Mathematical foundations of quantum theory

Milán Mosonyi

Contents

1 Topology

1.1 Metric spaces and topological spaces

Recall the following notions from the theory of metric spaces. A *metric* d on a set $\mathcal{X}$ is a function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ such that

- $d(x, y) \geq 0$ for all $x, y \in \mathcal{X}$, and $d(x, y) = 0 \iff x = y$,
- $d(x, y) = d(y, x)$ for all $x, y \in \mathcal{X}$,
- $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in \mathcal{X}$ (triangle inequality).

For any $x \in \mathcal{X}$ and $\varepsilon > 0$, the *open ε-ball* $B(x, \varepsilon)$ of radius ε around x is $B(x, \varepsilon) := \{y \in \mathcal{X} : d(x, y) < \varepsilon\}$. A set $G \subseteq \mathcal{X}$ is called *open w.r.t. the metric d* if for every $x \in G$, there exists an $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq G$. We denote the set of open sets w.r.t. d by τ_d. A function $f : (\mathcal{X}, d_{\mathcal{X}}) \to (\mathcal{Y}, d_{\mathcal{Y}})$ is called *continuous* if

$$\forall x \in \mathcal{X}, \quad \forall \varepsilon > 0 : \quad \exists \delta > 0 \text{ s.t. } d(y, x) < \delta \implies d(f(y), f(x)) < \varepsilon.$$

Exercise 1.1. Show that a function $f : (\mathcal{X}, d_{\mathcal{X}}) \to (\mathcal{Y}, d_{\mathcal{Y}})$ is continuous if and only if $\forall G \in \tau_{d_{\mathcal{Y}}}$: $f^{-1}(G) \in \tau_{d_{\mathcal{X}}}$, i.e., if the inverse image of any open set in $\mathcal{Y}$ is an open set in $\mathcal{X}$.

The above example shows that as far as we are only interested in the concept of continuity of functions, it is enough to consider the collection of open sets, without reference to the metric generating it. The study of continuity, convergence, etc., in terms of open sets, leads to the concept of topological spaces.

Definition 1.2. Let $\mathcal{X}$ be a non-empty set. A set system $\tau \subseteq \mathcal{P}(\mathcal{X})$ is a *topology* on $\mathcal{X}$, if

- $\emptyset, \mathcal{X} \in \tau$,
- $\{G_i\}_{i \in \mathcal{I}} \subseteq \tau \implies \cup_{i \in \mathcal{I}} G_i \in \tau$, where $\mathcal{I}$ can be any index set,
  (closedness under arbitrary union)
- $\{G_i\}_{i \in \mathcal{I}} \subseteq \tau \implies \cap_{i \in \mathcal{I}} G_i \in \tau$ if $\mathcal{I}$ is finite.
  (closedness under finite intersection)

Elements of τ are called *open sets* w.r.t. the topology τ.

Example 1.3. $\{\emptyset, \mathcal{X}\}$ and $\mathcal{P}(\mathcal{X})$ are both topologies on $\mathcal{X}$, called the *anti-discrete topology* and the *discrete topology*, respectively.

Example 1.4. Let $(\mathcal{X}, d)$ be a metric space, and τ_d the collection of all open sets w.r.t. d. Show that τ_d is a topology. We call τ_d the *topology induced by the metric* d.

Exercise 1.5. Let $\mathcal{X}$ be a non-empty set. Find a metric d such that $\tau_d = \mathcal{P}(\mathcal{X})$, the discrete topology. Is such a metric unique?

Exercise 1.6. Let $\mathcal{X}$ be a non-empty set. Does there exist a metric that induces the anti-discrete topology?

