot||$ is.

8.2 Entropies

8.2.1 Information theoretic origin

8.2.2 Mathematical origin

axiomatization ..

$$S(\rho) = -\frac{\partial}{\partial p}\mathrm{tr}\left[\rho^p\right]\Big|_{p=1}. \tag{8.15}$$

8.2.3 Physical origin

Boltzmann, time asymmetry & co., relation to the above, statistical interpretation

8.3 Majorization

Majorization is a stronger means of saying that a probability distribution/density operator is more mixed than another one. While entropies and the respective norms provide an ordering in the space of probability distributions/density operators, majorization leaves us with a partial order. We will, in fact, see that one distribution is more mixed than another in the majorization sense iff this holds true for all entropy-type functionals.

Consider two vectors $x, y \in \mathbb{R}^d$. We say that x is majorized by y, and write $x \prec y$, iff for the sum of decreasingly ordered components we have

$$\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow, \quad \forall k = 1, \ldots, d \text{ with equality for } k = d. \tag{8.16}$$

Note that if x describes any discrete probability distribution, then

$$\left(\frac{1}{d},\ldots,\frac{1}{d}\right) \prec x \prec (1,0,\ldots,0), \tag{8.17}$$

consistent with our intuition about what is the 'maximally mixed' distribution and which is a most 'pure' distribution. If only the d inequalities in Eq.(8.16) are satisfied but equality for $k = d$ is not required, then we say that x is *weakly submajorized* by y and we write $x \prec_w y$.

The basic results in the context of majorization are:

Theorem 8.4 (Majorization) *Let $x, y \in \mathbb{R}^d$. The following statements are equivalent:*

1. *x is majorized by y,*
2. *$x = My$ for some doubly stochastic[5] matrix $M \in \mathcal{M}_d(\mathbb{R})$,*
3. *$x = My$ where $M_{kl} = O_{kl}^2$ for some real orthogonal matrix[6] $O \in \mathcal{M}_d(\mathbb{R})$,*
4. *$\sum_i f(x_i) \leq \sum_i f(y_i)$ for all convex functions $f : \mathbb{R} \to \mathbb{R}$.*

Theorem 8.5 (Submajorization) *Let $x, y \in \mathbb{R}^d$. The following statements are equivalent:*

1. *$x \prec_w y$*
2. *$x = My$ for some doubly substochastic[7] matrix $M \in \mathcal{M}_d(\mathbb{R})$,*
3. *There exists a vector $z \in \mathbb{R}^d$ such that $z \prec y$ and $x_i \leq z_i$ for all i,*
4. *$\sum_i f(x_i) \leq \sum_i f(y_i)$ for all monotonically increasing convex functions $f : \mathbb{R} \to \mathbb{R}$.*

Theorem 8.6 (Birkhoff) *The set of doubly stochastic matrices in $M_d(\mathbb{R})$ is a convex polytope. Its extreme points are the permutation matrices. Similarly, the set of doubly substochastic matrices is a convex polytope. Its extreme points are all matrices which have at most one entry 1 in each column and each row and all other entries zero.*

Excursion 8.1 (Diagonals, eigenvalues, singular values) *Consider a Hermitian matrix $H \in \mathcal{M}_d(\mathbb{C})$ with eigenvalues $\{\lambda_i\}$ and diagonal elements $D_i := H_{ii}$. Using that we can diagonalize H via a unitary U, we get the relation $D_i = \sum_j |U_{ij}|^2 \lambda_j$ and since $[|U_{ij}|^2]$ is doubly stochastic we have*

$$\lambda \succ D. \tag{8.18}$$

[5]Doubly stochastic means that M is non-negative ($M_{kl} \geq 0$) and $\sum_k M_{kl} = \sum_j M_{ij} = 1$ for all l, i.

[6]If $U \in \mathcal{M}_d(\mathbb{C})$ is any unitary, then the matrix with components $M_{kl} = |U_{kl}|^2$ is doubly stochastic and called a *unistochastic* matrix. If U real orthogonal, then M is called *orthostochastic*.

[7]Doubly substochastic means that M is non-negative and all columns and rows have a sum which is at most 1.

Given vectors $D,\lambda \in \mathbb{R}^d$ Eq.(8.18) is thus a necessary condition for the existence of a Hermitian matrix with eigenvalues λ and diagonal entries D. As shown by Horn it is also a sufficient one. If $\lambda, D \in \mathbb{C}^d$ are complex, then Mirsky showed that there exists a complex matrix in $\mathcal{M}_d$ with eigenvalues λ_i and diagonal entries D_i iff $\sum_i \lambda_i = \sum_i D_i$, i.e., the trace remains the only relevant condition. In a similar vein we can ask for the relation between eigenvalues and singular values and obtain again a majorization like condition: let $s \in \mathbb{R}^d_+$ and $\lambda \in \mathbb{C}^d$, then there exists a matrix with singular values s_i and eigenvalues λ_i iff

$$\prod_{i=1}^k s_i \geq \Big|\prod_{i=1}^k \lambda_i\Big| \quad \forall k = 1,\ldots,d \quad \textit{with equality for } k=d, \tag{8.19}$$

where both s and $|\lambda|$ are arranged in decreasing order. That is, if there is no zero element we can express this as $\log s \succ \log|\lambda|$ which, in particular implies $|\lambda| \prec_w s$.

The concept of majorization can be applied to pairs of matrices in $\mathcal{M}_d(\mathbb{C})$ by applying it to the vectors containing either the singular values or (in particular, for Hermitian matrices) the eigenvalues in decreasing order. Let for instance $s(A), s(B) \in \mathbb{R}^d$ be the vectors containing the singular values of two arbitrary complex matrices $A, B \in \mathcal{M}_d(\mathbb{C})$. By Ky Fan dominance (Thm.8.2) we get that

$$s(A) \succ s(B) \;\Leftrightarrow ||A||_1 = ||B||_1 \text{ and } ||A|| \geq ||B|| \text{ for all u.i. norms.} \tag{8.20}$$

In the following we will apply majorization to Hermitian matrices and write $A \succ B$ iff the corresponding majorization relation holds for the eigenvalues of A and B.

Theorem 8.7 (Majorization between density matrices) *Let ρ, σ be two density matrices in $\mathcal{M}_d(\mathbb{C})$. Then the following are equivalent:*

1. *$\rho \succ \sigma$,*
2. *$\mathrm{tr}\,[f(\rho)] \geq \mathrm{tr}\,[f(\sigma)]$ for all convex functions $f : \mathbb{R} \to \mathbb{R}$,*
3. *there exists a set of unitaries $\{U_i \in \mathcal{M}_d(\mathbb{C})\}$ and probabilities p_i such that*

$$\sigma = \sum_i p_i U_i \rho U_i^\dagger. \tag{8.21}$$

Proof The equivalence 1. $\leftrightarrow$ 2. is an immediate application of Thm.8.4. For 3. $\to$ 1. we can w.l.o.g. assume that ρ and σ are both diagonal since the diagonalizing unitaries can be incorporated in the U_i's. Then $\sigma_{kk} = \sum_l M_{kl}\rho_{ll}$ where $M_{kl} = \sum_i p_i |\langle k|U_i|l\rangle|^2$ is doubly stochastic and thus $\rho \succ \sigma$ by Thm.8.4. For the converse 1. $\to$ 3. we again use the freedom of assuming the ρ, σ are both diagonal. Then by Thm.8.4 and Birkhoff's Thm.8.6 we can write $\sigma_{kk} = \sum_{i,k,l} p_i \langle k|P_i|l\rangle^2 \rho_{ll}$ where the P_i's are permutations and the p_i's probabilities. Written in terms of the matrices this is $\sigma = \sum_i p_i P_i \rho P_i^T$ which completes the proof. □

A similar result holds if we go from a fixed pair of density operators to all pairs which are related by a positive map:

Theorem 8.8 (Majorization and doubly stochastic positive maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive linear map. Then the following are equivalent:*

1. *T is trace preserving and unital,*
2. *For all density operators ρ it holds that $\rho \succ T(\rho)$,*
3. *For all density operators ρ and convex functions $f : \mathbb{R} \to \mathbb{R}$ it holds that $\mathrm{tr}\,[f(\rho)] \geq \mathrm{tr}\,\big[f\big(T(\rho)\big)\big]$.*

Proof The equivalence 2. $\leftrightarrow$ 3. follows from Thm.8.4 in the same way as it was used for the previous Thm.8.7. For 2. $\to$ 1. note that the trace preserving property is implied by the equality condition in the definition of majorization (see Eq.(8.16)). Unitaliy follows from assuming that $\mathbb{1} \succ T(\mathbb{1})$ together with the fact that the maximally mixed distribution only majorizes itself (see Eq.(8.17)). Finally, in order to show 1. $\to$ 2. let us consider an arbitrary density operator ρ with eigenvalue decomposition $\rho = V^\dagger \lambda(\rho) V$ and similarly $T(\rho) = U^\dagger \lambda(T(\rho)) U$, where $U, V \in \mathcal{M}_d(\mathbb{C})$ are unitaries and the λ's denote the diagonal matrices containing the respective eigenvalues. Define a map $T' : \mathcal{M}_d \to \mathcal{M}_d$ via $T'(X) = UT(VXV^\dagger)U^\dagger$. By construction we have now $\lambda(T(\rho)) = T'(\lambda(\rho))$ which becomes $\lambda(T(\rho))_{ii} = \sum_j M_{ij} \lambda(\rho)_{jj}$ if we introduce $M_{ij} := \langle i | T'(|j\rangle\langle j|) | i \rangle$. The assertion now follows from the fact that M is a doubly stochastic matrix, which is in turn implied by T' being a trace-preserving, unital and positive map. □

As a consequence we get that doubly stochastic quantum channels are exactly those for which the entropy is non-decreasing:

Corollary 8.1 (Entropy increasing maps) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive and trace-preserving linear map and $h : \mathbb{R} \to \mathbb{R}$ any strictly concave function.*[8] *Then T is unital iff for all density operators $\rho \in \mathcal{M}_d(\mathbb{C})$ we have that*

$$\mathrm{tr}\,\big[h\big(T(\rho)\big)\big] \geq \mathrm{tr}\,[h(\rho)]\,. \tag{8.22}$$

Proof If T is unital, then the assertion follows from Thm.8.8 for every (not necessarily strictly) concave function h. Conversely, if h is strictly concave, then $\mathrm{tr}\,[h(\rho)]$ has $\rho = \mathbb{1}/d$ as a unique maximizer within the set of density operators. So if we assume Eq.(8.22), then necessarily $T(\mathbb{1}) = \mathbb{1}$. □

We will now use the developed tools in order to formalize the statement that any classical probability distribution obtained from measuring a quantum state is always at least as mixed as the density operator from which it has been obtained:

Theorem 8.9 (Measurements and convex functions)
Let $\{P_i \in \mathcal{M}_d(\mathbb{C})\}_{i=1,\ldots,m}$ with $m \geq d$ be a POVM (i.e., $P_i \geq 0$ and $\sum_i P_i = \mathbb{1}$)

[8] 'Strictly concave' means that for all $x_1 \neq x_2$ and $\lambda \in (0,1)$ it holds that $h\big(\lambda x_1 + (1-\lambda)x_2\big) > \lambda h(x_1) + (1-\lambda)h(x_2)$.

and $\rho \in \mathcal{M}_d(\mathbb{C})$ a density operator. Assume that the POVM is symmetric in the sense that $\mathrm{tr}[P_i] = d/m$ and denote the probabilities obtained from a measurement on ρ by $p_i := \mathrm{tr}[P_i\rho]$. For every convex function $f : \mathbb{R} \to \mathbb{R}$ which satisfies $f(0) \leq 0$ we have

$$\mathrm{tr}[f(\rho)] \geq \sum_{i=1}^{m} f(p_i). \tag{8.23}$$

For $m = d$ the same holds true without $f(0) \leq 0$ being required.

