CS 766/QIC 820 Theory of Quantum Information (Fall 2011)

Lecture 2: Mathematical preliminaries (part 2)

This lecture represents the second half of the discussion that we started in the previous lecture concerning basic mathematical concepts and tools used throughout the course.

2.1 The singular-value theorem

The spectral theorem, discussed in the previous lecture, is a valuable tool in quantum information theory. The fact that it is limited to normal operators can, however, restrict its applicability.

The *singular value theorem*, which we will now discuss, is closely related to the spectral theorem, but holds for arbitrary operators—even those of the form $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ for different spaces $\mathcal{X}$ and $\mathcal{Y}$. Like the spectral theorem, we will find that the singular value decomposition is an indispensable tool in quantum information theory. Let us begin with a statement of the theorem.

Theorem 2.1 (Singular Value Theorem). *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be a nonzero operator, and let $r = \mathrm{rank}(A)$. There exist positive real numbers $s_1, \ldots, s_r$ and orthonormal sets $\{x_1, \ldots, x_r\} \subset \mathcal{X}$ and $\{y_1, \ldots, y_r\} \subset \mathcal{Y}$ such that*

$$A = \sum_{j=1}^{r} s_j y_j x_j^*. \tag{2.1}$$

An expression of a given matrix A in the form of (2.1) is said to be a *singular value decomposition* of A. The numbers $s_1, \ldots, s_r$ are called *singular values* and the vectors $x_1, \ldots, x_r$ and $y_1, \ldots, y_r$ are called *right* and *left singular vectors*, respectively.

The singular values $s_1, \ldots, s_r$ of an operator A are uniquely determined, up to their ordering. Hereafter we will assume, without loss of generality, that the singular values are ordered from largest to smallest: $s_1 \geq \cdots \geq s_r$. When it is necessary to indicate the dependence of these singular values on A, we denote them $s_1(A), \ldots, s_r(A)$. Although technically speaking 0 is not usually considered a singular value of any operator, it will be convenient to also define $s_k(A) = 0$ for $k > \mathrm{rank}(A)$. The notation $s(A)$ is used to refer to the vector of singular values $s(A) = (s_1(A), \ldots, s_r(A))$, or to an extension of this vector $s(A) = (s_1(A), \ldots, s_k(A))$ for $k > r$ when it is convenient to view it as an element of $\mathbb{R}^k$ for $k > \mathrm{rank}(A)$.

There is a close relationship between singular value decompositions of an operator A and spectral decompositions of the operators A^*A and AA^*. In particular, it will necessarily hold that

$$s_k(A) = \sqrt{\lambda_k(AA^*)} = \sqrt{\lambda_k(A^*A)} \tag{2.2}$$

for $1 \leq k \leq \mathrm{rank}(A)$, and moreover the right singular vectors of A will be eigenvectors of A^*A and the left singular vectors of A will be eigenvectors of AA^*. One is free, in fact, to choose the left singular vectors of A to be any orthonormal collection of eigenvectors of AA^* for which the corresponding eigenvalues are nonzero—and once this is done the right singular vectors will be uniquely determined. Alternately, the right singular vectors of A may be chosen to be

any orthonormal collection of eigenvectors of A^*A for which the corresponding eigenvalues are nonzero, which uniquely determines the left singular vectors.

In the case that $\mathcal{Y} = \mathcal{X}$ and A is a normal operator, it is essentially trivial to derive a singular value decomposition from a spectral decomposition. In particular, suppose that

$$A = \sum_{j=1}^{n} \lambda_j x_j x_j^*$$

is a spectral decomposition of A, and assume that we have chosen to label the eigenvalues of A in such a way that $\lambda_j \neq 0$ for $1 \leq j \leq r = \operatorname{rank}(A)$. A singular value decomposition of the form (2.1) is obtained by setting

$$s_j = |\lambda_j| \qquad \text{and} \qquad y_j = \frac{\lambda_j}{|\lambda_j|} x_j$$

for $1 \leq j \leq r$. Note that this shows, for normal operators, that the singular values are simply the absolute values of the nonzero eigenvalues.

2.1.1 The Moore-Penrose pseudo-inverse

Later in the course we will occasionally refer to the *Moore-Penrose pseudo-inverse* of an operator, which is closely related to its singular value decompositions. For any given operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, we define the Moore-Penrose pseudo-inverse of A, denoted $A^+ \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$, as the unique operator satisfying these properties:

1. $AA^+A = A$,
2. $A^+AA^+ = A^+$, and
3. AA^+ and A^+A are both Hermitian.

It is clear that there is at least one such choice of A^+, for if

$$A = \sum_{j=1}^{r} s_j y_j x_j^*$$

is a singular value decomposition of A, then

$$A^+ = \sum_{j=1}^{r} \frac{1}{s_j} x_j y_j^*$$

satisfies the three properties above.

The fact that A^+ is uniquely determined by the above equations is easily verified, for suppose that $X, Y \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ both satisfy the above properties:

1. $AXA = A = AYA$,
2. $XAX = X$ and $YAY = Y$, and
3. AX, XA, AY, and YA are all Hermitian.