Just as for the previously studied types of set systems, it is trivial to see that the intersection of any collection of topologies on a given set $\mathcal{X}$ is again a topology. In particular, for any set system $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$, we can define the topology generated by $\mathcal{A}$ as

$$\tau(\mathcal{A}) := \cap \{\kappa \subseteq \mathcal{P}(\mathcal{X}) : \mathcal{A} \subseteq \kappa, \kappa \text{ is a topology on } \mathcal{X}\}.$$

Definition 1.7. Let $\{(\mathcal{X}_i, \tau_i)\}_{i\in\mathcal{I}}$ be an arbitrary collection of topological spaces. We call the elements of $(\times)_{i\in\mathcal{I}} \tau_i$ *open boxes*. The *product topology* $\times_{i\in\mathcal{I}}\tau_i$ is the topology on $\times_{i\in\mathcal{I}}\mathcal{X}_i$ generated by all open boxes, i.e.,

$$\otimes_{i\in\mathcal{I}}\tau_i := \tau\left((\times)_{i\in\mathcal{I}} \tau_i\right).$$

We call $\otimes_{i\in\mathcal{I}}(\mathcal{X}_i, \tau_i) := (\times_{i\in\mathcal{I}}\mathcal{X}_i, \otimes_{i\in\mathcal{I}}\tau_i)$ the product of the topological spaces $\{(\mathcal{X}_i, \tau_i)\}_{i\in\mathcal{I}}$.

Exercise 1.8. Show that

$$\otimes_{i\in\mathcal{I}}\tau_i = \cup\left((\times)_{i\in\mathcal{I}} \tau_i\right),$$

i.e., a set is open in the product topology if and only if it is the union of open boxes. Conclude that

$$\tau_{\mathbb{R}^d} = \otimes_{i=1}^{d}\tau_{\mathbb{R}},$$

where for any $d \in \mathbb{N}$, $\tau_{\mathbb{R}^d}$ denotes the usual Euclidean topology of $\mathbb{R}^d$, generated by the Euclidean metric. More generally, for any $d_1, \ldots, d_r \in \mathbb{N}$,

$$\tau_{\mathbb{R}^{d_1+\ldots+d_r}} = \otimes_{i=1}^{d}\tau_{\mathbb{R}^{d_i}}.$$

Exercise 1.9. Let $\{(\mathcal{X}_i, d_i)\}_{i\in\mathcal{I}}$ be metric spaces, where $\mathcal{I}$ is a finite index set.

(i) Let

$$(\times_{i\in\mathcal{I}}d_i)_1(x, y) := \sum_{i\in\mathcal{I}} d(x_i, y_i),$$

$$(\times_{i\in\mathcal{I}}d_i)_2(x, y) := \left(\sum_{i\in\mathcal{I}} d(x_i, y_i)^2\right)^{1/2},$$

$$(\times_{i\in\mathcal{I}}d_i)_\infty(x, y) := \max_{i\in\mathcal{I}} d(x_i, y_i), \qquad x, y \in \times_{i\in\mathcal{I}}\mathcal{X}_i.$$

Show that all of the above are metrics on $\times_{i\in\mathcal{I}}\mathcal{X}_i$, and they all generate the product topology $\otimes_{i\in\mathcal{I}}\tau_{d_i}$.

(ii) Show that for any $1 \leq p < +\infty$,

$$(\times_{i\in\mathcal{I}} d_i)_p(x,y) := \left(\sum_{i\in\mathcal{I}} d(x_i,y_i)^p\right)^{1/p}, \qquad x,y \in \times_{i\in\mathcal{I}}\mathcal{X}_i$$

also generates $\otimes_{i\in\mathcal{I}}\tau_{d_i}$. (Hint: Use the Hölder inequality to show the triangle inequality.)

1.2 Separation properties

Definition 1.10. A topological space $(\mathcal{X},\tau)$ is

- *Hausdorff (T2)* if any two points can be separated by open sets, i.e., for any $x,y\in\mathcal{X}$, $x\neq y$, there exist $U\in\tau_x$ and $V\in\tau_y$ such that $U\cap V=\emptyset$.
- *regular*, if any closed set $F\subseteq\mathcal{X}$ and any point $x\in\mathcal{X}\setminus F$ can be separated by open sets, i.e., there exist $U\in\tau_x$ and $V\in\tau$ such that $V\supseteq F$ and $U\cap V=\emptyset$.
- *normal*, if any two disjoint closed sets can be separated by open sets, i.e., for any two disjoint closed F_1,F_2 there exist open sets $U_1,U_2\in\tau$ such that $U_1\supseteq F_1$, $U_2\supseteq F_2$ and $U_1\cap U_2=\emptyset$.