Proof We want to make use of the implication 1. → 3. in Thm.8.8. To this end we first embed everything in an m-dimensional space and set for instance $\rho' := \rho \oplus 0 \in \mathcal{M}_m$. Moreover, we define a linear map $T : \mathcal{M}_m \to \mathcal{M}_m$ via

$$T(X) := \frac{\mathbb{1}_m}{m}\mathrm{tr}[X(0 \oplus \mathbb{1}_{m-d})] + \sum_{i=1}^{m} |i\rangle\langle i| \, \mathrm{tr}[X(P_i \oplus 0)]. \tag{8.24}$$

Note that T is positive, unital and trace-preserving as required by Thm.8.8. Then

$$\begin{aligned} \mathrm{tr}[f(\rho)] &\geq \mathrm{tr}[f(\rho')] \geq \mathrm{tr}\left[f\big(T(\rho')\big)\right], \\ &= \sum_{i=1}^{m} f(p_i), \end{aligned} \tag{8.25}$$

where the first inequality comes from $f(0) \leq 0$ which is only required if $m > d$ since $\rho' = \rho$ for $m = d$. □

Note that for von Neumann measurements, i.e., when the P_i's are one-dimensional orthogonal projections, the above theorem essentially reduces to saying that the eigenvalues of ρ majorize the diagonal entries (see Eq.(8.18)). We also remark that Thm.8.9 is no longer true in general if we drop any of the three assumptions: $f(0) \leq 0$, $m \geq d$ or $\mathrm{tr}[P_i] = d/m$.

8.4 Divergences and quasi-relative entropies

In classical probability theory, one way of quantifying how two probability distributions p and q differ from each other (other than using norms and metrics) is to use *Csiszar's f-divergence* measures

$$C_f(p|q) := \sum_{x \in X} q_x f\Big(\frac{p_x}{q_x}\Big), \tag{8.26}$$

where $f : \mathbb{R}^+ \to \mathbb{R}$ is any continuous convex function. Every such C_f is jointly convex in its arguments and monotone in the sense that $C_f(p|q) \geq C_f\big(t(p)|t(q)\big)$ for any map $t : p \mapsto p'$ characterized by a partitioning of the set of events X into disjoint X_i's so that $p'_i = \sum_{x \in X_i} p_x$. If $f(1) = 0$ then $C_f(p|p) = 0$ and if $f(x)$ is strictly convex at $x = 1$ then $C_f(p|q) = 0$ implies $p = q$. Prominent examples of f-divergences are:

1. *Kullback Leibler divergence* or *relative entropy* for $f(x) = x \log x$:

$$Q_{KL}(p|q) = \sum_x p_x \log \frac{p_x}{q_x}, \tag{8.27}$$

2. *Relative α-entropy*[9] for $f(x) = (1 - x^\alpha)/(\alpha(1-\alpha))$ with $\alpha \in (0, 1)$:

$$Q_\alpha(p|q) = \Big(1 - \sum_x p_x^\alpha q_x^{1-\alpha}\Big)/(\alpha(1-\alpha)). \tag{8.28}$$

   This converges to the Kullback Leibler divergence as $\lim_{\alpha\to 0} Q_\alpha(p|q) = Q_{KL}(q|p)$ and $\lim_{\alpha\to 1} Q_\alpha(p|q) = Q_{KL}(p|q)$.

3. χ^2-*divergence* for $f(x) = x^2 - 1$:

$$Q_{\chi^2}(p|q) = \sum_x \frac{(p_x - q_x)^2}{q_x} = \Big(\sum_x \frac{p_x^2}{q_x}\Big) - 1. \tag{8.29}$$

   Note that by Taylor expansion the χ^2-divergence locally approximates every f-divergence which is twice differentiable at $x = 1$. More precisely $C_f(p|q)/C_{\chi^2}(p|q) \to \frac{1}{2} f''(1)$ as $p \to q$.

4. *Hellinger divergence* for $f(x) = (\sqrt{x} - 1)^2/2$:

$$Q_{h^2}(p|q) = \Big(1 - \sum_x \sqrt{p_x q_x}\Big) = \frac{1}{2}\sum_x (\sqrt{p_x} - \sqrt{q_x})^2. \tag{8.30}$$

   The square root of Q_{h^2} is a metric since it is proportional to the norm difference in l_2.

5. l_1-*norm* is the only bare metric among the f-divergences and is obtained for $f(x) = |x - 1|$:

$$Q_{l_1}(p|q) = \sum_x |p_x - q_x|. \tag{8.31}$$

In the following we will discuss generalizations of these 'distance' measures to density operators.

8.4.1 χ^2 divergence

A simple generalization of the χ^2 divergence in Eq.(8.29) to the non-commutative context of density operators ρ and σ is

$$\begin{aligned} \chi^2_\alpha(\rho|\sigma) &:= \operatorname{tr}\left[(\rho - \sigma)\sigma^{-\alpha}(\rho - \sigma)\sigma^{\alpha-1}\right], \quad \alpha \in \mathbb{R} \\ &= \operatorname{tr}\left[\rho\sigma^{-\alpha}\rho\sigma^{\alpha-1}\right] - 1. \end{aligned} \tag{8.32}$$

[9] In fact, there are many variants of relative α-entropies in the literature—most of them are slight modifications of others. The *relative Renyi entropy* for instance is usually defined as $S_\alpha(p|q) = (\alpha - 1)^{-1} \log \sum_x p_x^\alpha q_x^{1-\alpha}$ with $\alpha \geq 0$. In the limit $\alpha \to 1$ this becomes $Q_{KL}(p|q)$ again.

8.4.2 Quasi-relative entropies

The relative entropy has generalizing relatives in very much the same way as the von Neumann entropy has Renyi and Tsallis type cousins. The most prominent ones are functions depending on the one-parameter family $\mathrm{tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]$ where $\alpha \in \mathbb{R}$. These quantities appear naturally in the context of hypothesis testing (see Sec.8.5). The Tsallis relative entropy $S_{(\alpha)}$ and the Renyi relative entropy S_α are defined as

$$S_{(\alpha)}(\rho|\sigma) \quad := \quad \frac{1}{1-\alpha}\big(1 - \mathrm{tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]\big), \tag{8.33}$$

$$S_\alpha(\rho|\sigma) \quad := \quad \frac{1}{\alpha-1} \log \mathrm{tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]. \tag{8.34}$$

Both are non-negative for $\alpha \geq 0$ and coincide with the relative entropy $S(\rho|\sigma) = \mathrm{tr}\left[\rho(\log\rho - \log\sigma)\right]$ in the limit $\alpha \to 1$.[10] One of the main properties of these generalized relative entropies follows from Thm.5.15 which implies that the map $(\rho,\sigma) \mapsto \mathrm{tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]$ is jointly concave for $\alpha \in [0,1]$ and jointly convex for $\alpha \in [-1,0]$. Exploiting that joint convexity implies monotonicity under completely positive trace-preserving maps (see Thm.5.16) we arrive at:

Theorem 8.10 (Monotonicity of quasi-relative entropies) *Let $\rho,\sigma \in \mathcal{M}_d(\mathbb{C})$ be two density matrices. Then for all completely positive trace-preserving linear maps $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ and $\alpha \in [0,1]$*

$$\mathrm{tr}\left[\rho^\alpha \sigma^{1-\alpha}\right] \leq \mathrm{tr}\left[T(\rho)^\alpha T(\sigma)^{1-\alpha}\right], \tag{8.35}$$

and the reverse inequality holds for $\alpha \in [-1,0]$.

This means that for $\alpha \in [0,1]$ both Renyi and Tsallis relative entropy are non-increasing under the action of a quantum channel.

A useful tool for deriving properties of the quasi-relative entropies in Eqs.(8.33,8.34) is a simple method which relates them to their classical counterparts: consider the spectral decompositions $\rho = \sum_i \lambda_i |x_i\rangle\langle x_i|$ and $\sigma = \sum_j \mu_j |y_j\rangle\langle y_j|$ and define $p_{ij} := \lambda_i |\langle x_i|y_j\rangle|^2$ and $q_{ij} := \mu_j |\langle x_i|y_j\rangle|^2$. Then $p,q \in \mathbb{R}^{2d}$ are classical probability distributions and straight forward calculation shows that

$$\mathrm{tr}\left[\rho^\alpha \sigma^{1-\alpha}\right] \quad = \quad \sum_x q_x \left(\frac{p_x}{q_x}\right)^\alpha, \tag{8.36}$$

$$S(\rho|\sigma) \quad = \quad \sum_x p_x \log \frac{p_x}{q_x}. \tag{8.37}$$

That is, under the mapping $(\rho,\sigma) \to (p,q)$ the quantum relative entropies coincide with their classical analogues. This allows us to easily prove some of their properties—or simply to import them from the world of classical probability distributions:

[10] Outside the interval $\alpha \in [0,1]$ inverses are take on the support and the relative entropies are set to ∞ if the kernel of the operator to be inverted is not contained in the kernel of the other operator.

Theorem 8.11 (Relation among relative entropies) *Let $\rho, \sigma \in \mathcal{M}_d(\mathbb{C})$ be a pair of positive semi-definite operators and $S_{(\alpha)}(\rho|\sigma) := \big(\mathrm{tr}\left[\rho - \rho^\alpha \sigma^{1-\alpha}\right]\big)/(1-\alpha)$. Then for all $\alpha \leq 1$ and $\beta \geq 1$ we have*

$$S_{(\alpha)}(\rho|\sigma) \leq S(\rho|\sigma) \leq S_{(\beta)}(\rho|\sigma). \tag{8.38}$$

Proof The result follows from the fact that for all positive numbers t and s it holds that $(1-t^{-s})/s \leq \log t \leq (t^s - 1)/s$. This gives

$$\frac{p_x}{s}\big(1 - p_x^{-s} q_x^s\big) \;\leq\; p_x \log \frac{p_x}{q_x} \;\leq\; \frac{p_x}{s}\big(p_x^s q_x^{-s} - 1\big). \tag{8.39}$$

Using the relations in Eqs.(8.36,8.37) the desired result follows by setting $s = 1-\alpha$ on the l.h.s. and $s = \beta - 1$ on the right. □

Note that other properties and uses of the functional $\mathrm{tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]$ will be discussed in Sec.8.5 in the context of hypothesis testing.

