Quantum information theory (MAT4430) **Spring 2021**

Lecture 2: Toolbox for quantum mechanics

Lecturer: Alexander Müller-Hermes

In this lecture, we will collect some tools needed to study quantum information theory. We will briefly introduce complex Euclidean spaces and classes of normal operators on these spaces. Particular emphasize will be put on positive operators, which will represent states in quantum mechanics.

1 Complex Euclidean spaces

Modern quantum mechanics is defined in the language of functional analysis, using the notion of complex *Hilbert spaces*. Since large parts of quantum information theory can be studied using finite-dimensional spaces, we will restrict to this case throughout this course. We start with the following definition:

Definition 1.1 (Complex Euclidean space)**.** *A complex Euclidean space is a finite-dimensional vector space $\mathcal{H}$ equipped with a complex form $\langle\cdot|\cdot\rangle : \mathcal{H}\times\mathcal{H}\to\mathbb{C}$ with the following properties:*

1. *We have $\langle x|x\rangle \geqslant 0$ for every $x\in\mathcal{H}$, with equality if and only if $x=0$.*
2. *We have $\langle x+\lambda y|z\rangle = \langle x|y\rangle + \overline{\lambda}\langle y|z\rangle$ for every $x,y,z\in\mathcal{H}$ and every $\lambda\in\mathbb{C}$.*
3. *We have $\langle x|y\rangle = \overline{\langle y|x\rangle}$ for every $x,y\in\mathcal{H}$.*

These properties guarantee that the function $x\mapsto\sqrt{\langle x|x\rangle}$ defines a norm on $\mathcal{H}$. This norm is called the Euclidean norm *or* 2-norm *and it is denoted by $\|\cdot\|_2$.*

Note that we define the inner product to be conjugate-linear in the first argument and linear in the second argument, which is the usual convention in the physics literature (contrary to the mathematics literature).

Every vector space has a basis, but on Euclidean spaces there is a stronger notion:

Definition 1.2 (Orthonormal basis)**.** *An* orthonormal basis *is a set of vectors $\{x_1,\dots,x_n\}$ that is maximal with the following properties*[1]*:*

1. *We have $\|x_i\|_2 = \langle x_i|x_i\rangle = 1$ for all $i\in\{1,\dots,n\}$.*
2. *We have $\langle x_i|x_j\rangle = 0$ whenever $i\neq j$.*

An orthonormal basis is, in particular, a basis. Hence, the cardinality $n\in\mathbb{N}$ of an orthonormal basis of $\mathcal{H}$ coincides with the dimension $\dim(\mathcal{H})$. Given an orthonormal basis $\{x_1,\dots,x_n\}$ of a complex Euclidean space $\mathcal{H}$, we can write any element $y\in\mathcal{H}$ as a linear combination

$$y = \sum_{i=1}^{n}\langle x_i|y\rangle x_i. \tag{1}$$

Another way of saying this, is that the identity map $\mathbb{1}_{\mathcal{H}}:\mathcal{H}\to\mathcal{H}$ admits the decomposition

$$\mathbb{1}_{\mathcal{H}} = \sum_{i=1}^{n}\langle x_i|\cdot\rangle x_i. \tag{2}$$

Recall the following theorem:

[1] i.e., it cannot be extended by an additional vector x_{n+1} such that the properties still hold.

Theorem 1.3. *Every complex Euclidean space has an orthonormal basis.*

By choosing an orthonormal basis, any complex Euclidean $\mathcal{H}$ can be identified with $\mathbb{C}^d$ equipped with the standard inner product and for $d = \dim(\mathcal{H})$. *Throughout this course, we will always make this identification implicitely, and when saying "complex Euclidean space" we always mean $\mathbb{C}^d$ for some $d \in \mathbb{N}$.* We will usually denote complex Euclidean spaces by $\mathcal{H}$ and we will be using indices such as $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_{AB}, \ldots$ to keep track of different Euclidean spaces. Moreover, we will sometimes omit the "complex" in "complex Euclidean spaces" for brevity: When saying "Euclidean space" we will always mean "complex Euclidean space" in these lecture notes.

2 Bra-ket notation

A fundamental property of (complex) Euclidean spaces is a canonical identification with their dual space:

Definition 2.1 (Dual space)**.** *For a complex Euclidean space $\mathcal{H}$ the* dual space *is given by*

$$\mathcal{H}^* = \{f : \mathcal{H} \to \mathbb{C} \ : \ f \text{ linear}\}.$$

The following theorem is a fundamental result in the theory of complex Euclidean spaces (and more general Hilbert spaces).

Theorem 2.2 (Baby version of Riesz representation)**.** *For any $f \in \mathcal{H}^*$ there is a unique $y \in \mathcal{H}$ satisfying*

$$f(x) = \langle y|x\rangle,$$

for all $x \in \mathcal{H}$.