Lemma 1.11. A topological space $(\mathcal{X},\tau)$ is regular if and only if for any $x\in\mathcal{X}$ and $G\in\tau_x$ there exists a $U\in\tau_x$ such that $\overline{U}\subseteq G$.

Proof. Assume that $(\mathcal{X},\tau)$ is regular, and $G\in\tau_x$. Then $x\notin\mathcal{X}\setminus G=:F$, and hence there exist disjoint open sets $U\in\tau_x$ and $V\supseteq F$. Thus $\overline{U}\subseteq\mathcal{X}\setminus V\subseteq\mathcal{X}\setminus F=G$.

Conversely, if F is closed and $x\in\mathcal{X}\setminus F=:G$ then any $U\in\tau_x$ such that $\overline{U}\subseteq G$ defines an open $V:=\mathcal{X}\setminus\overline{U}$ containing F and disjoint from U. □

Lemma 1.12. Any closed subset of a compact topological space is compact.

Proof. Let $(\mathcal{X},\tau)$ be compact and $F\subseteq\mathcal{X}$ be closed. If $F\subseteq\cup_{i\in\mathcal{I}}U_i$ is an open cover of F then $(\mathcal{X}\setminus F)\cup(\cup_{i\in\mathcal{I}}U_i)$ is an open cover of $\mathcal{X}$ and hence there exists a finite subcover $\mathcal{X}\subseteq(\mathcal{X}\setminus F)\cup(\cup_{i\in\mathcal{I}_0}U_i)$, where $\mathcal{I}_0\subseteq\mathcal{I}$ is finite. Obviously, $F\subseteq\cup_{i\in\mathcal{I}_0}U_i$. □

Lemma 1.13. A compact Hausdorff space is normal.

Proof. Let $(\mathcal{X},\tau)$ be a compact Hausdorff space and $F_1,F_2\subseteq\mathcal{X}$ be disjoint closed sets; then they are also compact by Lemma 1.12. By the Husdorff property, for any $x\in F_1$, $y\in F_2$, there exist $U_{x,y}\in\tau_x$ and $V_{x,y}\in\tau_y$ disjoint open sets. For a fixed $x\in F_1$, $\{V_{x,y}\}_{y\in F_2}$ is an open cover of F_2, and hence there exist $y_1,\ldots,y_m\in F_2$

such that $F_2 \subseteq \cup_{i=1}^m V_{x,y_i} =: V_x$. Let $U_x := \cap_{i=1}^m U_{x,y_i}$, so that U_x is an open set and $U_x \cap V_x = \emptyset$. Then $\{U_x\}_{x\in\mathcal{X}}$ is an open cover of F_1, and hence there exists an finite subcover $\{U_{x_j}\}_{j=1}^n$. Let $U := \cup_{j=1}^n U_{x_j}$ and $V := \cap_{j=1}^n V_{x_j}$. Then U is an open set containing F_1, V is an open set containing F_2, and $U \cap V = \emptyset$. □

Definition 1.14. A topological space $(\mathcal{X}, \tau)$ is *locally compact* if every neighbourhood of x contains a compact neighbourhood of x, or equivalently, if for every $x \in \mathcal{X}$ and every $U \in \tau_x$ there exists a compact set K such that $x \in \operatorname{int} K \subseteq K \subseteq U$.

Lemma 1.15. A Hausdorff topological space is locally compact if and only if for any $x \in \mathcal{X}$ and $U \in \tau_x$ there exists a $V \in \tau_x$ such that $\overline{V}$ is compact and $\overline{V} \subseteq U$.

Proof. The "if" direction is obvious. Assume now that $(\mathcal{X}, \tau)$ is locally compact, so that for any $x \in \mathcal{X}$ and $U \in \tau_x$ there exists a compact set K such that $x \in \operatorname{int} K \subseteq K \subseteq U$. Then $V := \operatorname{int} K$ is open, and $\overline{V} \subseteq K$ is a closed subset of a compact set, and hence it is compact by Lemma 1.12. □

Corollary 1.16. A locally compact Hausdorff space is regular.