8.4.3 Fidelity

8.5 Hypothesis testing

Hypothesis testing is the simplest task in the context of statistical inference: given two possibilities, find out which one is actually realized by performing a statistical test. General examples are questions of the form "was there a Higgs particle generated?" or "can we identify a tumor in our tomographic data?". In our context the two hypotheses correspond to two density operators and we will see that several of the previously discussed distance measures for density operators will get an operational meaning as a means of quantifying how distinguishable the two states are.

Suppose we are given one out of two possible sources which produce states characterized by density operators ρ and σ respectively. Assume further that the a priori probability for ρ is p so that $(1-p)$ is the prior for σ. Imagine we want to decide whether the actual source produces ρ or σ by performing a single two-outcome measurement with effect operators $\{P, \mathbb{1} - P\}$. We then assign one of the measurement outcomes, say the one corresponding to P, to the hypothesis ρ while the other outcome leads us to guess σ. In this way, the average probability for a false conclusion will be

$$e(P) := p\,\mathrm{tr}\left[\rho(\mathbb{1} - P)\right] + (1-p)\mathrm{tr}\left[\sigma P\right]. \tag{8.40}$$

The following tells us how to minimize this average error probability and provides the optimal measurement:

Theorem 8.12 (Quantum Neyman-Pearson) *Within the above context the error probability satisfies the bound*

$$e(P) \geq \frac{1}{2}\Big(1 - \big|\big|p\rho - (1-p)\sigma\big|\big|_1\Big), \tag{8.41}$$

where equality is achieved iff P is a projection onto the positive part of $[p\rho - (1-p)\sigma]$.

Proof We have to minimize $e(P)$ w.r.t. all $0 \leq P \leq \mathbb{1}$, i.e.,

$$\min_P e(P) = p - \max_P \operatorname{tr}\left[\big(p\rho - (1-p)\sigma\big)P\right]. \tag{8.42}$$

Decomposing $(p\rho - (1-p)\sigma) =: A$ into orthogonal positive and negative parts as $A = A_+ - A_-$ it becomes clear that the maximum is attained iff $PA_+ = A_+$ and $PA_- = 0$. Eq.(8.41) then follows by using that for Hermitian operators $\operatorname{tr}[A_+] = (||A||_1 + \operatorname{tr}[A])/2$. □

This gives a simple operational meaning to the trace norm distance between density operators: it quantifies how good the states can be distinguished by a single measurement.

Let us now go one step further and assume that we can measure not only on a single state but on $n \in \mathbb{N}$ copies of the system. That is, we let the source produce n identical state (either ρ or σ) before we perform a global measurement. According to the above Thm.8.12 an optimal measurement leads to an error probability

$$e_n := \frac{1}{n}\Big(1 - \big|\big|p\rho^{\otimes n} - (1-p)\sigma^{\otimes n}\big|\big|_1\Big). \tag{8.43}$$

This turns out to decrease exponentially as $\exp(-\xi n)$ at a rate which is asymptotically independent of p (unless $p = 0, 1$) and given in the following theorem:

Theorem 8.13 (Quantum Chernoff bound) *For every non-trivial a priori probability $p \neq 0, 1$ an optimal sequence of measurements yields*

$$\xi := \lim_{n\to\infty}\Big(-\frac{1}{n}\log e_n\Big) = -\log\Big(\inf_{s\in[0,1]} \operatorname{tr}\left[\rho^{1-s}\sigma^s\right]\Big). \tag{8.44}$$

In some contexts it is natural to treat the two errors (1. we guess ρ while it was σ and 2. we guess σ while it was ρ) in an asymmetric way.[11] Assume we again perform a two-outcome measurement described by effect operators $P_n \in \mathcal{M}_{d^n}(\mathbb{C})$ and $(\mathbb{1} - P_n)$ on n copies of the system. Consider the task of minimizing the error probability

$$\beta_n(P_n) := \operatorname{tr}\left[\sigma^{\otimes n}P_n\right] \quad \text{under the constraint} \quad \operatorname{tr}\left[\rho^{\otimes n}(\mathbb{1} - P_n)\right] \leq \epsilon. \tag{8.45}$$

For a sequence of optimal measurements the error β_n will again decrease exponentially with n and the rate will turn out to be the relative entropy $S(\rho|\sigma)$. The analogous classical result is sometimes called *Stein's Lemma* where the Kullback-Leibler divergence appears as the optimal rate function. Before we come to the quantum version of this result we state a crucial ingredient for its proof which eventually enables us again to reduce the quantum to the classical case:

[11] In classical statistical analysis this is natural for instance in many medical contexts: there we want to be sure not to overlook a certain decease, i.e., the probability for a false negative test should be small while one may accept a larger fraction of false positive tests.

Theorem 8.14 (Hiai-Petz) *Let $\rho, \sigma \in \mathcal{M}_d(\mathbb{C})$ be two density operators and $n \in \mathbb{N}$. Define a map $T : \mathcal{M}_{d^n}(\mathbb{C}) \to \mathcal{M}_{d^n}(\mathbb{C})$ via $T(X) := \sum_{i=1}^k P_i X P_i$ from the spectral decomposition $\sigma^{\otimes n} = \sum_{i=1}^k \lambda_i P_i$ where the sum runs over k distinct eigenvalues and the P_i's project onto the corresponding eigenspaces. Then*

$$S\big(T(\rho^{\otimes n})\big|\sigma^{\otimes n}\big) \;\leq\; nS(\rho|\sigma) \;\leq\; S\big(T(\rho^{\otimes n})\big|\sigma^{\otimes n}\big) + d\log(n+1), \tag{8.46}$$

where $S(\rho|\sigma) = \mathrm{tr}\left[\rho(\log\rho - \log\sigma)\right]$ is the relative entropy.

Proof The left inequality follows from monotonicity of the relative entropy under trace-preserving completely positive maps together with additivity $S\big(\rho^{\otimes n}\big|\sigma^{\otimes n}\big) = nS(\rho|\sigma)$ and the invariance $T(\sigma^{\otimes n}) = \sigma^{\otimes n}$.

For the right inequality we use again additivity and proceed as follows:

$$\begin{aligned} n\, S(\rho|\sigma) \;\; &\leq \;\; \mathrm{tr}\left[\rho^{\otimes n}\big(\log T(\rho^{\otimes n}) - \log\sigma^{\otimes n}\big)\right] + \log k & (8.47)\\ &\leq \;\; S\Big(T(\rho^{\otimes n})\big|\sigma^{\otimes n}\Big) + d\log\,(n+1). & (8.48)\end{aligned}$$

Here the first inequality uses $\log\rho^{\otimes n} \leq \log T(\rho^{\otimes n}) + \log k$ which follows from Lemma 8.2 together with the operator monotonicity of the logarithm. The second inequality exploits that (i) T is a conditional expectation (see Eq.(1.43)), so that in particular $\mathrm{tr}\left[AT(B)\right] = \mathrm{tr}\left[T(A)T(B)\right]$ for any A, B and (ii) the simple combinatorial bound $k \leq (n+1)^d$. □

Note that Eq.(8.46) implies in particular that

$$S(\rho|\sigma) = \lim_{n\to\infty} \frac{1}{n} S\big(T(\rho^{\otimes n})\big|\sigma^{\otimes n}\big). \tag{8.49}$$

Moreover since $\sigma^{\otimes n}$ commutes with $T(\rho^{\otimes n})$ the relative entropy on the r.h.s. equals the one of the classical probability distributions which are obtained from measuring the two states in the basis in which they are simultaneously diagonal. Since the relative entropy (or Kullback-Leibler divergence) is the optimal rate function appearing in the classical Stein's lemma, the above Thm.8.14 implies that $S(\rho|\sigma)$ is an achievable rate in the quantum context as well. The following shows that it is indeed the optimal rate:

Theorem 8.15 (Quantum Stein Lemma) *Consider the task of distinguishing two quantum states $\rho, \sigma \in \mathcal{M}_d(\mathbb{C})$. Let β_n be the error probability as defined in Eq.(8.45), minimized over all measurements. For every $\epsilon \in (0,1)$ we have that*

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_n = S(\rho|\sigma). \tag{8.50}$$

Proof As discussed above achievability of the rate $S(\rho|\sigma)$ follows from Thm.8.14 together with the classical Stein's lemma. What remains to prove is thus an upper bound on the optimal rate, i.e., a lower bound on the error probability

β_n. To this end we apply Lemma 8.1 to $A = \rho^{\otimes n}$ and $B = e^{\lambda n}\sigma^{\otimes n}$ for some $\lambda \in \mathbb{R}$ to be chosen later. For $s \in [0,1]$ this gives

$$e^{-s\lambda n}\mathrm{tr}\left[\rho^{1+s}\sigma^{-s}\right]^n \geq \mathrm{tr}\left[\left(\rho^{\otimes n} - e^{\lambda n}\sigma^{\otimes n}\right)P_n\right] \quad (8.51)$$

$$\geq (1-\epsilon) - e^{n\lambda}\beta_n(P_n), \quad (8.52)$$

where the second inequality follows from Eq.(8.45). Rewriting these inequalities we obtain

$$\beta_n \geq e^{-n\lambda}\left[(1-\epsilon) - e^{-n\left(\lambda s - f(s)\right)}\right], \quad (8.53)$$

where we set $f(s) := \log \mathrm{tr}\left[\rho^{1+s}\sigma^{-s}\right]$. Since $f(0) = 0$ and $f'(0) = S(\rho|\sigma)$ the choice $\lambda = S(\rho|\sigma) + \delta$ will for any $\delta > 0$ guarantee that there is an $s \in (0,1]$ such that $\lambda s > f(s)$. Thus, $\lim_{n\to\infty} -\frac{1}{n}\beta_n \leq S(\rho|\sigma) + \delta$ and since this holds for arbitrary $\delta > 0$ the relative entropy is indeed the optimal asymptotic rate. □

Lemma 8.1 *Let $A, B \in \mathcal{M}_d(\mathbb{C})$ be two positive semi-definite operators and denote by $(A-B)_+$ the positive part of the Hermitian matrix $(A-B)$. Then for all $s \in [0,1]$:*

$$||A-B||_1 \geq \mathrm{tr}\left[A+B\right] - 2\mathrm{tr}\left[A^s B^{1-s}\right], \quad \textit{and} \quad (8.54)$$

$$\mathrm{tr}\left[(A-B)_+\right] \leq \mathrm{tr}\left[A^{1+s}B^{-s}\right]. \quad (8.55)$$