Using these properties, we observe that

$$\begin{aligned} X &= XAX = (XA)^*X = A^*X^*X = (AYA)^*X^*X = A^*Y^*A^*X^*X \\ &= (YA)^*(XA)^*X = YAXAX = YAX = YAYAX = Y(AY)^*(AX)^* \\ &= YY^*A^*X^*A^* = YY^*(AXA)^* = YY^*A^* = Y(AY)^* = YAY = Y, \end{aligned}$$

which shows that $X = Y$.

2.2 Linear mappings on operator algebras

Linear mappings of the form

$$\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y}),$$

where $\mathcal{X}$ and $\mathcal{Y}$ are complex Euclidean spaces, play an important role in the theory of quantum information. The set of all such mappings is sometimes denoted $\mathrm{T}(\mathcal{X}, \mathcal{Y})$, or $\mathrm{T}(\mathcal{X})$ when $\mathcal{X} = \mathcal{Y}$, and is itself a linear space when addition of mappings and scalar multiplication are defined in the straightforward way:

1. Addition: given $\Phi, \Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, the mapping $\Phi + \Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is defined by

   $$(\Phi + \Psi)(A) = \Phi(A) + \Psi(A)$$

   for all $A \in \mathrm{L}(\mathcal{X})$.
2. Scalar multiplication: given $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ and $\alpha \in \mathbb{C}$, the mapping $\alpha\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is defined by

   $$(\alpha\Phi)(A) = \alpha(\Phi(A))$$

   for all $A \in \mathrm{L}(\mathcal{X})$.

For a given mapping $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, the *adjoint* of Φ is defined to be the unique mapping $\Phi^* \in \mathrm{T}(\mathcal{Y}, \mathcal{X})$ that satisfies

$$\langle \Phi^*(B), A \rangle = \langle B, \Phi(A) \rangle$$

for all $A \in \mathrm{L}(\mathcal{X})$ and $B \in \mathrm{L}(\mathcal{Y})$.

The transpose

$$T : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{X}) : A \mapsto A^{\mathsf{T}}$$

is a simple example of a mapping of this type, as is the trace

$$\mathrm{Tr} : \mathrm{L}(\mathcal{X}) \to \mathbb{C} : A \mapsto \mathrm{Tr}(A),$$

provided we make the identification $\mathrm{L}(\mathbb{C}) = \mathbb{C}$.

2.2.1 Remark on tensor products of operators and mappings

Tensor products of operators can be defined in concrete terms using the same sort of Kronecker product construction that we considered for vectors, as well as in more abstract terms connected with the notion of multilinear functions. We will briefly discuss these definitions now, as well as their extension to tensor products of mappings on operator algebras.

First, suppose $A_1 \in \mathrm{L}(\mathcal{X}_1, \mathcal{Y}_1), \ldots, A_n \in \mathrm{L}(\mathcal{X}_n, \mathcal{Y}_n)$ are operators, for complex Euclidean spaces

$$\mathcal{X}_1 = \mathbb{C}^{\Sigma_1}, \ldots, \mathcal{X}_n = \mathbb{C}^{\Sigma_n} \quad \text{and} \quad \mathcal{Y}_1 = \mathbb{C}^{\Gamma_1}, \ldots, \mathcal{Y}_n = \mathbb{C}^{\Gamma_n}.$$

We define a new operator

$$A_1 \otimes \cdots \otimes A_n \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n),$$

in terms of its matrix representation, as

$$(A_1 \otimes \cdots \otimes A_n)((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = A_1(a_1, b_1) \cdots A_n(a_n, b_n)$$

(for all $a_1 \in \Gamma_1, \ldots, a_n \in \Gamma_n$ and $b_1 \in \Sigma_1, \ldots, b_n \in \Gamma_n$).

It is not difficult to check that the operator $A_1 \otimes \cdots \otimes A_n$ just defined satisfies the equation

$$(A_1 \otimes \cdots \otimes A_n)(u_1 \otimes \cdots \otimes u_n) = (A_1 u_1) \otimes \cdots \otimes (A_n u_n) \tag{2.3}$$

for all choices of $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$. Given that $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ is spanned by the set of all elementary tensors $u_1 \otimes \cdots \otimes u_n$, it is clear that $A_1 \otimes \cdots \otimes A_n$ is the only operator that can satisfy this equation (again, for all choices of $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$). We could, therefore, have considered the equation (2.3) to have been the defining property of $A_1 \otimes \cdots \otimes A_n$.