Proof. Immediate from Lemmas 1.11 and 1.15. □

Lemma 1.17. Let $(\mathcal{X}, \tau)$ be a locally compact Hausdorff topological space. Then for every compact set K contained in an open set G, there exists an open set U such that $\overline{U}$ is compact, and $K \subseteq U \subseteq \overline{U} \subseteq G$.

Proof. By Lemma 1.15, for any $x \in K$ there exists a $U_x \in \tau_x$ such that $\overline{U}_x$ is compact and $\overline{U}_x \subseteq G$. Obviously, $\{U_x\}_{x\in\mathcal{X}}$ is an open cover of K, and hence there exist $x_1, \ldots, x_n \in K$ such that $K \subseteq \cup_{i=1}^n U_{x_i} =: U$. Then $\overline{U} \subseteq \cup_{i=1}^n \overline{U}_{x_i} \subseteq G$, and $\overline{U}$, as a closed subset of the compact set $\cup_{i=1}^n \overline{U}_{x_i}$, is itself compact. □

Theorem 1.18. (Urysohn's lemma) A topological space $(\mathcal{X}, \tau)$ is normal if and only if any two disjoint closed sets can be separated by a continuous function in the sense that if F_1, F_2 are disjoint closed sets then there exists a continuous function $f : \mathcal{X} \to [0, 1]$ such that $f|_{F_1} \equiv 1$ and $f|_{F_2} \equiv 0$.

Proposition 1.19. Let $(\mathcal{X}, \tau)$ be a locally compact Hausdorff space, and $K \subseteq G \in \tau$, where K is compact. Then there exists a continuous function $f : \mathcal{X} \to [0, 1]$ with compact support such that $f|_K \equiv 1$ and $\operatorname{supp} f \subseteq G$.

Proof. By Lemma 1.17, there exists an open set U such that $\overline{U}$ is compact, and $K \subseteq U \subseteq \overline{U} \subseteq G$. Now, $\overline{U}$ is a compact Hausdorff space, and hence it is normal by Lemma 1.13. By Urisohn's lemma, there exists a continuous function $f_0 : \overline{U} \to [0, 1]$

such that $f_0|_K \equiv 1$ and $f_0|_{\overline{U}\setminus U} \equiv 0$. Let us extend f_0 to $\mathcal{X} \setminus \overline{U}$ to be constant 0, and let f denote the resulting function on $\mathcal{X}$. For any closed set $B \subseteq \mathbb{R}$, we have

$$f^{-1}(B) = \begin{cases} f_0^{-1}(B), \, 0 \notin B, \\ f_0^{-1}(B) \cup (\mathcal{X} \setminus U). \end{cases}$$

Here $f_0^{-1}(B)$ is a closed subset of $\overline{U}$ in the subspace topology of $\overline{U}$, and hence there exists a closed set $F \subseteq \mathcal{X}$ such that $f_0^{-1}(B) = \overline{U} \cap F$; therefore $f_0^{-1}(B)$ is closed in $\mathcal{X}$. Since $\mathcal{X} \setminus U$ is also closed, we see that $f^{-1}(B)$ is closed for any closed subset $B \subseteq \mathbb{R}$, i.e., f is continuous on $\mathcal{X}$. The rest of the properties are immediate from the construction of f. $\square$

2 Measure theory primer

2.1 Introduction

The aim of a mathematical model in physics is to correctly reproduce measurement statistics. Assume that the results of some measurements can take values in some set $\mathcal{X}$; this will most often be the real line $\mathbb{R}$, in which case we talk about a real-valued measurement, or $\mathbb{R}^d$ for $d = 2, 3$. Imagine, for instance, that we want to determine the position of a particle, e.g., a photon. What we can do is that we set up a detector that "clicks" when it absorbs a photon, so if our detector clicks in the experiment, we know that our photon was at the location of the detector, and if it does not click then we know that it was somewhere else. By setting up multiple detectors, we can obtain more detailed information about the location of the particle, namely, whether it was at the same place as detector $D_1, \dots, D_r$, or somewhere else.