Proof In order to prove Eq.(8.55) we show first that it is sufficient to consider diagonal A and B. To this end we exploit that $\mathrm{tr}\left[A^{1+s}B^{-s}\right]$ is non-increasing under completely positive, trace-preserving maps (see e.g. Thm.8.10). We apply this to the map $T(\cdot) := \sum_i P_i \cdot P_i$ which is constructed from the spectral decomposition $(A-B) =: \sum_i \lambda_i P_i$ where the P_i's are one-dimensional projections. Then since $\mathrm{tr}\left[(A-B)_+\right] = \mathrm{tr}\left[\left(T(A) - T(B)\right)_+\right]$ and now $[T(A), T(B)] = 0$ it is indeed sufficient to consider diagonal A and B. For those the assertion follows from the simple inequality $a - b \leq a(a/b)^s$ which holds for numbers $a \geq b \geq 0$. □

Lemma 8.2 *Let $\rho \in \mathcal{M}_d(\mathbb{C})$ be a density operator and $\{P_i \in \mathcal{M}_d(\mathbb{C})\}_{i=1,\ldots,k}$ a POVM where each P_i is a projection. If $T(X) := \sum_{i=1}^k P_i X P_i$, then*

$$\rho \leq kT(\rho). \quad (8.56)$$

Proof Due to linearity it is sufficient to consider pure states $\rho = |\psi\rangle\langle\psi|$. Then for every $|\phi\rangle \in \mathbb{C}^d$

$$\langle\phi|kT(\rho) - \rho|\phi\rangle = \left(k\sum_{i=1}^k |\langle\phi|P_i|\psi\rangle|^2\right) - \left|\langle\phi|\sum_{i=1}^k P_i|\psi\rangle\right|^2 \geq 0, \quad (8.57)$$

where the inequality follows from Chauchy-Schwarz when applied to a vector with components $\langle\phi|P_i|\psi\rangle$ and the other with components k-times 1. □

8.6 Hilbert's projective metric

Let $\mathcal{V}$ be a real, finite-dimensional vector space (e.g., the space of Hermitian matrices in $\mathcal{M}_d(\mathbb{C})$). A *convex cone* $\mathcal{C} \subset \mathcal{V}$ is a subset for which $\alpha\mathcal{C} + \beta\mathcal{C} \subseteq \mathcal{C}$ for all $\alpha, \beta \geq 0$. We will call a convex cone in $\mathcal{V}$ a *proper cone* if it is closed, has a non-empty interior and satisfies $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$. Examples of proper cones are sets of non-negative polynomials, vectors or matrices with non-negative entries, positive semi-definite matrices, and the cones of un-normalized density matrices which are separable, have a positive partial transpose, are undistillable, etc.

Recall that the *dual cone* is defined as $C^* := \{c \in \mathcal{V} | \forall v \in \mathcal{C} : \langle v|c\rangle \geq 0\}$ and that by the *bipolar theorem* $\mathcal{C}^{**} = \mathcal{C}$ for closed convex cones. This implies that

$$v \in \mathcal{C} \quad \Leftrightarrow \quad \langle v|c\rangle \geq 0 \quad \text{for all } c \in \mathcal{C}^*. \tag{8.58}$$

Every proper cone induces a (partial) order relation in $\mathcal{V}$ and we will write $a \geq_{\mathcal{C}} b$ meaning $a - b \in \mathcal{C}$. We will omit the subscript $_{\mathcal{C}}$ if the cone $\mathcal{C}$ under consideration is the set of positive semi-definite matrices. Also note that in this case we have $\mathcal{C} = \mathcal{C}^*$.

For every pair of non-zero elements $a, b \in \mathcal{C}$ define

$$\sup(a/b) := \sup_{c \in \mathcal{C}^*} \frac{\langle a|c\rangle}{\langle b|c\rangle}, \qquad \inf(a/b) := \inf_{c \in \mathcal{C}^*} \frac{\langle a|c\rangle}{\langle b|c\rangle}. \tag{8.59}$$

By construction $\sup(a/b) \geq \inf(a/b) \geq 0$ and $\sup(a/b) = 1/\inf(b/a)$. If $\mathcal{C}$ (and consequently also $\mathcal{C}^*$) is a proper cone we can use Eq.(8.58) and rewrite

$$\sup(a/b) \;=\; \inf\{\lambda \in \mathbb{R} | a \leq_{\mathcal{C}} \lambda b\}, \tag{8.60}$$

$$\inf(a/b) \;=\; \sup\{\lambda \in \mathbb{R} | \lambda b \leq_{\mathcal{C}} a\}. \tag{8.61}$$

This implies that $\inf(a/b)b \leq_{\mathcal{C}} a \leq_{\mathcal{C}} \sup(a/b)b$, where the latter makes sense only if $\sup(a/b)$ is finite. In other words, Eq.(8.59) provides the factors by which b has to be rescaled at least in order to become larger or smaller than a.

Hilbert's projective metric is then defined for non-zero $a, b \in \mathcal{C}$ as

$$D(a,b) := \ln\big[\sup(a/b)\sup(b/a)\big]. \tag{8.62}$$

Obviously, D is symmetric, non-negative and satisfies $D(a, \beta b) = D(a, b)$ for all $\beta > 0$. That is, D depends only on the 'direction' of its arguments. Since it turns out to fulfill the triangle inequality and $D(a, b) = 0$ implies that $a = \beta b$ for some $\beta > 0$, D is a *projective metric*. Hence, if we restrict the arguments a, b further to a subset which excludes multiples of elements (such as the unit sphere of a norm), then D becomes a metric on that space. With $\mathcal{C}$ the cone of positive semi-definite matrices D, for instance, is a metric on the set of density matrices. The following relates Hilbert's projective metric to metrics induced by norms.

Proposition 8.2 (Norm bound) *Consider a proper cone $\mathcal{C}$ in a real linear space. Let $A, B \in \mathcal{C}$ be non-zero elements and define $\nu := \inf(B/A)$ and $\mu := \inf(A/B)$. Then*

$$||A-B|| \;\leq\; |1-\nu|\;||A|| + |1-\mu|\;||B|| + \min\big[|1-\nu|\;||A||, |1-\mu|\;||B||\big], \tag{8.63}$$

for every norm for which $X \geq_{\mathcal{C}} Y \geq_{\mathcal{C}} 0$ implies $||X|| \geq ||Y||$.

Proof By definition $B \geq_{\mathcal{C}} \nu A$ which implies $A - B \leq_{\mathcal{C}} (1-\nu)A$ and similar holds for μ if only A and B are interchanged. This guarantees the existence of positive elements $P_\nu, P_\mu \geq_{\mathcal{C}} 0$ such that

$$A - B + P_\nu = (1-\nu)A \quad \text{and} \quad B - A + P_\mu = (1-\mu)B. \tag{8.64}$$

Using the triangular inequality for the norm $||A-B|| \leq ||A-B+P_\nu|| + ||P_\nu||$ and exploiting its assumed monotonicity by bounding $||P_\nu|| \leq ||P_\nu + P_\mu||$, Eq.(8.64) then leads to

$$||A-B|| \;\leq\; ||A-B+P_\nu|| + ||P_\nu + P_\mu|| \tag{8.65}$$
$$\leq\; 2|1-\nu|\;||A|| + |1-\mu|\;||B||. \tag{8.66}$$

Since the same reasoning and thus the same inequality applies with $\mu \leftrightarrow \nu$ and A and B interchanged, Eq.(8.63) then follows from taking the smaller bound. □

Corollary 8.2 (Norms vs. Hilbert's projective metric) *With the assumptions of Prop.8.2 and the additional requirement that $\mu, \nu \leq 1$ we have*

$$||A-B|| \leq 3\left(\frac{||A|| + ||B||}{2} - \sqrt{||A||\;||B||}\; e^{-D(A,B)/2}\right). \tag{8.67}$$

Proof Instead of taking the minimum in Eq.(8.63) we take the average and then use that for positive numbers $(a+b)/2 \geq \sqrt{ab}$. In this way we obtain

$$||A-B|| \;\leq\; \frac{3}{2}\big(||A|| + ||B|| - (\nu||A|| + \mu||B||)\big) \tag{8.68}$$
$$\leq\; 3\left(\frac{||A|| + ||B||}{2} - \sqrt{\nu||A||\mu||B||}\right), \tag{8.69}$$

which leads the sought result by using $\inf(A/B) = 1/\sup(B/A)$ together with the definition Eq.(8.62). □

Better bounds can be obtained by exploiting some additional structure. So let us now explicitly assume that $\mathcal{C}$ is the cone of positive semi-definite operators on a finite-dimensional Hilbert space and consider the trace norm:

Proposition 8.3 (Trace norm vs. Hilbert's projective metric) *Let $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ be two density matrices. Then*

$$\frac{1}{2}||\rho_1 - \rho_2||_1 \;\leq\; \big[1 + \inf(\rho_2/\rho_1)\big]^{-1} - \big[1 + \sup(\rho_2/\rho_1)\big]^{-1} \tag{8.70}$$
$$\leq\; \tanh\big[D(\rho_1,\rho_2)/4\big]. \tag{8.71}$$

Proof Recall that $||\rho_1-\rho_2||_1/2 = \mathrm{tr}\,[P(\rho_1-\rho_2)]$ for some Hermitian projection P. Exploiting that $x(2-x) \leq 1$ for all $x \in [0,2]$ and applying it to $x = \mathrm{tr}\,[P(\rho_1+\rho_2)]$ we obtain

$$\frac{1}{2}||\rho_1-\rho_2||_1 \leq \frac{\mathrm{tr}\,[P(\rho_1-\rho_2)]}{\mathrm{tr}\,[P(\rho_1+\rho_2)]\,\big(2-\mathrm{tr}\,[P(\rho_1+\rho_2)]\,\big)} \tag{8.72}$$

$$= \big(1+\mathrm{tr}\,[P\rho_2]\,/\mathrm{tr}\,[P\rho_1]\,\big)^{-1} - \big(1+\mathrm{tr}\,[P'\rho_2]\,/\mathrm{tr}\,[P'\rho_1]\,\big)^{-1} \tag{8.73}$$

where $P' = \mathbb{1} - P$ and the last equation follows from elementary algebra. Eq.(8.70) now follows from Eq.(8.73) by taking the supremum over all $P, P' \geq 0$ and inserting the definitions from Eq.(8.59).