When considering operator spaces as vector spaces, similar identities to the ones in the previous lecture for tensor products of vectors become apparent. For example,

$$\begin{aligned} A_1 \otimes \cdots \otimes A_{k-1} \otimes (A_k + B_k) \otimes A_{k+1} \otimes \cdots \otimes A_n & \\ = A_1 \otimes \cdots \otimes A_{k-1} \otimes A_k \otimes A_{k+1} \otimes \cdots \otimes A_n & \\ + A_1 \otimes \cdots \otimes A_{k-1} \otimes B_k \otimes A_{k+1} \otimes \cdots \otimes A_n. & \end{aligned}$$

In addition, for all choices of complex Euclidean spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$, $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, and $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$, and all operators $A_1 \in \mathrm{L}(\mathcal{X}_1, \mathcal{Y}_1), \ldots, A_n \in \mathrm{L}(\mathcal{X}_n, \mathcal{Y}_n)$ and $B_1 \in \mathrm{L}(\mathcal{Y}_1, \mathcal{Z}_1), \ldots, B_n \in \mathrm{L}(\mathcal{Y}_n, \mathcal{Z}_n)$, it holds that

$$(B_1 \otimes \cdots \otimes B_n)(A_1 \otimes \cdots \otimes A_n) = (B_1 A_1) \otimes \cdots \otimes (B_n A_n).$$

Also note that spectral and singular value decompositions of tensor products of operators are very easily obtained from those of the individual operators. This allows one to quickly conclude that

$$\|A_1 \otimes \cdots \otimes A_n\|_p = \|A_1\|_p \cdots \|A_n\|_p,$$

along with a variety of other facts that may be derived by similar reasoning.

Tensor products of linear mappings on operator algebras may be defined in a similar way to those of operators. At this point we have not yet considered concrete representations of such mappings, to which a Kronecker product construction could be applied, but we will later discuss such representations. For now let us simply define the linear mapping

$$\Phi_1 \otimes \cdots \otimes \Phi_n : \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \to \mathrm{L}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n),$$

for any choice of linear mappings $\Phi_1 : \mathrm{L}(\mathcal{X}_1) \to \mathrm{L}(\mathcal{Y}_1), \ldots, \Phi_n : \mathrm{L}(\mathcal{X}_n) \to \mathrm{L}(\mathcal{Y}_n)$, to be the unique mapping that satisfies the equation

$$(\Phi_1 \otimes \cdots \otimes \Phi_n)(A_1 \otimes \cdots \otimes A_n) = \Phi_1(A_1) \otimes \cdots \otimes \Phi_n(A_n)$$

for all operators $A_1 \in \mathrm{L}(\mathcal{X}_1), \ldots, A_n \in \mathrm{L}(\mathcal{X}_n)$.

Example 2.2 (The partial trace)**.** Let $\mathcal{X}$ be a complex Euclidean space. As mentioned above, we may view the trace as taking the form $\mathrm{Tr} : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathbb{C})$ by making the identification $\mathbb{C} = \mathrm{L}(\mathbb{C})$. For a second complex Euclidean space $\mathcal{Y}$, we may therefore consider the mapping

$$\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} : \mathrm{L}(\mathcal{X} \otimes \mathcal{Y}) \to \mathrm{L}(\mathcal{Y}).$$

This is the unique mapping that satisfies

$$(\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(A \otimes B) = \mathrm{Tr}(A) B$$

for all $A \in \mathrm{L}(\mathcal{X})$ and $B \in \mathrm{L}(\mathcal{Y})$. This mapping is called the *partial trace*, and is more commonly denoted $\mathrm{Tr}_{\mathcal{X}}$. In general, the subscript refers to the space to which the trace is applied, while the space or spaces that remains ($\mathcal{Y}$ in the case above) are implicit from the context in which the mapping is used.

One may alternately express the partial trace on $\mathcal{X}$ as follows, assuming that $\{x_a : a \in \Sigma\}$ is any orthonormal basis of $\mathcal{X}$:

$$\mathrm{Tr}_{\mathcal{X}}(A) = \sum_{a \in \Sigma} (x_a^* \otimes \mathbb{1}_{\mathcal{Y}}) A (x_a \otimes \mathbb{1}_{\mathcal{Y}})$$

for all $A \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$. An analogous expression holds for $\mathrm{Tr}_{\mathcal{Y}}$.

2.3 Norms of operators

The next topic for this lecture concerns norms of operators. As is true more generally, a *norm* on the space of operators $\mathrm{L}(\mathcal{X}, \mathcal{Y})$, for any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, is a function $\|\cdot\|$ satisfying the following properties:

1. Positive definiteness: $\|A\| \geq 0$ for all $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, with $\|A\| = 0$ if and only if $A = 0$.
2. Positive scalability: $\|\alpha A\| = |\alpha| \, \|A\|$ for all $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $\alpha \in \mathbb{C}$.
3. The triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$ for all $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$.

Many interesting and useful norms can be defined on spaces of operators, but in this course we will mostly be concerned with a single family of norms called *Schatten p-norms*. This family includes the three most commonly used norms in quantum information theory: the *spectral norm*, the *Frobenius norm*, and the *trace norm*.

2.3.1 Definition and basic properties of Schatten norms

For any operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and any real number $p \geq 1$, one defines the Schatten p-norm of A as

$$\|A\|_p = \left[\mathrm{Tr}\left((A^* A)^{p/2}\right)\right]^{1/p}.$$

We also define

$$\|A\|_\infty = \max\{\|Au\| : u \in \mathcal{X}, \ \|u\| = 1\}, \tag{2.4}$$

which happens to coincide with $\lim_{p \to \infty} \|A\|_p$ and therefore explains why the subscript ∞ is used.