Proof. Just evaluate $f \in \mathcal{H}^*$ on an orthonormal basis of $\mathcal{H}$, and use these values to build the vector $y \in \mathcal{H}$. □

Most researchers in quantum information theory use the so-called *bra-ket notation*, and we will do the same throughout the course. The basic idea of this notation is to cleverly encode the identification between a complex Euclidean space $\mathcal{H}$ and its dual space $\mathcal{H}^*$:

- Vectors $|x\rangle \in \mathcal{H}$ are called *kets*.
- Functionals $\langle y| \in \mathcal{H}^*$ are called *bras*.
- Applying a bra to a ket yields a *bra(c)ket* $\langle y|x\rangle \in \mathbb{C}$, i.e., the inner product on $\mathcal{H}$.
- We write $|x\rangle\langle y| : \mathcal{H} \to \mathcal{H}$ for the linear map acting as

  $$|x\rangle\langle y|\,(|z\rangle) = \langle y|z\rangle|x\rangle$$

  on vectors $|z\rangle \in \mathcal{H}$.

Note that Theorem 2.2 is build into this notation by the identification $|x\rangle \leftrightarrow \langle x|$. We will occasionally need to fix a particular orthonormal basis on complex Euclidean spaces. Since we always consider the concrete spaces $\mathbb{C}^d$, we can define a canonical basis on them:

Definition 2.3 (Computational basis)**.** *The computational basis of* $\mathbb{C}^d$ *is given by* $\{|1\rangle, \ldots, |d\rangle\}$*, where*

$$|i\rangle = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

with the 1 *appearing in the* i*th position.*

The term "computational basis" is inspired by quantum computing, where a quantum system is used to solve computational tasks and where the computational basis serves as a reference for information stored in the system. Using the computational basis, we can express the kets living in $\mathbb{C}^d$ as "column vectors", and consequently we should think of the corresponding bras as "row vectors". For vectors

$$|x\rangle = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \quad \text{and} \quad |y\rangle = \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$$

in $\mathbb{C}^d$ we can therefore compute their inner product in the usual way, by setting

$$\langle x|y\rangle = (\overline{x_1}, \ldots, \overline{x_d}) \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} = \sum_i \overline{x_i} y_i.$$

The bra-ket notation has many advantages when working with complex Euclidean spaces. For example, (2) can be written in a very memorable way as

$$\mathbb{1}_{\mathcal{H}} = \sum_{i=1}^{n} |x_i\rangle\langle x_i|,$$

whenever $\{|x_1\rangle, \ldots, |x_n\rangle\}$ is an orthonormal basis of $\mathcal{H}$, and this expression can easily be inserted into complicated equations. We will see more examples later.

3 Linear operators and the Hilbert-Schmidt inner product space

In the following, let $\mathcal{H} = \mathbb{C}^d$, $\mathcal{H}_A = \mathbb{C}^{d_A}$ and $\mathcal{H}_B = \mathbb{C}^{d_B}$ denote complex Euclidean spaces. For any linear operator $L : \mathcal{H}_A \to \mathcal{H}_B$, we define the *operator norm* as

$$\|L\| = \sup_{x \in \mathcal{H}_A} \frac{\|Lx\|_2}{\|x\|_2}.$$

We will denote by $B(\mathcal{H}_A, \mathcal{H}_B)$ the set of (bounded) linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$, and by $B(\mathcal{H})$ the set of (bounded) linear operators from $\mathcal{H}$ into itself. Of course, every linear operator between complex Euclidean spaces (which are by definition finite-dimensional) is automatically bounded. We choose this notation anyway, since it is standard in functional analysis. Using the computational basis, we can identify operators in $B(\mathbb{C}^{d_A}, \mathbb{C}^{d_B})$ with $d_B \times d_A$-matrices with complex entries. Throughout this course, we will often use these two descriptions interchangeably.

Let us briefly mention some operations on linear operators: Given linear operators $L_1, L_2 \in B(\mathcal{H})$, we denote their *commutator* by

$$[L_1, L_2] = L_1 L_2 - L_2 L_1,$$

and their *anticommutator* by

$$\{L_1, L_2\} = L_1 L_2 + L_2 L_1.$$

We say that the operators L_1 and L_2 *commute* if $[L_1, L_2] = 0$ and that they *anticommute* if $\{L_1, L_2\} = 0$. These notions will play an important role in the foundations of quantum mechanics. We will also need the trace functional: After fixing an orthonormal basis $\{x_1, \ldots, x_d\}$ of $\mathcal{H}$, we may define a functional $\mathrm{Tr} : B(\mathcal{H}) \to \mathbb{C}$ by

$$\mathrm{Tr}[L] = \sum_{i=1}^{d} \langle x_i | L | x_i \rangle,$$

for any $L \in B(\mathcal{H})$. It can be shown, that this definition does not depend on the choice of the orthonormal basis, and we might as well consider the computational basis. Therefore, we can take it as a general definition of the *trace*. Note that $\mathrm{Tr}[L] = \sum_{i=1}^{d} \lambda_i$, where λ_i are the eigenvalues of L, i.e., the d complex roots of the characteristic polynomial with multiplicities. This will sometimes be useful.