If we prepare many particles in the same way, and repeat the experiment many times (say, n times), then we obtain a measurement statistics $\frac{k_0}{n}, \frac{k_1}{n}, \dots, \frac{k_r}{n}$, where k_i is the number of times the particle was detected by detector i, when $i \in [n] := \{1, 2, \dots, n\}$, and k_0 is the number of times the particle was not detected by any of the detectors. We have $\sum_{i=0}^{n} \frac{k_i}{n} = 1$, i.e., the numbers $p_i = \frac{k_i}{n}$ form a probability distribution on the possible measurement outcomes. In this particular example the probabilities are rational number with denominator n, but since we want to build a model that correctly reproduces the measurement statistics for any number of repetitions, we should consider any rational numbers for p_i. For mathematical convenience, we will also consider probability distributions where the p_i can also be irrational numbers; from a physical point of view, this may be justified by the fact that any real number can be approximated by rational numbers with arbitrary precision.

Now, our aim is to find a mathematical model that correctly reproduces our measurement statistics, for all possible constellation of the detectors and all possible preparations of the photons. Note that the detectors are not point-like, but extended objects (and so are the particles, in fact), so the click of the k-th detector means that the particle was in a subset A_k of $\mathcal{X}$, where the latter may be $\mathbb{R}$, if the detectors are arranged along a line, or $\mathbb{R}^2$, if the detectors correspond to regions of a screen, or $\mathbb{R}^3$, if the detectors can be in a general position in the 3-dimensional space. Since we want to model the measurement statistics for all possible constellations of detectors, we need a model that assigns probabilities to subsets of $\mathcal{X}$ such that the probabilities sum to 1 on any finite collection of mutually exclusive events that is complete, i.e., if $A_1, \dots, A_n$ are disjoint subsets of $\mathcal{X}$ such that $\cup_{i=1}^{n} A_i = \mathcal{X}$ then $\sum_{i=1}^{n} p(A_i) = 1$.

In order to have a mathematically well-behaved theory, we will introduce some further requirements and restrictions: we will require the additivity of the probabilities also on any countably infinite family of mutually disjoint sets; this will guarantee

that our model has nice continuity properties. On the other hand, we will not necessarily assign a probability to all possible subsets of $\mathcal{X}$, but only those that are "nice enough" in some sense, but these will still contain all sets that we can imagine or describe in some constructive way. These requirements formally mean that the subsets to which we assign probabilities form a σ-algebra, that we define in Section 2.2. We give the mathematical definition of a probability measure on a σ-algebra in Section 2.3.

2.2 Measurable spaces

Our first goal is to discuss how to introduce a notion of volume for sets in the d-dimensional real Euclidean space $\mathbb{R}^d$ that is compatible with our everyday geometric intuition.

Let us introduce

$$\mathrm{Box}(\mathbb{R}^d) := \left\{ \times_{i=1}^d [a_i, b_i) : a_i, b_i \in \mathbb{R} \right\},$$

the elements of which are called *boxes*. It is clear that any reasonable notion of volume should satisfy

$$\mathrm{Vol}\left(\times_{i=1}^d [a_i, b_i) \right) := \prod_{i=1}^d (b_i - a_i).$$

Remark 2.1. We consider boxes whose sides are finite intervals that contain their left endpoints but not the right ones merely for notational convenience, so that we do not have to list all the other possibilities or invent cumbersome notations to cover them. It is anyway clear that the sets

$$\times_{i=1}^d (a_i, b_i), \qquad \times_{i=1}^d [a_i, b_i), \qquad \times_{i=1}^d (a_i, b_i], \qquad \times_{i=1}^d [a_i, b_i]$$

should have the same volume.

Let us now move on to what we require from a volume function on more general sets. It is again intuitively clear that the following should be satisfied:

- *finite additivity:* the volume of the union of finitely many disjoint sets should be the sum of the volumes of the individual sets:

$$\mathrm{Vol}\left(\cup_{i=1}^n A_i \right) = \sum_{i=1}^n \mathrm{Vol}(A_i).$$

- *translation-invariance:*

$$\mathrm{Vol}(A + x) = \mathrm{Vol}(A),$$

where for $A \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$, we define $A + x := \{a + x : a \in A\}$.