An equivalent way to define these norms is to consider the the vector $s(A)$ of singular values of A, as discussed at the beginning of the lecture. For each $p \in [1, \infty]$, it holds that the Schatten p-norm of A coincides with the ordinary (vector) p-norm of $s(A)$:

$$\|A\|_p = \|s(A)\|_p.$$

The Schatten p-norms satisfy many nice properties, some of which are summarized in the following list:

1. The Schatten p-norms are non-increasing in p. In other words, for any operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and for $1 \leq p \leq q \leq \infty$ we have
$$\|A\|_p \geq \|A\|_q.$$

2. For every $p \in [1,\infty]$, the Schatten p-norm is isometrically invariant (and therefore unitarily invariant). This means that
$$\|A\|_p = \|UAV^*\|_p$$
for any choice of linear isometries U and V (which include unitary operators U and V) for which the product UAV^* makes sense.
3. For each $p \in [1,\infty]$, one defines $p^* \in [1,\infty]$ by the equation
$$\frac{1}{p} + \frac{1}{p^*} = 1.$$
For every operator $A \in \mathrm{L}(\mathcal{X},\mathcal{Y})$, it holds that
$$\|A\|_p = \max\left\{ |\langle B, A\rangle| \,:\, B \in \mathrm{L}(\mathcal{X},\mathcal{Y}),\ \|B\|_{p^*} \leq 1 \right\}.$$
This implies that
$$|\langle B, A\rangle| \leq \|A\|_p \|B\|_{p^*},$$
which is *Hölder's inequality* for Schatten norms.
4. For any choice of linear operators $A \in \mathrm{L}(\mathcal{X}_1,\mathcal{X}_2)$, $B \in \mathrm{L}(\mathcal{X}_2,\mathcal{X}_3)$, and $C \in \mathrm{L}(\mathcal{X}_3,\mathcal{X}_4)$, and any choice of $p \in [1,\infty]$, we have
$$\|CBA\|_p \leq \|C\|_\infty \|B\|_p \|A\|_\infty.$$
It follows that
$$\|AB\|_p \leq \|A\|_p \|B\|_p \tag{2.5}$$
for all choices of $p \in [1,\infty]$ and operators A and B for which the product AB exists. The property (2.5) is known as *submultiplicativity*.
5. It holds that
$$\|A\|_p = \|A^*\|_p = \|A^{\mathsf{T}}\|_p = \|\overline{A}\|_p$$
for every $A \in \mathrm{L}(\mathcal{X},\mathcal{Y})$.

2.3.2 The trace norm, Frobenius norm, and spectral norm

The Schatten 1-norm is more commonly called the *trace norm*, the Schatten 2-norm is also known as the *Frobenius norm*, and the Schatten ∞-norm is called the *spectral norm* or *operator norm*. A common notation for these norms is:
$$\|\cdot\|_{\mathrm{tr}} = \|\cdot\|_1, \quad \|\cdot\|_{\mathrm{F}} = \|\cdot\|_2, \quad \text{and} \quad \|\cdot\| = \|\cdot\|_\infty.$$
In this course we will generally write $\|\cdot\|$ rather than $\|\cdot\|_\infty$, but will not use the notation $\|\cdot\|_{\mathrm{tr}}$ and $\|\cdot\|_{\mathrm{F}}$.

Let us note a few special properties of these three particular norms:

1. *The spectral norm.* The spectral norm $\|\cdot\| = \|\cdot\|_\infty$, also called the *operator norm*, is special in several respects. It is the norm *induced* by the Euclidean norm, which is its defining property (2.4). It satisfies the property
$$\|A^*A\| = \|A\|^2$$
for every $A \in \mathrm{L}(\mathcal{X},\mathcal{Y})$.

2. *The Frobenius norm.* Substituting $p = 2$ into the definition of $\|\cdot\|_p$ we see that the Frobenius norm $\|\cdot\|_2$ is given by

$$\|A\|_2 = [\mathrm{Tr}\,(A^*A)]^{1/2} = \sqrt{\langle A, A\rangle}.$$

It is therefore the norm defined by the inner product on $\mathrm{L}\,(\mathcal{X}, \mathcal{Y})$. In essence, it is the norm that one obtains by thinking of elements of $\mathrm{L}\,(\mathcal{X}, \mathcal{Y})$ as ordinary vectors and forgetting that they are operators:

$$\|A\|_2 = \sqrt{\sum_{a,b} |A(a,b)|^2},$$

where a and b range over the indices of the matrix representation of A.