Next, we will review the notion of adjoints:

Definition 3.1 (Adjoint operator)**.** *For any linear operator $L \in B(\mathcal{H}_A, \mathcal{H}_B)$ we define the* adjoint *$L^\dagger \in B(\mathcal{H}_B, \mathcal{H}_A)$ as the unique operator satisfying*

$$\langle y | L x \rangle = \langle L^\dagger y | x \rangle,$$

for every $|x\rangle \in \mathcal{H}_A$ and $|y\rangle \in \mathcal{H}_B$.

The adjoint operation takes a concrete form when expressed in the computational basis, and we have

$$L^\dagger = \overline{L}^T,$$

where $\overline{(\cdot)}$ denotes the entrywise complex conjugation, and $(\cdot)^T$ the matrix transpose in the computational basis. If not stated otherwise, all transposes and complex conjugations of operators should be understood in the computational basis (although it is usually possible to choose another basis without changing the definitions).

The vector space $B(\mathcal{H}_A, \mathcal{H}_B)$ is itself a complex Euclidean space when equipped with the *Hilbert-Schmidt* inner product

$$\langle L_1, L_2 \rangle_{HS} = \mathrm{Tr}\left[L_1^\dagger L_2\right],$$

for $L_1, L_2 \in B(\mathcal{H}_A, \mathcal{H}_B)$. Using the computational basis, we can define a canonical orthonormal basis for the Hilbert-Schmidt inner product space $B(\mathbb{C}^{d_A}, \mathbb{C}^{d_B})$:

Definition 3.2 (Matrix units)**.** *For $i \in \{1, \ldots, d_A\}$ and $j \in \{1, \ldots, d_B\}$ we define the* matrix units *as*

$$|i\rangle\langle j| = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

with the 1 appearing in the (i, j)th position. The matrix units form an orthonormal basis of $B(\mathbb{C}^{d_A}, \mathbb{C}^{d_B})$ with the Hilbert-Schmidt inner product.

The Hilbert-Schmidt inner product can be understood as the usual inner product on $\mathbb{C}^{d_A d_B}$ by "viewing matrices as vectors", i.e., by taking the entries of the corresponding matrix as the entries of a vector. The corresponding 2-norm, given by $\|L\|_{HS} = \sqrt{\langle L, L\rangle_{HS}}$, is called the *Hilbert-Schmidt norm* or *Frobenius norm*.

Many important classes of operators can be defined by properties involving their adjoints. A linear operator $L \in B(\mathcal{H})$ is called

- *normal*, if $\left[L, L^\dagger\right] = 0$.
- *selfadjoint*, if $L = L^\dagger$.
- *projection*[2], if $L = L^\dagger = L^2$.
- *unitary*, if $L^{-1} = L^\dagger$.

We will denote the set of selfadjoint operators on $\mathcal{H}$ by $B(\mathcal{H})_{sa}$, the set of unitary operators on $\mathcal{H}$ by $\mathcal{U}(\mathcal{H})$, and the set of projections by $\mathrm{Proj}(\mathcal{H})$. All of these are subsets of the normal operators. We will often need the following normal form (no pun intended) of normal operators:

Theorem 3.3 (Spectral decomposition)**.** *For any normal operator $L \in B(\mathcal{H})$, there exists an orthonormal basis $\{|x_1\rangle, \ldots, |x_d\rangle\}$ of $\mathcal{H}$ consisting of eigenvectors of L such that*

$$L = \sum_{i=1}^{d} \lambda_i |x_i\rangle\langle x_i|,$$

where $\lambda_i \in \mathbb{C}$ are the eigenvalues of L corresponding to the eigenvector $|x_i\rangle$.

The spectral decomposition is often referred to as a *diagonalization*, since the matrix corresponding to the operator L in the orthonormal basis $\{|x_1\rangle, \ldots, |x_d\rangle\}$ is diagonal. If two normal operators $L_1, L_2 \in B(\mathcal{H})$ commute, then there exists an orthonormal basis $\{|x_1\rangle, \ldots, |x_d\rangle\}$ of joint eigenvectors such that

$$L_1 = \sum_{i=1}^{d} \lambda_i |x_i\rangle\langle x_i| \quad \text{and} \quad L_2 = \sum_{i=1}^{d} \mu_i |x_i\rangle\langle x_i|,$$

where $\lambda_i, \mu_i \in \mathbb{C}$ are the eigenvalues of L_1 and L_2, respectively. In this case, we also say that L_1 and L_2 are *simultaneously diagonalizable*.