- *rotational invariance:*

$$\mathrm{Vol}(\mathcal{R}(A)) = \mathrm{Vol}(A), \qquad \mathcal{R} \in R(d),$$

where $\mathcal{R}(A) := \{\mathcal{R}(a) : a \in A\}$, and $R(d)$ denotes the set of rotations in $\mathbb{R}^d$ (a subset of the special orthogonal group $SO(d)$.)

The second and the third requirements together may be expressed as the requirement that the volume function is *invariant under rigid motions*.

Now, one runs into an unexpected difficulty: Intuitive though the above requirements may seem, it is not possible to define a volume function on *all* subsets of $\mathbb{R}^d$ that satisfies all of them, at least if one accepts the so-called *axiom of choice*, one of the key elements of the set-theoretic axiomatization of Mathematics. Indeed, Banach and Tarski showed that for any $d \geq 3$, a solid box in $\mathbb{R}^d$ can be decomposed into finitely many disjoint subsets such that applying only rotations and translations to these sets, it is possible to assemble two identical copies of the original box. This phenomenon, called the *Banach-Tarski paradox*, is clearly extremely counter-intuitive, and prompted some mathematicians to reject the axiom of choice. However, it is an indispensable tool to prove many important results in mathematics, some of which we will also encounter in these notes. Hence, the overwhelming majority of mathematicians accepts the axiom of choice, and gives up the existence of a volume function on all subsets of $\mathbb{R}^d$ with the above listed properties, and we will follow this majority here.

There are obviously two ways out of the above difficulty: one is to give up some of the requirements on the volume function formulated above; this however, would lead to an awkward notion of volume that we would prefer to avoid. The other option, which we will follow, is to keep the requirements, but to assign a volume to not every, but only some "nice enough" subsets of $\mathbb{R}^d$, including all boxes. In particular, the pieces into which a box is cut in the Banach-Tarski paradox will not be "nice enough" according to this definition.

To make this approach mathematically precise, we need to introduce a few new concepts. We will also take this opportunity to take our discussion onto a more general and abstract level, which will be very useful later.

Definition 2.2. For any set $\mathcal{X}$, let $\mathcal{P}(\mathcal{X}) := \{A \subseteq \mathcal{X}\}$ be the collection of all subsets of $\mathcal{X}$ (including the empty set $\emptyset$ and $\mathcal{X}$ itself). $\mathcal{P}$ stands for "potenz" in German, meaning "power" in English, and $\mathcal{P}(\mathcal{X})$ is called the *power set* of $\mathcal{X}$.

Next, we postulate what properties the collection of sets with a volume should have; we will call such sets *measurable*. We have already stated that the union of finitely many disjoint measurable sets should be measurable. It is also quite intuitive to require that if $A_1, A_2 \subseteq \mathbb{R}^d$ are measurable then so is their intersection $A_1 \cap A_2$,

as well as $A_1 \setminus (A_1 \cap A_2) = A_1 \setminus A_2$ (and then of course also $A_2 \setminus A_1$). Finally, we require that the increasing union of measurable sets is again measurable (see below). While this last requirement may seem sligthly less intuitive than the preceding ones, it will be very useful in building a theory where limits can be handled well.

Definition 2.3. Let $\mathcal{X} \neq \emptyset$ be a set and $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ be non-empty. We say that $\mathcal{A}$ is a *σ-ring* on $\mathcal{X}$ if it is closed under the following set operations:

(i) *set difference:*

$$A_1, A_2 \in \mathcal{A} \Longrightarrow A_2 \setminus A_1 \in \mathcal{A}. \tag{2.1}$$

(ii) *finite disjoint union:*

$$(A_i)_{i=1}^n \subseteq \mathcal{A}, \quad A_i \cap_{i \neq j} A_j = \emptyset \quad \Longrightarrow \quad \cup_{i=1}^n A_n \in \mathcal{A}. \tag{2.2}$$

(iii) *union of increasing sequences:*

$$(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A}, \quad A_n \subseteq A_{n+1},\, n \in \mathbb{N}, \quad \Longrightarrow \quad \cup_{n \in \mathbb{N}} A_n \in \mathcal{A}. \tag{2.3}$$

A σ-ring $\mathcal{A}$ is called a *σ-algebra* if $\mathcal{X} \in \mathcal{A}$.