3. *The trace norm.* Substituting $p = 1$ into the definition of $\|\cdot\|_p$ we see that the trace norm $\|\cdot\|_1$ is given by

$$\|A\|_1 = \mathrm{Tr}\left(\sqrt{A^*A}\right).$$

A convenient expression of $\|A\|_1$, for any operator of the form $A \in \mathrm{L}\,(\mathcal{X})$, is

$$\|A\|_1 = \max\{|\langle A, U\rangle| \,:\, U \in \mathrm{U}\,(\mathcal{X})\}.$$

Another useful fact about the trace norm is that it is *monotonic*:

$$\|\mathrm{Tr}_{\mathcal{Y}}(A)\|_1 \leq \|A\|_1$$

for all $A \in \mathrm{L}\,(\mathcal{X} \otimes \mathcal{Y})$. This is because

$$\|\mathrm{Tr}_{\mathcal{Y}}(A)\|_1 = \max\left\{|\langle A, U \otimes \mathbb{1}_{\mathcal{Y}}\rangle| \,:\, U \in \mathrm{U}\,(\mathcal{X})\right\}$$

while

$$\|A\|_1 = \max\left\{|\langle A, U\rangle| \,:\, U \in \mathrm{U}\,(\mathcal{X} \otimes \mathcal{Y})\right\};$$

and the inequality follows from the fact that the first maximum is taken over a subset of the unitary operators for the second.

Example 2.3. Consider a complex Euclidean space $\mathcal{X}$ and any choice of unit vectors $u, v \in \mathcal{X}$. We have

$$\|uu^* - vv^*\|_p = 2^{1/p}\sqrt{1 - |\langle u, v\rangle|^2}. \tag{2.6}$$

To see this, we note that the operator $A = uu^* - vv^*$ is Hermitian and therefore normal, so its singular values are the absolute values of its nonzero eigenvalues. It will therefore suffice to prove that the eigenvalues of A are $\pm\sqrt{1 - |\langle u, v\rangle|^2}$, along with the eigenvalue 0 occurring with multiplicity $n - 2$, where $n = \dim(\mathcal{X})$. Given that $\mathrm{Tr}(A) = 0$ and $\mathrm{rank}(A) \leq 2$, it is evident that the eigenvalues of A are of the form $\pm\lambda$ for some $\lambda \geq 0$, along with eigenvalue 0 with multiplicity $n - 2$. As

$$2\lambda^2 = \mathrm{Tr}(A^2) = 2 - 2|\langle u, v\rangle|^2$$

we conclude $\lambda = \sqrt{1 - |\langle u, v\rangle|^2}$, from which (2.6) follows:

$$\|uu^* - vv^*\|_p = \left(2\left(1 - |\langle u, v\rangle|^2\right)^{p/2}\right)^{1/p} = 2^{1/p}\sqrt{1 - |\langle u, v\rangle|^2}.$$

2.4 The operator-vector correspondence

It will be helpful throughout this course to make use of a simple correspondence between the spaces $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $\mathcal{Y} \otimes \mathcal{X}$, for given complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$.

We define the mapping

$$\mathrm{vec} : \mathrm{L}(\mathcal{X}, \mathcal{Y}) \to \mathcal{Y} \otimes \mathcal{X}$$

to be the linear mapping that represents a change of bases from the standard basis of $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ to the standard basis of $\mathcal{Y} \otimes \mathcal{X}$. Specifically, we define

$$\mathrm{vec}(E_{b,a}) = e_b \otimes e_a$$

for all $a \in \Sigma$ and $b \in \Gamma$, at which point the mapping is determined for every $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ by linearity. In the Dirac notation, this mapping amounts to flipping a bra to a ket:

$$\mathrm{vec}(|b\rangle \langle a|) = |b\rangle |a\rangle .$$

(Note that it is only standard basis elements that are flipped in this way.)

The vec mapping is a linear bijection, which implies that every vector $u \in \mathcal{Y} \otimes \mathcal{X}$ uniquely determines an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ that satisfies $\mathrm{vec}(A) = u$. It is also an isometry, in the sense that

$$\langle A, B \rangle = \langle \mathrm{vec}(A), \mathrm{vec}(B) \rangle$$

for all $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$. The following properties of the vec mapping are easily verified:

1. For every choice of complex Euclidean spaces $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_1$, and $\mathcal{Y}_2$, and every choice of operators $A \in \mathrm{L}(\mathcal{X}_1, \mathcal{Y}_1)$, $B \in \mathrm{L}(\mathcal{X}_2, \mathcal{Y}_2)$, and $X \in \mathrm{L}(\mathcal{X}_2, \mathcal{X}_1)$, it holds that

$$(A \otimes B)\,\mathrm{vec}(X) = \mathrm{vec}(AXB^{\mathsf{T}}). \tag{2.7}$$

2. For every choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and every choice of operators $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, the following equations hold:

$$\mathrm{Tr}_{\mathcal{X}}\left(\mathrm{vec}(A)\,\mathrm{vec}(B)^*\right) = AB^*, \tag{2.8}$$

$$\mathrm{Tr}_{\mathcal{Y}}\left(\mathrm{vec}(A)\,\mathrm{vec}(B)^*\right) = \left(B^*A\right)^{\mathsf{T}}. \tag{2.9}$$

3. For $u \in \mathcal{X}$ and $v \in \mathcal{Y}$ we have

$$\mathrm{vec}(uv^*) = u \otimes \overline{v}. \tag{2.10}$$

   This includes the special cases $\mathrm{vec}(u) = u$ and $\mathrm{vec}(v^*) = \overline{v}$, which we obtain by setting $v = 1$ and $u = 1$, respectively.