4 Applying functions to (normal) operators

How can we apply functions such as exp or log to operators in $B(\mathcal{H})$? A priori, there is no inherently correct way of defining an expression like $\exp(X)$ for a linear operator $X \in B(\mathcal{H})$, and it would be nice if this could be done in some consistent way preserving the usual composition rules of functions. Constructions to achieve this goal run under the name "functional calculus", and they differ in the kinds of functions and subsets of $B(\mathcal{H})$ they apply to. For this course, we will mostly need to apply functions on normal operators, and here the spectral decomposition will lead to a general theory. For completeness, we will briefly mention two classes of functions where it is intuitively clear how to apply them to operators.

[2]Some authors would call this an *orthogonal projection*, and they would refer to an operator satisfying $L = L^2$ as a projection. Throughout our course, we will only consider orthogonal projections and we will simply call them "projections" for brevity.

First, consider a function $f : \mathbb{C} \to \mathbb{C}$ given by a polynomial, i.e.,

$$f(x) = a_k x^k + a_{k-1} x^{k-1} + \cdots a_1 x + a_0,$$

with complex coefficients $a_0, \ldots, a_k \in \mathbb{C}$. Since $B(\mathcal{H})$ is an algebra with unit element $\mathbb{1}_{\mathcal{H}}$, it is clear how to define the operator $f(X) \in B(\mathcal{H})$ for any linear operator $X \in B(\mathcal{H})$. We simply set

$$f(X) = a_k X^k + a_{k-1} X^{k-1} + \cdots a_1 X + a_0 \mathbb{1}_{\mathcal{H}} \in B(\mathcal{H}).$$

From this example it is also clear how to apply certain non-polynomial functions on operators. Consider an entire function $f : \mathbb{C} \to \mathbb{C}$, i.e., a function f that is analytic everywhere in $\mathbb{C}$. Such a function can be expressed as a power series $f(x) = \sum_{k=0}^{\infty} a_k x^k$ which is absolutely convergent in $\mathbb{C}$. Consider a linear operator $X \in B(\mathcal{H})$. How should we define $f(X)$? To make this expression consistent with the expression for polynomials, we should define

$$f(X) = \lim_{N \to \infty} \sum_{k=0}^{N} a_k X^k =: \sum_{k=0}^{\infty} a_k X^k,$$

where we set $X^0 = \mathbb{1}_{\mathcal{H}}$. Does this define a linear operator? Yes, it does! It is easy to check that

$$\|f(X)\| \leqslant \sum_{k=0}^{\infty} |a_k| \|X\|^k < \infty.$$

An example where this construction can be applied is the exponential function $\exp(X)$, which is defined for any linear operator $X \in B(\mathcal{H})$ on the Euclidean space $\mathcal{H}$.

Obviously, not all complex functions are entire and during the course we will occasionally apply the logarithm to linear operators. Luckily, we do not need to apply such functions to any linear operator $X \in B(\mathcal{H})$, but only to normal operators. Here, the spectral theorem can be applied:

Definition 4.1 (Functions of normal operators)**.** *Let $D \subseteq \mathbb{C}$ denote some subset of complex numbers, $f : D \to \mathbb{C}$ a complex function and $X \in B(\mathcal{H})$ a normal operator with spectrum $\sigma(X) \subseteq D$. Then, we define*

$$f(X) = \sum_{i=1}^{d} f(\lambda_i) |x_i\rangle\langle x_i|,$$

where $\{|x_1\rangle, \ldots, |x_d\rangle\}$ is an orthonormal basis of eigenvectors of X (which exists by Theorem 3.3).

The reader should convince themselves that Definition 4.1 for a normal operator X gives the same operator $f(X)$ as derived above when f is a polynomial or an entire function.

5 Positive operators

We will also need the set of positive semidefinite operators:

Definition 5.1 (Positive operators)**.** *A linear operator $P \in B(\mathcal{H})$ is called* positive semidefinite *if*

$$\langle x|P|x\rangle \geqslant 0,$$

for any $|x\rangle \in \mathcal{H}$. If the previous inequality is always strict, we will call the operator positive definite*.*

For brevity, we will usually refer to positive semidefinite operators simply as *positive operators*. Note that this includes the zero operator! This is in contrast to real numbers, where we usually say *non-negative* to denote the positive real numbers and zero. The following theorem collects alternative characterizations of positive operators:

Theorem 5.2 (Characterizing positive operators)**.** *For a linear operator $P \in B(\mathcal{H})$ the following are equaivalent:*

1. *P is positive.*
2. *P is selfadjoint and has non-negative eigenvalues.*
3. *There exists a positive operator $Q \in B(\mathcal{H})$ such that $P = Q^2$.*
4. *There exists an operator $X \in B(\mathcal{H})$ such that $P = X^\dagger X$.*
5. *There exists an operator $Y \in B(\mathcal{H}, \mathcal{H}')$ for some Euclidean space $\mathcal{H}'$ such that $P = Y^\dagger Y$.*