In order to arrive at Eq.(8.71) we use that $\inf(\rho_2/\rho_1) = 1/\sup(\rho_1/\rho_2)$ and abbreviate $x := \sup(\rho_1/\rho_2)$, $y := \sup(\rho_2/\rho_1)$. In this way we can rewrite and bound the r.h.s. of Eq.(8.70) as

$$\frac{xy-1}{(x+1)(y+1)} \leq \frac{\sqrt{xy}-1}{\sqrt{xy}+1}, \tag{8.74}$$

where the inequality follows from applying the inequality $(x+1)(y+1) \geq (\sqrt{xy}+1)^2$, which holds for all $x, y \geq 0$. Using that according to definition (8.62) we have $xy = \exp\,[D(\rho_1,\rho_2)]$ then completes the proof. □

Exploiting that $2\mathrm{tanh}(x/2) \leq x$ for $x \geq 0$ we obtain in particular

$$||\rho_1-\rho_2||_1 \leq D(\rho_1,\rho_2)/2. \tag{8.75}$$

Note that a non-trivial lower bound on norms cannot exist: suppose that for some vector ψ we have $\langle\psi|\rho_1|\psi\rangle > 0$ but $\langle\psi|\rho_2|\psi\rangle = 0$. Then $\sup(\rho_1/\rho_2) = \infty$ according to Eq.(8.60) so that $D(\rho_1,\rho_2) = \infty$ although the two states could be arbitrarily close in norm. However, if the supports of ρ_1 and ρ_2 are the same, Hilbert's metric is finite and can be explicitly expressed in terms of operator norms:

Proposition 8.4 *Let $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ be density matrices and consider the positive semi-definite cone $\mathcal{C}$. Then with $^{-1}$ denoting the pseudo inverse (inverse on the support) we have*

$$\sup(\rho_1/\rho_2) = \begin{cases} ||\rho_2^{-1/2}\rho_1\rho_2^{-1/2}||_\infty, & \textit{if } \mathrm{supp}[\rho_1] \subseteq \mathrm{supp}[\rho_2] \\ \infty & \textit{otherwise} \end{cases} \tag{8.76}$$

$$\inf(\rho_1/\rho_2) = \begin{cases} ||\rho_1^{-1/2}\rho_2\rho_1^{-1/2}||_\infty^{-1}, & \textit{if } \mathrm{supp}[\rho_2] \subseteq \mathrm{supp}[\rho_1] \\ 0 & \textit{otherwise} \end{cases} \tag{8.77}$$

$$D(\rho_1,\rho_2) = \begin{cases} \ln\Big[||\rho_1^{-1/2}\rho_2\rho_1^{-1/2}||_\infty||\rho_2^{-1/2}\rho_1\rho_2^{-1/2}||_\infty, & \textit{if } \mathrm{supp}[\rho_2] = \mathrm{supp}[\rho_1] \\ \infty & \textit{otherwise} \end{cases}$$

Proof We only have to prove the relation for $\sup(\rho_1/\rho_2)$ since this implies the other two by $\inf(\rho_1/\rho_2) = 1/\sup(\rho_2/\rho_1)$ and definition (8.62) respectively.

Assume that $\text{supp}[\rho_1] \not\subseteq \text{supp}[\rho_2]$. Then there is a vector ψ for which $\langle\psi|\rho_2|\psi\rangle = 0$ while $\langle\psi|\rho_1|\psi\rangle > 0$ so that the infimum in Eq.(8.60) is over an empty set and thus by the usual convention ∞. If, however, $\text{supp}[\rho_1] \subseteq \text{supp}[\rho_2]$, then $\rho_1 \leq \lambda\rho_2$ is equivalent to $\rho_2^{-1/2}\rho_1\rho_2^{-1/2} \leq \lambda\mathbb{1}$ and the smallest λ for which this holds is the operator norm. □

An immediate consequence of this is a simple bound on the quantum χ^2-divergence:

Corollary 8.3 *Let $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ be density matrices. Then*

$$\chi^2_{1/2}(\rho_1|\rho_2) \leq \sup(\rho_1/\rho_2) - 1. \tag{8.78}$$

Proof Recall the definition of the quantum χ^2-divergence from Eq.(8.32): $\chi^2(\rho_1|\rho_2) = \text{tr}\left[\rho_1\rho_2^{-1/2}\rho_1\rho_2^{-1/2}\right] - 1$. From here the assertion follows via Eq.(8.76) by using that $\text{tr}\left[\rho_1\rho_2^{-1/2}\rho_1\rho_2^{-1/2}\right] \leq ||\rho_2^{-1/2}\rho_1\rho_2^{-1/2}||_\infty$. □

8.7 Contractivity and the increase of entropy

Trace norm

Theorem 8.16 (Trace norm contractivity) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ be a trace preserving positive linear map. Then for all Hermitian $H \in \mathcal{M}_d(\mathbb{C})$:*

$$||T(H)||_1 \leq ||H||_1. \tag{8.79}$$

Proof Let $T(H) = Q_+ - Q_-$ and $H = P_+ - P_-$ be decompositions into orthogonal parts $Q_\pm \geq 0$ and $P_\pm \geq 0$. Projecting the equation $Q_+ - Q_- = T(P_+) - T(P_-)$ onto the support space of Q_+ (or Q_-) and exploiting positivity of the involved expressions leads to $\text{tr}[Q_+] \leq \text{tr}[T(P_+)]$ (or $\text{tr}[Q_-] \leq \text{tr}[T(P_-)]$). Using the trace-preserving property of T then leads to $\text{tr}[Q_+ + Q_-] \leq \text{tr}[P_+ + P_-]$ which is a reformulation of Eq.(8.79). □

Note that this means in particular that for all pairs of density operators ρ_1 and ρ_2 we have

$$||T(\rho_1) - T(\rho_2)||_1 \leq ||\rho_1 - \rho_2||_1. \tag{8.80}$$

For a generic map T this inequality can in principle be sharpened in the sense that the r.h.s. is multiplied by a factor smaller than one. The best possible such factor which is independent of the ρ_i's is called *contraction coefficient*. For the trace norm its computation is simplified due to the following:

Lemma 8.3 (Trace norm contraction coefficient) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ be a linear map. Then*

$$\sup_{\rho_1,\rho_2} \frac{||T(\rho_1) - T(\rho_2)||_1}{||\rho_1 - \rho_2||_1} = \frac{1}{2}\sup_{\psi\perp\phi} \left|\left|T\big(|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\big)\right|\right|_1, \tag{8.81}$$

where the supremum on the l.h.s. is taken over all density matrices in $\mathcal{M}_d(\mathbb{C})$ and the one on the r.h.s. over all pairs of orthogonal state vectors in $\mathbb{C}^d$.

Proof The r.h.s. in Eq.(8.81) is certainly a lower bound since the set over which the supremum is taken is reduced. So we have to show that it is also an upper bound. To this end, consider any difference of density operators with $\rho_1 - \rho_2 = P_+ - P_-$ being a decomposition into orthogonal positive and negative parts (i.e., $P_\pm \geq 0$ and $P_+P_- = 0$). By rescaling both numerator and denominator of $||T(P_+ - P_-)||_1/||P_+ - P_-||_1$ we can achieve that $\mathrm{tr}\,[P_\pm] = 1$ so that on the l.h.s. of Eq.(8.81) we can w.l.o.g. assume that the two density matrices have orthogonal supports. In this case $||P_+ - P_-||_1 = 2$. A convex decomposition then leads to $P_+ - P_- = \sum_i \lambda_i \big(|\psi_i\rangle\langle\psi_i| - |\phi_i\rangle\langle\phi_i|\big)$ with $\psi_i \perp \phi_i$ for all i. Exploiting convexity of the norm then finally leads to the equality in Eq.(8.81). □

We will see below that non-trivial upper bounds for the trace-norm contraction coefficient, and thus improvements on inequality (8.80), can for instance be obtained from the *projective diameter* and the fidelity contraction coefficient. Another possibility to improve on Eq.(8.80) is to formalize the intuition that if T is strictly inside the cone of positive maps, then it should be a strict contraction. In the case of classical channels (i.e., stochastic matrices) this is the content of *Doeblin's theorem* of which a quantum counterpart looks as follows:

Theorem 8.17 (Quantum version of Doeblin's theorem) *Let $T, T' : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_{d'}(\mathbb{C})$ be two trace-preserving and Hermiticity-preserving linear maps of which T' is such that $T'(X) = \mathrm{tr}\,[X]\,Y$ for some $Y \in \mathcal{M}_{d'}(\mathbb{C})$. If $T - \epsilon T'$ is positive for some $\epsilon \geq 0$, then for all density matrices $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$:*

$$||T(\rho_1) - T(\rho_2)||_1 \leq (1-\epsilon)||\rho_1 - \rho_2||_1. \tag{8.82}$$

Proof We will use a variant of Eq.(8.10), namely that for any traceless Hermitian matrix X one can express the trace-norm as

$$||X||_1 = \sup_{0 \leq P \leq \mathbb{1}} 2\mathrm{tr}\,[PX]\,. \tag{8.83}$$

Applying this to $X = T(\rho_1 - \rho_2)$ and using that $T'(\rho_1 - \rho_2) = 0$ we obtain

$$||T(\rho_1) - T(\rho_2)||_1 = \sup_{0 \leq P \leq \mathbb{1}} 2\mathrm{tr}\,[(\rho_1 - \rho_2)(T - \epsilon T')^*(P)]\,. \tag{8.84}$$

Note that $P' := (T - \epsilon T')^*(P)$ is positive if, as assumed, $T - \epsilon T'$ is a positive map. Moreover, the fact that $P \leq \mathbb{1}$ together with the trace-preserving property of T and T' implies that $P' \leq (T - \epsilon T')^*(\mathbb{1}) = (1-\epsilon)\mathbb{1}$ so that finally

$$||T(\rho_1) - T(\rho_2)||_1 \leq (1-\epsilon) \sup_{0 \leq P \leq \mathbb{1}} 2\mathrm{tr}\,[(\rho_1 - \rho_2)P]\,, \tag{8.85}$$

from which the claimed inequality follows by exploiting Eq.(8.83) once again. □

While positivity is difficult to decide, complete positivity is easily seen from the Choi-Jamiolkowski operator (see Prop.2.1). So if $\tau = (T \otimes \mathrm{id}_d)(|\Omega\rangle\langle\Omega|)$ is the Choi-Jamiolkowski operator of T, then Eq.(8.82) holds for every $\epsilon \geq 0$ for which there is a Hermitian Y with $\mathrm{tr}\,[Y] = 1$ so that

$$\tau \geq \frac{\epsilon}{d}\big(Y \otimes \mathbb{1}\big). \tag{8.86}$$

If T is highly contractive in the sense that it maps everything close to a highly mixed density operator ρ_∞, then $Y = \rho_\infty$ can be a reasonable choice. Also note that for a given Y the largest ϵ satisfying Eq.(8.86) can be computed explicitly following the approach of Prop.8.4. Choosing $Y = \mathbb{1}/d$ enables us to relate the spectra of a channel and it Jamiolkowski state:[12]