Remark 2.4. Any σ-ring contains the empty set; indeed, for any $A \in \mathcal{A}$,

$$\emptyset = A \setminus A \in \mathcal{A}.$$

Remark 2.5. A σ-algebra $\mathcal{A}$ is also closed under *complement*:

$$A \in \mathcal{A} \quad \Longrightarrow \quad \mathcal{X} \setminus A \in \mathcal{A}.$$

Example 2.6. There are two trivial σ-algebras on any non-empty set $\mathcal{X}$: The minimal one $\{\emptyset, \mathcal{X}\}$, and the maximal one $\mathcal{P}(\mathcal{X})$. (Exercise: Check that these are indeed σ-algebras.)

In view of the above considerations, our aim is to find an extension of the volume function from the set of boxes to a σ-ring that contains all the boxes, such that the extension satisfies the requirements postulated above. First, note the following:

Remark 2.7. Any σ-ring $\mathcal{A}$ on $\mathbb{R}^d$ that contains $\mathrm{Box}(\mathbb{R}^d)$ is also a σ-algebra. Indeed, let $B_n := \times_{i \in [d]} [-n, n) \in \mathrm{Box}(\mathbb{R}^d) \subseteq \mathcal{A}$, $n \in \mathbb{N}$; then $\mathbb{R}^d = \cup_{n \in \mathbb{N}} B_n \in \mathcal{A}$, according to (2.3).

Remark 2.8. The above observation is one of the reasons why we will mainly be interested in σ-algebras, and not in the more general structure of σ-rings.

Of course, the smaller the σ-algebra $\mathcal{A}$ contaning $\mathrm{Box}(\mathbb{R}^d)$, the easier to find an extension of the volume function to $\mathcal{A}$ with the desired properties. Thus, the most economic approach is to look for an extension onto the *smallest* σ-algebra containing $\mathrm{Box}(\mathbb{R}^d)$; the following simple observations guarantee that it indeed exists.

Lemma 2.9. The intersection of any collection of σ-rings/σ-algebras on the same set is again a σ-ring/σ-algebra.

Proof. Trivial, exercise. □

Corollary 2.10. For any collection of subsets $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$, there exists a smallest σ-algebra containing $\mathcal{A}$, which we denote by $\sigma(\mathcal{A})$, and call it the *σ-algebra generated by $\mathcal{A}$*.

Proof. Let $\sigma(\mathcal{A}) := \cap\{\mathcal{B} \subseteq \mathcal{P}(\mathcal{X}) : \mathcal{B}\ \sigma\text{-algebra},\ \mathcal{A} \subseteq \mathcal{B}\}$. Then $\sigma(\mathcal{A})$ is a σ-algebra by Lemma 2.9, and it is clearly a subset of any σ-algebra containing $\mathcal{A}$. □

Definition 2.11. The *Borel σ-algebra* of $\mathbb{R}^d$ (denoted as $\mathcal{B}(\mathbb{R}^d)$) is the smallest σ-algebra on $\mathbb{R}^d$ containing all the boxes in $\mathbb{R}^d$, i.e.,

$$\mathcal{B}(\mathbb{R}^d) := \sigma(\mathrm{Box}(\mathbb{R}^d)).$$

The elements of $\mathcal{B}(\mathbb{R}^d)$ are called *Borel sets* or *Borel measurable sets*.

Hence, our aim will be to find an extension of the volume function onto the Borel σ-algebra, with the required properties. We will do this in the next section; in the rest of this section, we explore some further properties of σ-rings and σ-algebras that will be useful later.

We start with the following simple but very useful disjunctization lemma:

Lemma 2.12. The countable union of elements of a σ-ring can be written as the countable union of disjoint elements of the σ-ring, and also as the union of an increasing chain of elements in the σ-ring.

Moreover, the union is again an element of the σ-ring.