The operator Q in 3. *is unique and it is called the* positive square root *of P. We will write $\sqrt{P}$ or $P^{1/2}$ to denote it.*

Proof. We will first show that 1. and 2. are equivalent. Assume that P is positive as in Definition 5.1. For any $|x\rangle, |y\rangle \in \mathcal{H}$ we have

$$\langle x + iy|P|x + iy\rangle = \langle x|P|x\rangle + \langle y|P|y\rangle + i\left(\langle x|P|y\rangle - \langle y|P|x\rangle\right) \geqslant 0,$$

and we conclude that $\langle x|P|y\rangle - \langle y|P|x\rangle = 0$. Therefore, P is selfadjoint. Clearly, P has positive eigenvalues, since otherwise, there would be a normalized eigenvector $|x\rangle$ satisfying $\langle x|P|x\rangle < 0$ and contradicting our assumption.

For the other direction, assume that P is selfadjoint (and in particular normal) and has positive eigenvalues. By the spectral decomposition, Theorem 3.3, we can write

$$P = \sum_{i=1}^{d} \lambda_i |x_i\rangle\langle x_i|, \tag{3}$$

for an orthonormal basis $\{|x_1\rangle, \ldots, |x_d\rangle\}$ of $\mathcal{H}$ consisting of eigenvectors of P, and with the positive eigenvalues $\lambda_i \geqslant 0$. For any $|y\rangle \in \mathcal{H}$ we can verify that

$$\langle y|P|y\rangle = \sum_{i=1}^{d} \lambda_i |\langle y|x_i\rangle|^2 \geqslant 0,$$

and hence P is positive.

Now, we show the rest of the equivalences. Assume that P is selfadjoint and has positive eigenvalues, and consider the spectral decomposition (3). We can now define a linear operator $Q \in B(\mathcal{H})$ by

$$Q = \sum_{i=1}^{d} \sqrt{\lambda_i} |x_i\rangle\langle x_i|.$$

By the equivalence of 1. and 2., we conclude that Q is positive. Furthermore, it is easy to check that $Q^2 = P$ showing that 3. holds.

It is clear that 3. implies 4. and that 4. implies 5.. Finally, we assume that there exists an operator $Y \in B(\mathcal{H}, \mathcal{H}')$ for some Euclidean space $\mathcal{H}'$ such that $P = Y^\dagger Y$. For any $|x\rangle \in \mathcal{H}$ we have

$$\langle x|P|x\rangle = \langle x|Y^\dagger Y|x\rangle = \langle Yx|Yx\rangle \geqslant 0,$$

and hence P is positive.

Finally, we need to show that the positive square root in 3. is unique. By polynomial interpolation, there exists a polynomial $f \in \mathbb{R}[X]$ with real coefficients satisfying $f(\lambda_i) = \sqrt{\lambda_i}$ for all eigenvalues $\lambda_1, \ldots, \lambda_d \geqslant 0$ of P. By the spectral theorem, we conclude that

$$Q = f(P). \tag{4}$$

Assume now, that there is another positive operator $\tilde{Q} \in B(\mathcal{H})^+$ satisfying $\tilde{Q}^2 = P$. By the spectral theorem we have $\tilde{Q} = \sum_{i=1}^d \mu_i |v_i\rangle\langle v_i|$ and therefore $P = \sum_{i=1}^d \mu_i^2 |v_i\rangle\langle v_i|$ for some $\mu_i \geqslant 0$ and an orthonormal basis of eigenvectors $\{|v_1\rangle, \ldots, |v_d\rangle\} \subset \mathcal{H}$. Since $\left[\tilde{Q}, P\right] = 0$, we conclude, by (4), that

$$\left[Q, \tilde{Q}\right] = 0.$$

Now, we know that there exists a common orthonormal basis of eigenvectors $\{|w_1\rangle, \ldots, |w_d\rangle\} \subset \mathcal{H}$ such that $Q = \sum_{i=1}^d \nu_i |w_i\rangle\langle w_i|$ and $\tilde{Q} = \sum_{i=1}^d \mu_i |w_i\rangle\langle w_i|$. By assumption, we have $\nu_i^2 = \mu_i^2$ for every $i \in \{1, \ldots, d\}$. Since $\nu_i, \mu_i \geqslant 0$, we conclude that $\mu_i = \mu_i'$ for all $i \in \{1, \ldots, d\}$ and therefore $Q = \tilde{Q}$. □

Based on the notion of positivity, there is another useful decomposition of general linear operators:

Theorem 5.3 (Singular value decomposition)**.** *For any linear operator $L \in B(\mathcal{H}_A, \mathcal{H}_B)$ there exist orthonormal bases $\{|x_1\rangle, \ldots, |x_{d_A}\rangle\}$ of $\mathcal{H}_A$ and $\{|y_1\rangle, \ldots, |y_{d_B}\rangle\}$ of $\mathcal{H}_B$ such that*

$$L = \sum_{i=1}^{R} s_i |y_i\rangle\langle x_i|,$$

for strictly positive $s_i \in \mathbb{R}^+$ called singular values *of L, and where R equals the rank of L.*

The singular value decomposition can also be written in a non-reduced form by setting

$$L = \sum_{i=1}^{D} s_i |y_i\rangle\langle x_i|,$$

with $D = \min(d_A, d_B)$ and where some singular values may be zero. Sometimes this is convenient, e.g., when we want to avoid mentioning the rank of L.

Proof. Consider the positive operators $L^\dagger L \in B(\mathcal{H}_A)$ and $LL^\dagger \in B(\mathcal{H}_B)$ and assume without loss of generality that $d_A \geqslant d_B$ (otherwise, exchange the roles of L and $L^\dagger$ in the following proof). By the spectral decomposition, there exists an orthonormal basis $\{|x_1\rangle, \ldots, |x_{d_A}\rangle\}$ of $\mathcal{H}_A$ consisting of eigenvectors of $L^\dagger L$ such that

$$L^\dagger L = \sum_{i=1}^{d_A} \lambda_i |x_i\rangle\langle x_i|,$$

where $\lambda_i \geqslant 0$ are the eigenvalues of $L^\dagger L$. We set

$$R = \#\{i \in \{1, \ldots, d_A\} \ : \ \lambda_i \neq 0\},$$

and we assume without loss of generality that $\lambda_1, \ldots, \lambda_R \neq 0$ and $\lambda_i = 0$ for any $i > R$. Since

$$\|L|x_i\rangle\|_2^2 = \langle x_i | L^\dagger L | x_i \rangle = \lambda_i,$$

we conclude that $L|x_i\rangle = 0$ if and only if $\lambda_i = 0$, and therefore R coincides with the rank of the linear operator L. Next, observe that

$$LL^\dagger\left(L|x_i\rangle\right) = L\left(L^\dagger L|x_i\rangle\right) = \lambda_i L|x_i\rangle,$$

for any $i \in \{1,\ldots,d_A\}$. For any $i \in \{1,\ldots,R\}$ we can define an eigenvector

$$|y_i\rangle := \frac{L|x_i\rangle}{\sqrt{\lambda_i}} \neq 0,$$

of the operator $LL^\dagger$ corresponding to eigenvalue λ_i. With this definition, we can verify that

$$\langle y_i|y_j\rangle = \frac{\langle x_i|L^\dagger L|x_j\rangle}{\sqrt{\lambda_i\lambda_j}} = \delta_{ij},$$

for all $i,j \in \{1,\ldots,R\}$, and we conclude that $\{|y_1\rangle,\ldots,|y_R\rangle\}$ is a set of orthonormal vectors in $\mathcal{H}_B$. Extending this set gives an orthonormal basis $\{|y_1\rangle,\ldots,|y_{d_B}\rangle\}$. Finally, we can check that

$$L|x_i\rangle = \sqrt{\lambda_i}|y_i\rangle,$$

for all $i \in \{1,\ldots,R\}$ and, since R is the rank of L, we obtain the singular value decomposition with singular values $s_i = \sqrt{\lambda_i}$.

□

6 Tensor products

The second construction we need are tensor products, which can be defined in various ways, and usually this is done quite abstractly. We will use a more concrete definition, which is sometimes called the *Kronecker product*:

Definition 6.1 (Tensor product)**.** *For complex Euclidean spaces $\mathcal{H}_A = \mathbb{C}^{d_A}$ and $\mathcal{H}_B = \mathbb{C}^{d_B}$ the* tensor product *is given by*

$$\mathcal{H}_A \otimes \mathcal{H}_B = \Big\{f : \{1,\ldots,d_A\} \times \{1,\ldots,d_B\} \to \mathbb{C}\Big\},$$

where the vector addition is point-wise and we have the inner product

$$\langle f,g\rangle = \sum_{i,j} \overline{f(i,j)}g(i,j).$$

For vectors

$$|x\rangle = \begin{pmatrix} x_1 \\ \vdots \\ x_{d_A} \end{pmatrix} \quad \text{and} \quad |y\rangle = \begin{pmatrix} y_1 \\ \vdots \\ y_{d_B} \end{pmatrix}$$

we define the tensor product $|x\rangle \otimes |y\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ of vectors as

$$\left(|x\rangle \otimes |y\rangle\right)(i,j) = x_i y_j,$$

for any $i \in \{1,\ldots,d_A\}$ and $j \in \{1,\ldots,d_B\}$. The vectors $|x\rangle \otimes |y\rangle$ are also called elementary tensors *and they span the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$.*