Example 2.4 (The Schmidt decomposition)**.** Suppose $u \in \mathcal{Y} \otimes \mathcal{X}$ for given complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$. Let $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be the unique operator for which $u = \mathrm{vec}(A)$. There exists a singular value decomposition

$$A = \sum_{i=1}^{r} s_i y_i x_i^*$$

of A. Consequently

$$u = \mathrm{vec}(A) = \mathrm{vec}\left(\sum_{i=1}^{r} s_i y_i x_i^*\right) = \sum_{i=1}^{r} s_i\,\mathrm{vec}\left(y_i x_i^*\right) = \sum_{i=1}^{r} s_i y_i \otimes \overline{x_i}.$$

The fact that $\{x_1, \ldots, x_r\}$ is orthonormal implies that $\{\overline{x_1}, \ldots, \overline{x_r}\}$ is orthonormal as well.

We have therefore established the validity of the *Schmidt decomposition*, which states that every vector $u \in \mathcal{Y} \otimes \mathcal{X}$ can be expressed in the form

$$u = \sum_{i=1}^{r} s_i y_i \otimes z_i$$

for positive real numbers $s_1, \ldots, s_r$ and orthonormal sets $\{y_1, \ldots, y_r\} \subset \mathcal{Y}$ and $\{z_1, \ldots, z_r\} \subset \mathcal{X}$.

2.5 Analysis

Mathematical analysis is concerned with notions of limits, continuity, differentiation, integration and measure, and so on. As some of the proofs that we will encounter in the course will require arguments based on these notions, it is appropriate to briefly review some of the necessary concepts here.

It will be sufficient for our needs that this summary is narrowly focused on Euclidean spaces (as opposed to infinite dimensional spaces). As a result, these notes do not treat analytic concepts in the sort of generality that would be typical of a standard analysis book or course. If you are interested in such a book, the following one is considered a classic:

- W. Rudin. *Principles of Mathematical Analysis*. McGraw–Hill, 1964.

2.5.1 Basic notions of analysis

Let $\mathcal{V}$ be a real or complex Euclidean space, and (for this section only) let us allow $\|\cdot\|$ to be any choice of a fixed norm on $\mathcal{V}$. We may take $\|\cdot\|$ to be the Euclidean norm, but nothing changes if we choose a different norm. (The validity of this assumption rests on the fact that Euclidean spaces are finite-dimensional.)

The *open ball* of radius r around a vector $u \in \mathcal{X}$ is defined as

$$\mathcal{B}_r(u) = \{v \in \mathcal{X} : \|u - v\| < r\},$$

and the *sphere* of radius r around u is defined as

$$\mathcal{S}_r(u) = \{v \in \mathcal{X} : \|u - v\| = r\}.$$

The *closed ball* of radius r around u is the union $\mathcal{B}_r(u) \cup \mathcal{S}_r(u)$.

A set $\mathcal{A} \subseteq \mathcal{X}$ is *open* if, for every $u \in \mathcal{A}$ there exists a choice of $\epsilon > 0$ such that $\mathcal{B}_\epsilon(u) \subseteq \mathcal{A}$. Equivalently, $\mathcal{A} \subseteq \mathcal{X}$ is open if it is the union of some collection of open balls. (This can be an empty, finite, or countably or uncountably infinite collections of open balls.) A set $\mathcal{A} \subseteq \mathcal{X}$ is *closed* if it is the complement of an open set.

Given subsets $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{X}$, we say that $\mathcal{B}$ is open or closed *relative to* $\mathcal{A}$ if $\mathcal{B}$ is the intersection of $\mathcal{A}$ and some open or closed set in $\mathcal{X}$, respectively.

Let $\mathcal{A}$ and $\mathcal{B}$ be subsets of a Euclidean space $\mathcal{X}$ that satisfy $\mathcal{B} \subseteq \mathcal{A}$. Then the *closure* of $\mathcal{B}$ relative to $\mathcal{A}$ is the intersection of all subsets $\mathcal{C}$ for which $\mathcal{B} \subseteq \mathcal{C}$ and $\mathcal{C}$ is closed relative to $\mathcal{A}$. In other words, this is the smallest set that contains $\mathcal{B}$ and is closed relative to $\mathcal{A}$. The set $\mathcal{B}$ is *dense* in $\mathcal{A}$ if the closure of $\mathcal{B}$ relative to $\mathcal{A}$ is $\mathcal{A}$ itself.

Suppose $\mathcal{X}$ and $\mathcal{Y}$ are Euclidean spaces and $f : \mathcal{A} \to \mathcal{Y}$ is a function defined on some subset $\mathcal{A} \subseteq \mathcal{X}$. For any point $u \in \mathcal{A}$, the function f is said to be *continuous* at u if the following holds: for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|f(v) - f(u)\| < \varepsilon$$

for all $v \in \mathcal{B}_\delta(u) \cap \mathcal{A}$. An alternate way of writing this condition is

$$(\forall \varepsilon > 0)(\exists \delta > 0)[f(\mathcal{B}_\delta(u) \cap \mathcal{A}) \subseteq \mathcal{B}_\varepsilon(f(u))].$$

If f is continuous at every point in $\mathcal{A}$, then we just say that f is *continuous on* $\mathcal{A}$.