Corollary 8.4 (Channel spectrum vs. Jamiolkowski spectrum) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a completely positive and trace-preserving linear map with corresponding Jamiolkowski state τ. If we denote by μ the largest modulus of eigenvalues of T which are strictly inside the unit disc and write $\lambda_{min}(\tau)$ for the smallest eigenvalue of τ, then*

$$\mu \leq 1 - d^2 \lambda_{min}(\tau). \tag{8.87}$$

PROOF Applying Thm.8.17 with $Y = \mathbb{1}/d$ to T^n for $n \in \mathbb{N}$ leads to $||T^n(\rho_1 - \rho_2)||_1 \leq (1 - d^2\lambda_{min}(\tau))^n ||\rho_1 - \rho_2||_1$. As we will see in Prop.8.6 and Thm.8.23 below, $\sup_{\rho_1,\rho_2} ||T^n(\rho_1 - \rho_2)||_1$ is, up to a constant, asymptotically lower bounded by μ^n so that Eq.(8.87) follows when considering $n \to \infty$. □

Hilbert's projective metric While investigating Perron-Frobenius theory of positive maps [**true?**], Birkhoff [**?**] observed that all strictly positive maps are contractive w.r.t. Hilbert's projective metric. This allowed him in essence to view convergence under subsequent application of a positive map as a special case of the Banach contraction mapping theorem. We will use the notation and terminology of Sec.8.6 and assume throughout that all cones are embedded in finite-dimensional real vector spaces. In order to state Birkhoff's observation we need to define the *projective diameter* $\Delta(T)$ of a positive, i.e., cone-preserving, map $T : \mathcal{C} \to \mathcal{C}$ on a proper cone $\mathcal{C}$ (e.g., the cone of positive semi-definite matrices in $\mathcal{M}_d(\mathbb{C})$):

$$\Delta(T) := \sup_{a,b \in \mathcal{C}^\circ} D\big(T(a), T(b)\big), \tag{8.88}$$

where the supremum runs over all elements a, b in the interior of $\mathcal{C}$, denoted by $\mathcal{C}^\circ$.

Theorem 8.18 (Birkhoff's contraction theorem) *Let $\mathcal{C}$ be a proper cone and $T : \mathcal{C} \to \mathcal{C}$ a positive (i.e., cone-preserving) linear map. Then*

$$\sup_{a,b \in \mathcal{C}^\circ} \frac{D\big(T(a), T(b)\big)}{D(a,b)} \leq \tanh\big(\Delta(T)/4\big). \tag{8.89}$$

In fact, Hilbert's metric is the essentially unique *projective metric* which allows to identify positive maps with contractive maps [**?**]. In order to formalize this, recall that a symmetric functional $D' : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_+ \cup \infty$ is called *projective metric* if it satisfies $D'(a,b) \leq D'(a,c) + D'(c,b)$ for all $a,b,c \in \mathcal{C}$ and $D'(a,b) = 0$ iff $a = \lambda b$ for some $\lambda > 0$.

[12] For another relation of that kind see Prop.6.5.

Theorem 8.19 (Uniqueness of Hilbert's projective metric) *Let $\mathcal{C}$ be a proper cone and D' a projective metric such that every strictly positive linear map $T : \mathcal{C} \to \mathcal{C}^\circ$ is a strict contraction w.r.t. to D'. Then there exists a continuous and strictly increasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $D' = f(D)$ where D is Hilbert's projective metric. Moreover, for any linear $T : \mathcal{C} \to \mathcal{C}^\circ$ we have*

$$\sup_{a,b\in\mathcal{C}^\circ} \frac{D\big(T(a),T(b)\big)}{D(a,b)} \leq \sup_{a,b\in\mathcal{C}^\circ} \frac{D'\big(T(a),T(b)\big)}{D'(a,b)}. \tag{8.90}$$

The projective diameter which appears in Birkhoff's theorem can in turn be lower bounded by the "second largest eigenvalue" [**?**]:

Theorem 8.20 (Spectral bound on projective diameter) *Let $\mathcal{C}$ be a proper cone and $T : \mathcal{C} \to \mathcal{C}$ a positive linear map for which $T(c) = c$ for some non-zero $c \in \mathcal{C}$. If $T(a) = \lambda a$ for some $\lambda \in \mathbb{C}$ and a not proportional to c, then*

$$|\lambda| \leq \tanh\big(\Delta(T)/4\big). \tag{8.91}$$

*[**I dropped the $\Delta < \infty$ and the $c \in \mathbb{C}^\circ$ requirement. Both seem superfluous .. proof? idea: perturb T in order to make it primitive and use that the spectrum changes continuously and that $\Delta(T) < \infty$ iff T is primitive (which should mean something like $T : \mathcal{C}\backslash 0 \to \mathcal{C}^\circ$)**]*

For the particular case of $\mathcal{C}$ being the cone of positive semi-definite matrices in $\mathcal{M}_d(\mathbb{C})$ we can show that the projective diameter also bounds the contraction w.r.t. the trace norm:

Theorem 8.21 (Trace norm contraction incl. projective diameter) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ be a positive, trace-preserving linear map. Then for all pairs of density matrices $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$*

$$\big|\big|T(\rho_1) - T(\rho_2)\big|\big|_1 \leq ||\rho_1 - \rho_2||_1 \tanh\big(\Delta(T)/4\big). \tag{8.92}$$

Proof Replacing $\rho_i \to T(\rho_i)$ in Prop.8.3 and taking the supremum over all full-rank density operators ρ_1, ρ_2 we obtain $\sup ||T(\rho_1 - \rho_2)||/2 \leq \tanh\big(\Delta(T)/4\big)$. On the l.h.s. of this inequality we may as well take the supremum over all density matrices and then obtain a lower bound by restricting to orthogonal pure states. Following Lem.8.3 this lower bound equals the trace-norm contraction coefficient which then implies Eq.(8.92). □

Birkhoff's theorem implies that if for two pairs of positive operators ρ_1, ρ_2 and ρ_1', ρ_2' there is a positive map $T : \rho_i \mapsto \rho_i'$, then $D(\rho_1, \rho_2) \geq D(\rho_1', \rho_2')$. In the following we will see that the converse is true as well: if D decreases, then there exists a completely positive map which does the mapping up to normalization. The main ingredient will be the following relation with the *condition number* which is defined for symmetric matrices as $\kappa(X) := ||X||_\infty ||X^{-1}||_\infty$.

Lemma 8.4 *Let $P \in \mathcal{M}_d(\mathbb{C})$ and $P' \in \mathcal{M}_{d'}(\mathbb{C})$ be positive definite operators. Then there exists a completely positive and unital linear map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ such that $T(P) = \lambda P'$ for some $\lambda > 0$ iff*

$$\kappa(P) \geq \kappa(P'). \tag{8.93}$$

Proof Let us first prove the "if part" and let all norms be operator norms. Note that $\kappa(P) = 1$ is equivalent to $P \propto \mathbb{1}$ for which the statement is trivially true (since $P' \propto \mathbb{1}$ as well). So let us assume that P is not proportional to the identity. Define two functions on spaces of symmetric matrices as $E_1(X) := \mathbb{1} - X/||P||$ and $E_2(X) := X - \mathbb{1}/||P^{-1}||$. By construction $E_i(P) \geq 0$ and $\text{span}\{E_1(P), E_2(P)\} = \text{span}\{P, \mathbb{1}\}$. Moreover, Eq.(8.93) implies that there is a $\lambda > 0$ such that

$$\frac{||P||}{||P'||} \geq \lambda \geq \frac{||P'^{-1}||}{||P^{-1}||}, \tag{8.94}$$

from which $E_i(\lambda P') \geq 0$ follows. Let ψ_1 (ψ_2) be an eigenvector corresponding to the smallest (largest) eigenvalue of P and define a map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ via

$$T(X) = \sum_i \langle\psi_i|X|\psi_i\rangle \frac{E_i(\lambda P')}{\langle\psi_i|E_i(P)|\psi_i\rangle}. \tag{8.95}$$

Since $E_i(P)|\psi_j\rangle = 0$ for $i \neq j$ we have that $T : E_i(P) \mapsto E_i(\lambda P')$ and therefore $T(\mathbb{1}) = \mathbb{1}$ and $T(P) = \lambda P'$. Moreover, due to positivity of the involved expressions T is a completely positive map, as desired.

For the "only if part" assume that there is a positive unital map T which acts as $T(P) = \lambda P'$. Then, since $E_i(P) \geq 0$, we have to have that $T\big(E_i(P)\big) = E_i(\lambda P') \geq 0$. These two inequalities (for $i = 1, 2$) are, however, equivalent to Eq.(8.94) from which Eq.(8.93) follows. □

As a consequence we obtain the claimed converse to Birkhoff's theorem:

Theorem 8.22 *Let $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ and $\rho_1', \rho_2' \in \mathcal{M}_{d'}(\mathbb{C})$ be two pairs of density matrices so that $\text{supp}(\rho_1) = \text{supp}(\rho_2)$. There exists a completely positive map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ which acts as $T(\rho_i) = \lambda_i \rho_i'$ for some $\lambda_i > 0$ iff*

$$D(\rho_1, \rho_2) \geq D(\rho_1', \rho_2'). \tag{8.96}$$

Proof The 'only if' part follows from Birkhoff's theorem. For the converse assume that Eq.(8.96) holds. The assumed equal support of ρ_1 and ρ_2 together with Eq.(8.96) implies that ρ_1' and ρ_2' have equal support as well. We can therefore restrict the input and output spaces to the respective support spaces and assume w.l.o.g. that all involved states have full rank. For any map T we can then define a map $\tilde{T}$ as

$$\tilde{T}(X) := \rho_1'^{-1/2} T\Big(\rho_1^{1/2} X \rho_1^{1/2}\Big) \rho_1'^{-1/2}/\lambda_1.$$

Note that T is completely positive iff $\tilde{T}$ is. Moreover, if we define $P := \rho_1^{-1/2}\rho_2\rho_1^{-1/2}$, $P' := \rho_1'^{-1/2}\rho_2'\rho_1'^{-1/2}$ and $\lambda := \lambda_2/\lambda_1$ then $T(\rho_i) = \lambda_i \rho_i'$ becomes equivalent to $\tilde{T}(P) = \lambda P'$ and $\tilde{T}(\mathbb{1}) = \mathbb{1}$. Hence, we can apply Lem.8.4 and by using Prop.8.4 we see that Eq.(8.96) is a reformulation of Eq.(8.93). □

To conclude this discussion we will give an operational interpretation of this result: assume that for a finite set of pairs of density matrices $\{(\rho_i, \rho_i')\}$ there is a completely positive map $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_{d'}(\mathbb{C})$ such that $T(\rho_i) =$

$\lambda_i\rho_i'$ for some $\lambda_i > 0$. Then we can construct a new map $\tilde{T} : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_{d'}(\mathbb{C}) \otimes \mathbb{Z}_2$ which is completely positive and trace-preserving and such that (i) $\rho_i \mapsto \rho_i'$ conditioned on outcome '1' on the ancillary two-level system, and (ii) this outcome is obtained with non-zero probability. More explicitly, this is obtained by

$$\tilde{T}(\rho) := cT(\rho) \otimes |1\rangle\langle 1| + B\rho B^\dagger \otimes |0\rangle\langle 0|, \tag{8.97}$$

with $c := ||T^*(\mathbb{1})||_\infty^{-1}$ and $B := \sqrt{\mathbb{1} - cT^*(\mathbb{1})}$.