Note that the previous definition extends to tensor products of higher order, such as

$$\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C = \Big\{f : \{1,\ldots,d_A\} \times \{1,\ldots,d_B\} \times \{1,\ldots,d_C\} \to \mathbb{C}\Big\}$$

and it can be verified easily that the tensor product of vectors is associative. The following identities can be verified easily for all $\alpha,\beta \in \mathbb{C}$, all $|x\rangle, |x_1\rangle, |x_2\rangle \in \mathcal{H}_A$ and all $|y\rangle, |y_1\rangle, |y_2\rangle \in \mathcal{H}_B$:

- $(|x_1\rangle + \alpha|x_2\rangle) \otimes |y\rangle = |x_1\rangle \otimes |y\rangle + \alpha\,(|x_2\rangle \otimes |y\rangle)$.
- $|x\rangle \otimes (|y_1\rangle + \beta|y_2\rangle) = |x\rangle \otimes |y_1\rangle + \beta\,(|x\rangle \otimes |y_2\rangle)$.
- $\langle x_1 \otimes x_2|y_1 \otimes y_2\rangle = \langle x_1|y_1\rangle\langle x_2|y_2\rangle$.

Note that the tensor product is *not* commutative! The tensor product $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ can be identified with $\mathbb{C}^{d_A d_B}$ by identifying the computational basis of the tensor product with the tensor products of the computational bases of $\mathcal{H}_A = \mathbb{C}^{d_A}$ and $\mathcal{H}_B = \mathbb{C}^{d_B}$. This identification leads to the computational basis of $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ given by

$$\{|i\rangle \otimes |j\rangle \ : \ i \in \{1, \ldots, d_A\}, j \in \{1, \ldots, d_B\}\}.$$

The following theorem shows a fundamental property of tensor products, which is sometimes used to define the tensor product in the first place:

Theorem 6.2 (Universal property)**.** *Consider complex Euclidean spaces $\mathcal{H}_A$, $\mathcal{H}_B$ and $\mathcal{H}_C$ and a bilinear map $f : \mathcal{H}_A \times \mathcal{H}_B \to \mathcal{H}_C$, i.e., such that the maps*

$$|x\rangle \mapsto f(|x\rangle, |y'\rangle) \quad \text{and} \quad |y\rangle \mapsto f(|x'\rangle, |y\rangle),$$

are linear for every $|x'\rangle \in \mathcal{H}_A$ and every $|y'\rangle \in \mathcal{H}_B$. Then, there exists a unique linear map $F : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_C$ satisfying

$$F(|x\rangle \otimes |y\rangle) = f(|x\rangle, |y\rangle),$$

for every $|x\rangle \in \mathcal{H}_A$ and every $|y\rangle \in \mathcal{H}_B$.

Proof. Define the linear map $F : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_C$ by linearly extending

$$F(|i\rangle \otimes |j\rangle) = f(|i\rangle, |j\rangle).$$

Since $\{|i\rangle \otimes |j\rangle\}_{ij}$ is a basis of $\mathcal{H}_A \otimes \mathcal{H}_B$, this defines a unique linear map and the desired property can be checked easily. □

How do operators act between tensor products? Given operators $X \in B(\mathcal{H}_A, \mathcal{H}_C)$ and $Y \in B(\mathcal{H}_A, \mathcal{H}_D)$ we define their tensor product as the operator $X \otimes Y \in B(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_D)$ acting as

$$(X \otimes Y)(|x\rangle \otimes |y\rangle) = X|x\rangle \otimes Y|y\rangle,$$

and extended linearly (using the universal property or the computational basis). Concretely, for $X \in B(\mathbb{C}^{d_A}, \mathbb{C}^{d_C})$ given by the matrix $X = [x_{ij}]_{ij}$ and $Y \in B(\mathbb{C}^{d_B}, \mathbb{C}^{d_D})$ we have

$$X \otimes Y = \begin{pmatrix} x_{11}Y & x_{12}Y & \cdots & x_{1d_A}Y \\ x_{21}Y & x_{22}Y & \cdots & x_{2d_A}Y \\ \vdots & & \ddots & \vdots \\ x_{d_C1}Y & x_{d_C2}Y & \cdots & x_{d_Cd_A}Y \end{pmatrix},$$

in the computational basis of the tensor product spaces introduced above. With this expression, it is easy to check

$$\begin{aligned} &\operatorname{Tr}[X \otimes Y] = \operatorname{Tr}[X] \operatorname{Tr}[Y], \\ &\sigma(X \otimes Y) = \{\lambda\mu \ : \ \lambda \in \sigma(X), \mu \in \sigma(Y)\}, \end{aligned}$$

for any $X \in B(\mathcal{H}_A)$ and $Y \in B(\mathcal{H}_B)$.