The *preimage* of a set $\mathcal{B} \subseteq \mathcal{Y}$ under a function $f : \mathcal{A} \to \mathcal{Y}$ defined on some subset $\mathcal{A} \subseteq \mathcal{X}$ is defined as

$$f^{-1}(\mathcal{B}) = \{u \in \mathcal{A} \,:\, f(u) \in \mathcal{B}\}.$$

Such a function f is continuous on $\mathcal{A}$ if and only if the preimage of every open set in $\mathcal{Y}$ is open relative to $\mathcal{A}$. Equivalently, f is continuous on $\mathcal{A}$ if and only if the preimage of every closed set in $\mathcal{Y}$ is closed relative to $\mathcal{A}$.

A *sequence* of vectors in a subset $\mathcal{A}$ of a Euclidean space $\mathcal{X}$ is a function $s : \mathbb{N} \to \mathcal{A}$, where $\mathbb{N}$ denotes the set of natural numbers $\{1, 2, \ldots\}$. We usually denote a general sequence by $(u_n)_{n\in\mathbb{N}}$ or (u_n), and it is understood that the function s in question is given by $s : n \mapsto u_n$. A sequence $(u_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ *converges* to $u \in \mathcal{X}$ if, for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|u_n - u\| < \varepsilon$ for all $n \geq N$.

A sequence (v_n) is a *sub-sequence* of (u_n) if there is a strictly increasing sequence of nonnegative integers $(k_n)_{n\in\mathbb{N}}$ such that $v_n = u_{k_n}$ for all $n \in \mathbb{N}$. In other words, you get a sub-sequence from a sequence by skipping over whichever vectors you want, provided that you still have infinitely many vectors left.

2.5.2 Compact sets

A set $\mathcal{A} \subseteq \mathcal{X}$ is *compact* if every sequence (u_n) in $\mathcal{A}$ has a sub-sequence (v_n) that converges to a point $v \in \mathcal{A}$. In any Euclidean space $\mathcal{X}$, a set $\mathcal{A}$ is compact if and only if it is closed and bounded (which means it is contained in $\mathcal{B}_r(0)$ for some real number $r > 0$). This fact is known as the *Heine-Borel Theorem*.

Compact sets have some nice properties. Two properties that are noteworthy for the purposes of this course are following:

1. If $\mathcal{A}$ is compact and $f : \mathcal{A} \to \mathbb{R}$ is continuous on $\mathcal{A}$, then f achieves both a maximum and minimum value on $\mathcal{A}$.
2. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. If $\mathcal{A} \subset \mathcal{X}$ is compact and $f : \mathcal{X} \to \mathcal{Y}$ is continuous on $\mathcal{A}$, then $f(\mathcal{A}) \subset \mathcal{Y}$ is also compact.

2.6 Convexity

Many sets of interest in the theory of quantum information are *convex sets*, and when reasoning about some of these sets we will make use of various facts from the the theory of convexity (or convex analysis). Two books on convexity theory that you may find helpful are these:

- R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- A. Barvinok. *A Course in Convexity*. Volume 54 of *Graduate Studies in Mathematics*, American Mathematical Society, 2002.

2.6.1 Basic notions of convexity

Let $\mathcal{X}$ be any Euclidean space. A set $\mathcal{A} \subseteq \mathcal{X}$ is *convex* if, for all choices of $u, v \in \mathcal{A}$ and $\lambda \in [0,1]$, we have

$$\lambda u + (1-\lambda)v \in \mathcal{A}.$$

Another way to say this is that $\mathcal{A}$ is convex if and only if you can always draw the straight line between any two points of $\mathcal{A}$ without going outside $\mathcal{A}$.

A point $w \in \mathcal{A}$ in a convex set $\mathcal{A}$ is said to be an *extreme point* of $\mathcal{A}$ if, for every expression

$$w = \lambda u + (1-\lambda)v$$

for $u, v \in \mathcal{A}$ and $\lambda \in (0,1)$, it holds that $u = v = w$. These are the points that do not lie *properly* between two other points in $\mathcal{A}$.

A set $\mathcal{A} \subseteq \mathcal{X}$ is a *cone* if, for all choices of $u \in \mathcal{A}$ and $\lambda \geq 0$, we have that $\lambda u \in \mathcal{A}$. A *convex cone* is simply a cone that is also convex. A cone $\mathcal{A}$ is convex if and only if, for all $u, v \in \mathcal{A}$, it holds that $u + v \in \mathcal{A}$.

The intersection of any collection of convex sets is also convex. Also, if $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}$ are convex, then their sum and difference

$$\mathcal{A} + \mathcal{B} = \{u + v \,:\, u \in \mathcal{A},\, v \in \mathcal{B}\} \quad \text{and} \quad \mathcal{A} - \mathcal{B} = \{u - v \,:\, u \in \mathcal{A},\, v \in \mathcal{B}\}$$

are also convex.