In other words, Thm.8.22 shows that Hilbert's metric provides a necessary and sufficient condition for the existence of a probabilistic quantum operation which maps $\rho_i \mapsto \rho_i'$ upon success.

Asymptotic convergence and ergodic theory. In this paragraph we consider quantum channels which can be applied repeatedly, i.e., $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_d(\mathbb{C})$, and analyze the behavior of T^n as $n \in \mathbb{N}$ goes to infinity. Naturally, some of the spectral properties investigated in Chapter 6 will enter the discussion. For instance, depending on whether or not T has eigenvalues of modulus one which are different from one, T^n will or will not converge. In any case the image of T^n will converge to $\mathcal{X}_T$, the space spanned by all eigenvectors corresponding to eigenvalues of modulus one as defined in Eq.(6.65). Similarly, according to Prop.6.3 there will be an ascending subsequence $n_i \in \mathbb{N}$ such that $T^{n_i} \rightarrow T_\phi$ will converge to T_ϕ which is a projection onto χ_T. In other words

$$||T^n - T_\varphi^n|| \rightarrow 0 \tag{8.98}$$

where $T_\varphi := T_\phi T$ describes the "asymptotic dynamics" which is studied in greater detail in Thm.6.16. Recall from Chapter 6 that T_φ is defined in terms of the spectral decomposition of T where all eigenvalues of modulus smaller than one are set to zero while the peripheral eigensystem is kept unchanged. Similarly, T_ϕ is nothing but the projection corresponding to the peripheral eigenvectors.

The question which is addressed in the following is: how fast is the convergence in Eq.(8.98) ?

Clearly, some of the previous results on contractivity may be applied. For instance Thm.8.17 implies that

$$||T^n(\rho_1) - T^n(\rho_2)||_1 \leq (1-\epsilon)^n ||\rho_1 - \rho_2||_1,$$

but it leaves us with having to determine $(1-\epsilon)$.

A clearer picture about the asymptotic behavior[13] can be obtained by having a closer look at spectral properties: recall that the spectral radius of a positive and trace-preserving map is one, so that $(T - T_\varphi)$ has spectral radius equal to μ where $\mu := \sup_{\lambda \in \text{spec}(T)}\{|\lambda| < 1\}$ is the largest modulus of the eigenvalues of T which are in the interior of the unit disc. Note further that

$$T^n - T_\varphi^n = (T - T_\varphi)^n, \tag{8.99}$$

[13]which might, however, lead to quite ridiculous bounds for finite n

since $T_\varphi T = TT_\varphi = T_\varphi^2$. Taken together these facts suggest that μ governs the asymptotic behavior and we will see indeed that

$$||T^n - T_\varphi^n|| = \mathcal{O}(\nu^n), \quad \text{for all } \nu > \mu. \tag{8.100}$$

Before we proceed, some remarks are in order: (i) note that the choice of the norm is only relevant when it comes to decorating Eq.(8.100) with actual constants; most of the time we will consider $||\hat{T}^n - \hat{T}_\varphi^n||_\infty$, (ii) if T has only one eigenvalue of modulus one with corresponding fixed point density matrix ρ_∞, then $T_\varphi(\cdot) = \mathrm{tr}\,[\cdot]\,\rho_\infty$ and $T_\varphi = T_\phi = T^\infty$, (iii) it is not a typo that we do not have $\nu \geq \mu$ in Eq.(8.100). In order to understand the latter, consider the following example:

Example 8.1 (Slow convergence of non-diagonalizable channels) *A simple example of a channel whose convergence is slower than $\mathcal{O}(\mu^n)$ is a classical channel with $0 < \mu < 1$, characterized by a stochastic matrix $S \in \mathcal{M}_3(\mathbb{R}_+)$ of the form*

$$S = \begin{pmatrix} \mu & 0 & 0 \\ 1-\mu & \mu & 0 \\ 0 & 1-\mu & 1 \end{pmatrix}, \quad \textit{leading to } S^\infty = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}. \tag{8.101}$$

In this case we can compute $S^n - S^\infty$ explicitly and verify that $||S^n - S^\infty||$ is asymptotically bounded by $\mu^n n$ both, from below and from above. A general proof of such an asymptotic behavior is given Thm.8.23 where it is traced back to the appearance of non-trivial Jordan blocks.

Note that this example can be translated to the context of quantum channels by defining a quantum channel on $\mathcal{M}_3(\mathbb{C})$ as $T(\cdot) := \sum_{i,j} S_{i,j} \langle j| \cdot |j\rangle\ |i\rangle\langle i|$ which then displays the same type of asymptotic behavior since $\hat{T} = S \oplus 0$. The main purpose of this example is to show that quantum channels with large Jordan blocks exist and can be easily constructed: extending the example to $\mathcal{M}_d(\mathbb{C})$ leads to a Jordan block of size $d-1$.

The above example suggests that the asymptotic behavior is governed by μ *and* the size (denote it by d_μ) of the largest Jordan block corresponding to μ. In the following we will see that there are indeed n-independent constants C_1, C_2 such that

$$C_1 \mu^n n^{d_\mu - 1} \leq ||T^n - T_\varphi^n|| \leq C_2 \mu^n n^{d_\mu - 1}. \tag{8.102}$$

Proving this requires a Lemma:

Lemma 8.5 *Consider an upper-triangular matrix $X \in \mathcal{M}_D(\mathbb{C})$ with decomposition $X = \Lambda + N$ into a diagonal part Λ and a strictly upper-diagonal part N (i.e., $N_{i,j} = 0$ if $i \geq j$). If $||\cdot||$ is a sub-multiplicative norm, $||\Lambda|| \leq 1$ and $n \in \mathbb{N}$, then*

$$||X^n|| \leq ||\Lambda^n|| + C_{D,n} ||\Lambda||^{n-D+1} \max\{||N||, ||N||^{D-1}\}, \tag{8.103}$$

with $C_{D,n} = (D-1)n^{D-1}$. If in addition $2(D-1) \leq n$ then Eq.(8.103) holds with $C_{D,n} = (D-1)\binom{n}{D-1}$.

Similarly, if X is a Jordan block with diagonal part $\Lambda = \lambda\mathbb{1}$, then for all $n \in \mathbb{N}$ and natural numbers $k_0 \leq \min\{n, D-1\}$

$$|\lambda|^{n-k_0}\binom{n}{k_0} \leq ||X^n||_\infty \leq \sum_{k=0}^{\min\{n,D-1\}} |\lambda|^{n-k}\binom{n}{k}. \tag{8.104}$$

Proof Looking at the binomial expansion of $(\Lambda + N)^n$ we first observe that all monomials in which N appears more than $D-1$ times will vanish since they are products of more than $D-1$ strictly upper-triangular matrices. Hence, by using the basic properties of the norm we obtain

$$||X^n|| \leq ||\Lambda^n|| + \sum_{k=1}^{\min\{n,D-1\}} \binom{n}{k} ||N||^k ||\Lambda||^{n-k}, \tag{8.105}$$

which leads to Eq.(8.103) when bounding either $\binom{n}{k} \leq n^{D-1}$ or $\binom{n}{k} \leq \binom{n}{D-1}$ if $n \geq 2(D-1)$ and then choosing k extremal in each of the three factors.

Now assume that X is a Jordan block. Since then $[\Lambda, N] = 0$ we have $X^n = \sum_k \lambda^{n-k} N^k \binom{n}{k}$, where k again runs from 0 to $D-1$. X^n is thus a Topelitz matrix whose first row has entries of the form $\lambda^{n-k}\binom{n}{k}$. Since the largest singular value, i.e. the operator norm, is bounded from below by the largest modulus of all entries, we obtain the l.h.s. of Eq.(8.104). The r.h.s. is obtained by using the triangle inequality for the norm together with $||N||_\infty = 1$. □

Using this Lemma we will now derive two related results about the asymptotic behavior of T^n—based on either Jordan decomposition or Schur decomposition of T. While the first one allows us to pinpoint the exact asymptotic scaling, the second one enables us to provide explicit (albeit generally pessimistic) constants which depend only on n, μ and the underlying Hilbert space dimension.

For the following consider the Jordan decomposition $\hat{T} = A\big(\bigoplus_{k=1}^K J_k(\lambda_k)\big)A^{-1}$ where k labels the different eigenvalues λ_k of T according to their geometric multiplicity. That is, K is the number of eigenvectors. We define the *Jordan condition number* as $\kappa_T := \inf_A ||A||_\infty ||A^{-1}||_\infty$ where the infimum is taken over all similarity transformation A which bring $\hat{T}$ to Jordan normal form.

Theorem 8.23 (Asymptotic convergence I) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive and trace preserving linear map. Denote by $\mu := \sup_{\lambda \in \mathrm{spec}(T)}\{|\lambda| < 1\}$ the largest modulus of all eigenvalues which are in the interior of the unit disc, and by d_μ the dimension of the largest Jordan block corresponding to an eigenvalue of modulus μ. Then there exists a constant $C_1 > 0$ such that for all $n \in \mathbb{N}$*

$$C_1 \mu^n n^{d_\mu - 1} \leq ||\hat{T}^n - \hat{T}_\varphi^n||_\infty. \tag{8.106}$$

Moreover, this inequality holds for $C_1 = \begin{cases} \kappa_T^{-1} & \text{if } d_\mu = 1, \\ \kappa_T^{-1}\big(\mu(d_\mu - 1)\big)^{1-d_\mu} & \text{if } d_\mu \leq n+1. \end{cases}$

Similarly, there exists a $C_2 \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have

$$||\hat{T}^n - \hat{T}_\varphi^n||_\infty \leq C_2 \mu^n n^{d_\mu - 1}. \tag{8.107}$$

If $\hat{T}$ is diagonalizable (i.e., all Jordan blocks are one-dimensional), we can choose $C_2 = \kappa_T$.