7 Matrix calculus

Whenever we talk about matrices without specifying a basis, we will implicitely assume that they are expressed in the basis of matrix units. Moreover, whenever we consider the Euclidean spaces $\mathbb{C}^d$ explicitely, we will use the terms operators and vectors interchangably with their representations in the computational basis and the matrix units. We can now express all objects introduced in the previous sections on matrices:

- **Kets:** Column vectors.
- **Bras:** Row vectors.
- **Baby-Riesz representation:** $|x\rangle = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \leftrightarrow (\overline{x_1}, \ldots, \overline{x_d}) = \langle x|$.
- **Euclidean inner product:** $\langle x|y\rangle = (\overline{x_1}, \ldots, \overline{x_d}) \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} = \sum_i \overline{x_i} y_i$.
- **Euclidean norm:** $\||x\rangle\|_2 = \sqrt{\sum_i |x_i|^2}$.
- **Adjoint operator:** $X \mapsto X^\dagger = \overline{X}^T$, where T denotes the transposition in the computational basis, and the conjugation is entry-wise:

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \mapsto \begin{pmatrix} \overline{x_{11}} & \overline{x_{21}} \\ \overline{x_{12}} & \overline{x_{22}} \end{pmatrix} = X^\dagger.$$

- **Hilbert-Schmidt inner product:** For $X, Y \in B(\mathbb{C}^{d_A}, \mathbb{C}^{d_B})$ we have

$$\langle X, Y\rangle_{HS} = \mathrm{Tr}\left[X^\dagger Y\right] = \sum_{ij} \overline{x_{ij}} y_{ij}.$$

- **Hilbert-Schmidt norm:** $\|X\|_{HS} = \sqrt{\sum_{ij} |x_{ij}|^2}$.
- **Unitaries and orthonormal bases:** There is a 1-to-1 correspondence between orthonormal bases $\{|x_1\rangle, \ldots, |x_d\rangle\}$ of $\mathbb{C}^d$ and unitary operators $U \in \mathcal{U}\left(\mathbb{C}^d\right)$ such that $U|i\rangle = |x_i\rangle$ for every $i \in \{1, \ldots, d\}$. This can written explicitely by noting that

$$U = \left(|x_1\rangle, \ldots, |x_d\rangle\right),$$

  i.e., the vectors $|x_i\rangle$ appear as the columns of U, defines a unitary matrix.

- **Spectral decomposition:** For any normal operator $X \in B(\mathbb{C}^d)$ there exists a unitary $U \in \mathcal{U}\left(\mathbb{C}^d\right)$ such that

$$UXU^\dagger = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix},$$

  where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of X.

- **Singular value decomposition:** For any operator $X \in B(\mathbb{C}^{d_A}, \mathbb{C}^{d_B})$ there exist unitaries $U \in \mathcal{U}\left(\mathbb{C}^{d_A}\right)$ and $V \in \mathcal{U}\left(\mathbb{C}^{d_B}\right)$ such that

$$UXV^\dagger = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix},$$

where

$$S = \begin{pmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_d \end{pmatrix},$$

is the diagonal matrix containing the singular values on the diagonal.

- **Tensor product:** $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B} \simeq \mathbb{C}^{d_A d_B}$ and

$$|x\rangle \otimes |y\rangle = \begin{pmatrix} x_1|y\rangle \\ x_2|y\rangle \\ \vdots \\ x_{d_A}|y\rangle \end{pmatrix}.$$

- **Tensor product of operators:** For $X \in B(\mathbb{C}^{d_A}, \mathbb{C}^{d_C})$ and $Y \in B(\mathbb{C}^{d_B}, \mathbb{C}^{d_D})$ we have

$$X \otimes Y = \begin{pmatrix} x_{11}Y & x_{12}Y & \cdots & x_{1d_A}Y \\ x_{21}Y & x_{22}Y & \cdots & x_{2d_A}Y \\ \vdots & & \ddots & \vdots \\ x_{d_C1}Y & x_{d_C2}Y & \cdots & x_{d_Cd_A}Y \end{pmatrix}.$$

  This definition is sometimes called the *Kronecker product*.

- **Tensor-operators as block matrices:** Any $X \in B(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$ can be written as

$$X = \sum_{ij} |i\rangle\langle j| \otimes X_{ij} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1d_A} \\ X_{21} & X_{22} & \cdots & X_{2d_A} \\ \vdots & & \ddots & \vdots \\ X_{d_A1} & X_{d_A2} & \cdots & X_{d_Ad_A} \end{pmatrix},$$

  with operators $X_{ij} \in B\left(\mathbb{C}^{d_B}\right)$.

It might be helpful to keep these concrete representations in mind, and to use them to get a better feeling for the objects introduced in quantum information theory. At some point, it should come natural to you to do explicit computations, which is easiest when using the concrete representations introduced above.