Example 2.5. For any $\mathcal{X}$, the set $\mathrm{Pos}(\mathcal{X})$ is a convex cone. This is so because it follows easily from the definition of positive semidefinite operators that $\mathrm{Pos}(\mathcal{X})$ is a cone and $A + B \in \mathrm{Pos}(\mathcal{X})$ for all $A, B \in \mathrm{Pos}(\mathcal{X})$. The only extreme point of this set is $0 \in \mathrm{L}(\mathcal{X})$. The set $\mathrm{D}(\mathcal{X})$ of density operators on $\mathcal{X}$ is convex, but it is not a cone. Its extreme points are precisely those density operators having rank 1, i.e., those of the form uu^* for $u \in \mathcal{X}$ being a unit vector.

For any finite, nonempty set Σ, we say that a vector $p \in \mathbb{R}^\Sigma$ is a *probability vector* if it holds that $p(a) \geq 0$ for all $a \in \Sigma$ and $\sum_{a \in \Sigma} p(a) = 1$. A *convex combination* of points in $\mathcal{A}$ is any finite sum of the form

$$\sum_{a \in \Sigma} p(a) u_a$$

for $\{u_a \,:\, a \in \Sigma\} \subset A$ and $p \in \mathbb{R}^\Sigma$ a probability vector. Notice that we are speaking only of *finite* sums when we refer to convex combinations.

The *convex hull* of a set $\mathcal{A} \subseteq \mathcal{X}$, denoted $\mathrm{conv}(\mathcal{A})$, is the intersection of all convex sets containing $\mathcal{A}$. Equivalently, it is precisely the set of points that can be written as convex combinations of points in $\mathcal{A}$. This is true even in the case that $\mathcal{A}$ is infinite. The convex hull $\mathrm{conv}(\mathcal{A})$ of a closed set $\mathcal{A}$ need not itself be closed. However, if $\mathcal{A}$ is compact, then so too is $\mathrm{conv}(\mathcal{A})$. The *Krein-Milman theorem* states that every compact, convex set $\mathcal{A}$ is equal to the convex hull of its extreme points.

2.6.2 A few theorems about convex analysis in real Euclidean spaces

It will be helpful later for us to make use of the following three theorems about convex sets. These theorems concern just *real* Euclidean spaces $\mathbb{R}^\Sigma$, but this will not limit their applicability to quantum information theory: we will use them when considering spaces $\mathrm{Herm}\left(\mathbb{C}^\Gamma\right)$ of Hermitian operators, which may be considered as real Euclidean spaces taking the form $\mathbb{R}^{\Gamma \times \Gamma}$ (as discussed in the previous lecture).

The first theorem is *Carathéodory's theorem*. It implies that every element in the convex hull of a subset $\mathcal{A} \subseteq \mathbb{R}^\Sigma$ can always be written as a convex combination of a small number of points in $\mathcal{A}$ (where *small* means at most $|\Sigma| + 1$). This is true regardless of the size or any other properties of the set $\mathcal{A}$.

Theorem 2.6 (Carathéodory's theorem)**.** *Let $\mathcal{A}$ be any subset of a real Euclidean space $\mathbb{R}^\Sigma$, and let $m = |\Sigma| + 1$. For every element $u \in \operatorname{conv}(\mathcal{A})$, there exist m (not necessarily distinct) points $u_1, \ldots, u_m \in \mathcal{A}$, such that u may be written as a convex combination of $u_1, \ldots, u_m$.*

The second theorem is an example of a *minmax* theorem that is attributed to Maurice Sion. In general, minmax theorems provide conditions under which a minimum and a maximum can be reversed without changing the value of the expression in which they appear.

Theorem 2.7 (Sion's minmax theorem)**.** *Suppose that $\mathcal{A}$ and $\mathcal{B}$ are compact and convex subsets of a real Euclidean space $\mathbb{R}^\Sigma$. It holds that*

$$\min_{u \in \mathcal{A}} \max_{v \in \mathcal{B}} \langle u, v \rangle = \max_{v \in \mathcal{B}} \min_{u \in \mathcal{A}} \langle u, v \rangle .$$

This theorem is not actually as general as the one proved by Sion, but it will suffice for our needs. One of the ways it can be generalized is to drop the condition that one of the two sets is compact (which generally requires either the minimum or the maximum to be replaced by an infimum or supremum).

Finally, let us state one version of a *separating hyperplane* theorem, which essentially states that if one has a closed, convex subset $\mathcal{A} \subset \mathbb{R}^\Sigma$ and a point $u \in \mathbb{R}^\Sigma$ that is not contained in $\mathcal{A}$, then it is possible to cut $\mathbb{R}^\Sigma$ into two separate (open) half-spaces so that one contains $\mathcal{A}$ and the other contains u.

Theorem 2.8 (Separating hyperplane theorem)**.** *Suppose that $\mathcal{A}$ is a closed and convex subset of a real Euclidean space $\mathbb{R}^\Sigma$ and $u \in \mathbb{R}^\Sigma$ not contained in $\mathcal{A}$. There exists a vector $v \in \mathbb{R}^\Sigma$ for which*

$$\langle v, w \rangle > \langle v, u \rangle$$

for all choices of $w \in \mathcal{A}$.

There are other separating hyperplane theorems that are similar in spirit to this one, but this one will be sufficient for us.