PROOF By definition of T_φ we have that $\hat{T} - \hat{T}_\varphi = A\big(\bigoplus_{k:|\lambda_k|<1} J_k(\lambda_k)\big)A^{-1}$ (assuming, with some abuse of notation, that missing Jordan blocks are properly replaced by blocks of zero matrices). Exploiting Eq.(8.99) together with sub-multiplicativity of the norm we obtain

$$\kappa_T^{-1}||J^n||_\infty \leq ||\hat{T}^n - \hat{T}_\varphi^n||_\infty \leq \kappa_T||J^n||_\infty \quad \text{with} \quad J := \bigoplus_{k:|\lambda_k|<1} J_k(\lambda_k). \tag{8.108}$$

It follows from Eq.(8.104) that for every pair of eigenvalues with $|\lambda_i| < |\lambda_j|$ there is an $n_0 \in \mathbb{N}$ such that $||J_i(\lambda_i)^n||_\infty < ||J_j(\lambda_j)^n||_\infty$ holds for the norms of the corresponding Jordan blocks for all $n \geq n_0$. Moreover, if two eigenvalues have equal magnitude, then the Jordan block with the larger dimension has the larger norm.[14] Hence, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have

$$||J^n||_\infty \leq ||X^n||_\infty$$

where X is chosen to be the largest Jordan block corresponding to an eigenvalue of modulus μ. At this point the r.h.s. of Eq.(8.104) becomes applicable with $D = d_\mu$ and we arrive at the desired result in Eq.(8.107) by using that $\binom{n}{k} \leq n^{d_\mu - 1}$ is implied by $k \leq d_\mu - 1$. Note that we can drop the requirement that $n \geq n_0$ by appropriately rescaling C_2.

The lower bound stated in Eq.(8.106) is similarly obtained by using $||J^n||_\infty \geq ||X^n||_\infty$ and inserting the l.h.s. of Eq.(8.108) together with Eq.(8.104) after setting $k_0 = d_\mu - 1$ and bounding $\binom{n}{k_0} \geq (n/k_0)^{k_0}$. □

Note that if $\hat{T}$ is diagonalizable (as it holds for almost all channels), then the Jordan decomposition used in Thm.8.23 yields

$$\kappa_T^{-1}\mu^n \leq ||\hat{T}^n - \hat{T}_\varphi^n||_\infty \leq \kappa_T \mu^n \tag{8.109}$$

In some cases the Jordan condition number κ_T can be bounded in terms of the condition number $\kappa(\sigma) := ||\sigma||_\infty ||\sigma^{-1}||_\infty$ of a fixed point density matrix $\sigma = T(\sigma)$:

Proposition 8.5 (Jordan condition number and detailed balance) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a Hermiticity preserving linear map and $\Sigma : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ a linear map such that $\hat{\Sigma} > 0$ is a positive definite matrix. If*

$$\Sigma T^* = T\Sigma, \tag{8.110}$$

then the Jordan condition number satisfies $\kappa_T \leq \sqrt{\kappa(\hat{\Sigma})}$. In particular, if $\Sigma(X) = \sqrt{\sigma}X\sqrt{\sigma}$ for some positive definite $\sigma \in \mathcal{M}_d(\mathbb{C})$ (which is a fixed point of T if $T^*(\mathbb{1}) = \mathbb{1}$), then $\kappa_T \leq \sqrt{\kappa(\sigma)}$.

[14] This can be seen by first noting that for Jordan blocks of equal size $||J(\lambda)^n||_\infty = ||J(|\lambda|)^n||_\infty$ and then using that $J(|\lambda|)^n$ contains any smaller Jordan block with eigenvalue $|\lambda|$ as principal submatrix whose operator norm thus forms a lower bound.

PROOF Expressed in terms of the matrix representation $\hat{T}$ the "detailed balance" type condition of Eq.(8.110) reads $\hat{\Sigma}\hat{T}^\dagger = \hat{T}\hat{\Sigma}$. Since $\hat{\Sigma} > 0$ this is equivalent to $\hat{\Sigma}^{1/2}\hat{T}^\dagger\hat{\Sigma}^{-1/2} = \hat{\Sigma}^{-1/2}\hat{T}\hat{\Sigma}^{1/2}$ which is thus a Hermitian matrix so that $\hat{T}$ can be diagonalized by a similarity transformation $\hat{\Sigma}^{1/2}U$ for some unitary U. Since multilpying with the latter doesn't change the operator norm, we obtain $\kappa_T \leq \kappa\big(\hat{\Sigma}^{1/2}\big) = \sqrt{\kappa\big(\hat{\Sigma}\big)}$ which equals $\sqrt{\kappa(\sigma)}$ if $\hat{\Sigma} = \sigma^{1/2} \otimes \bar{\sigma}^{1/2}$. □

For the general non-diagonalizable case, the constant C_2 (constant w.r.t. n) in Thm.8.23 is somewhat elusive so that Eq.(8.107) becomes a rather weak statement. A look at the proof reveals that we can obtain an explicit C_2 by either making it dependent on the separation of the moduli of eigenvalues, or by replacing d_μ in Eq.(8.107) by the dimension of the largest Jordan block corresponding to *any* eigenvalue in the interior of the unit disk. In the latter case C_2 would still involve the Jordan condition number κ_T which in turn may be bounded in terms of the separation of eigenvalues (see [?]). The subsequent theorem will follow an alternative route which, to some extent, allows to circumvent such issues. It is based on the Schur decomposition $(\hat{T} - \hat{T}_\varphi) = U(\Lambda + N)U^\dagger$ where U is unitary, Λ is diagonal and N is strictly upper triangular.

Theorem 8.24 (Asymptotic convergence II) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive and trace preserving linear map. Denote by $\mu := \sup_{\lambda\in\mathrm{spec}(T)}\{|\lambda| < 1\}$ the largest modulus of all eigenvalues which are in the interior of the unit disc. Then for any $n \in \mathbb{N}$*

$$||\hat{T}^n - \hat{T}^n_\varphi||_\infty \leq \mu^n + C_{d,n}\mu^{n-d^2+1}\max\{||N||_\infty, ||N||_\infty^{d^2-1}\}, \tag{8.111}$$

with $C_{d,n} = (d^2-1)n^{d^2-1}$. If in addition $2(d^2-1) \leq n$ then Eq.(8.111) holds with $C_{d,n} = (d^2-1)\binom{n}{d^2-1}$. Moreover, we can bound $||N||_\infty \leq (\mu + 2\sqrt{d})$.

PROOF Eq.(8.111) is an immediate consequence of Eq.(8.103) together with the spectral properties of T. The operator norm bound for N follows from $||N||_\infty \leq ||\Lambda||_\infty + ||\hat{T}||_\infty + ||\hat{T}_\varphi||_\infty$ since $||\Lambda||_\infty = \mu$ and the singular values of positive, trace-preserving linear maps on $\mathcal{M}_d(\mathbb{C})$ are bounded by $\sqrt{d}$. □

We will now relate the operator norm convergence of maps, which we have discussed in the last paragraphs, to the trace norm convergence of states. Recall that $\mathcal{X}_T := \mathrm{span}\{X \in \mathcal{M}_d(C) | \exists\varphi \in \mathbb{R} : T(X) = e^{i\varphi}X\}$ is the complex linear span of all eigenvectors of T with corresponding eigenvalue of modulus one. According to Prop.6.12 $\mathcal{X}_T$ admits a basis consisting of density operators and satisfies $\mathcal{X}_T = T(\mathcal{X}_T) = T_\phi\big(\mathcal{M}_d(\mathbb{C})\big)$.

Proposition 8.6 (Convergence towards asymptotic states) *Let $T : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be a positive and trace preserving linear map and denote by $\Delta_T(\rho) := ||\rho - T_\phi(\rho)||_1$ the trace-norm distance between $\rho \in \mathcal{M}_d(\mathbb{C})$ and its projection onto $\mathcal{X}_T$. Then for any density operator $\rho \in \mathcal{M}_d(C)$ and $n \in \mathbb{N}$*

$$\Delta_T\big(T^n(\rho)\big) \leq \sqrt{d/2}\; ||\hat{T}^n - \hat{T}^n_\varphi||_\infty\; \Delta_T(\rho). \tag{8.112}$$

Conversely, if the supremum is taken over all density operators $\rho \in \mathcal{M}_d(C)$, then

$$\sup_\rho \Delta_T\big(T^n(\rho)\big) \geq \frac{1}{4\sqrt{d}} ||\hat{T}^n - \hat{T}^n_\varphi||_\infty. \tag{8.113}$$

Proof First note that since $T_\phi T = TT_\phi = T_\varphi = T_\varphi T_\phi$ we have $\Delta_T\big(T^n(\rho)\big) = ||T^n\big(\rho - T_\phi(\rho)\big)||_1 = ||(T^n - T^n_\varphi)\big(\rho - T_\phi(\rho)\big)||_1$. So for any density operator ρ we obtain

$$\begin{aligned}
\frac{\Delta_T\big(T^n(\rho)\big)}{\Delta_T(\rho)} &= \frac{||(T^n - T^n_\varphi)\big(\rho - T_\phi(\rho)\big)||_1}{||\rho - T_\phi(\rho)||_1} && (8.114)\\
&\leq \frac{\sqrt{d}}{2} \sup_{\psi\perp\phi} ||(T^n - T^n_\varphi)\big(|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\big)||_2 && (8.115)\\
&\leq \sqrt{d/2} \sup_{X\in\mathcal{M}_d(\mathbb{C})} \frac{||(T^n - T^n_\varphi)(X)||_2}{||X||_2} && (8.116)\\
&= \sqrt{d/2}\, ||(T^n - T^n_\varphi)||_{2\to 2}. && (8.117)
\end{aligned}$$

Here the first inequality uses Lemma 8.3 (where the optimization is over all pairs of orthogonal unit vectors in $\mathbb{C}^d$) together with $||\cdot||_1 \leq \sqrt{d}||\cdot||_2$ from Eq.(8.7). The second inequality uses that $||\ |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\ ||_2 = \sqrt{2}$. Since $||(T^n - T^n_\varphi)||_{2\to 2} = ||(\hat{T}^n - \hat{T}^n_\varphi)||_\infty$ we finally arrive at the inequality in Eq.(8.112).

For the converse statement recall that $||(T^n - T^n_\varphi)||_{2\to 2} = \sup_{X:||X||_2\leq 1} ||(T^n - T^n_\varphi)(X)||_2$. Any such X can be decomposed into four positive parts so that $X = \sum_{i=1}^4 c_i P_i$ with $|c_i| = 1$ and $P_i \geq 0$, $||P_i||_2 \leq 1$. Together with $||\cdot||_2 \leq ||\cdot||_1$ this implies

$$\frac{1}{4}||(T^n - T^n_\varphi)||_{2\to 2} \leq \sup_{P\geq 0} \big\{||(T^n - T^n_\varphi)(P)||_1 \big|\ ||P||_2 \leq 1\big\}.$$

Eq.(8.113) then follows by observing that $||P||_2 \leq 1$ implies $\mathrm{tr}\,[P] \leq \sqrt{d}$. □

8.8 Continuity bounds

Chapter 9

Symmetries

Chapter 10

Special Channels

Chapter 11

Quantum Spin Chains

Bibliography

Index