Proof. Let $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, and for every $n \in \mathbb{N}$, let $\tilde{A}_n := A_n \setminus \left(\cup_{i=1}^{n-1} A_i\right)$, $n \in \mathbb{N}$. Then $\tilde{A}_1 = A_1 \in \mathcal{A}$, $\tilde{A}_2 = A_2 \setminus A_1 \in \mathcal{A}$ by (2.1), and $A_1 \cup A_2 = \tilde{A}_1 \cup \tilde{A}_2 \in \mathcal{A}$ by (2.2). Continuing this argument, we see that $\tilde{A}_n \in \mathcal{A}$, $n \in \mathbb{N}$, and $\cup_{i=1}^{n} A_i = \cup_{i=1}^{n} \tilde{A}_i \in \mathcal{A}$ by (2.2). Finally,

$$\cup_{n\in\mathbb{N}} A_n = \cup_{n\in\mathbb{N}} \tilde{A}_n = \cup_{n\in\mathbb{N}} \left(\cup_{i=1}^{n} \tilde{A}_i\right) \in \mathcal{A},$$

where the last step is due to (2.3), since $\cup_{i=1}^{n} \tilde{A}_i \subseteq \cup_{i=1}^{n+1} \tilde{A}_i$, $n \in \mathbb{N}$. □

Corollary 2.13. A σ-ring $\mathcal{A}$ is also closed under

- *countable union:*

$$(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A} \implies \cup_{n\in\mathbb{N}} A_n \in \mathcal{A}, \tag{2.4}$$

  by Lemma 2.12;

- *countable intersection*; indeed

$$(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}, \implies \cap_{n\in\mathbb{N}} A_n = \left(\bar{A}\right) \setminus \left(\cup_{n\in\mathbb{N}}(\bar{A} \setminus A_n)\right) \in \mathcal{A},$$

  where $\bar{A} := \cup_{n\in\mathbb{N}} A_n$, and we used (2.4) and (2.1).

Remark 2.14. Obviously, a σ-algebra is closed under finite intersections and unions as well, since we can always take arbitrarily many of the A_n to be $\emptyset$ or the whole set $\mathcal{X}$.

Exercise 2.15. Let $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ be closed under set difference. Show that the following are equivalent:

(i) $\mathcal{A}$ is closed under countable unions.

(ii) • $\mathcal{A}$ is closed under *finite disjoint union*, i.e., $(A_i)_{i=1}^n \subseteq \mathcal{A}$, $A_i \cap_{i\neq j} = \emptyset$ implies $\cup_{i=1}^n A_i \in \mathcal{A}$, and

• $\mathcal{A}$ is closed under the *union of increasing sequences*, i.e., if $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ is such that $A_n \subseteq A_{n+1}$, $n \in \mathbb{N}$, then $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

The above exercise immediately yields the following

Corollary 2.16. A non-empty $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ is a σ-ring if and only if it is closed under set difference and countable union.

We will often restrict our considerations to a given subset of $\mathbb{R}^d$, e.g., some interval in $\mathbb{R}$. In this case it is useful to introduce the following:

Definition 2.17. Let $A \in \mathcal{B}(\mathbb{R}^d)$ be a Borel set. The Borel σ-algebra on A is simply the collection of Borel sets in A, i.e.,

$$\mathcal{B}(A) := \{B \in \mathcal{B}(\mathbb{R}^d) : B \subseteq A\}.$$

It is easy to see that $\mathcal{B}(A)$ is a σ-algebra on A.

We will sometimes need to consider functions taking values in the extended real line

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}.$$

The Borel σ algebra in this case is defined to be

$$\mathcal{B}(\overline{\mathbb{R}}) := \{A, A \cup \{+\infty\}, A \cup \{-\infty\}, A \cup \{\pm\infty\} : A \in \mathcal{B}(\mathbb{R})\}.$$

It is often useful to have simple generating sets for σ-algebras. For the Borel σ-algebra of $\mathbb{R}$ (resp. $\overline{\mathbb{R}}$), the following will be useful:

Exercise 2.18. Show that

$$\begin{aligned}
\mathcal{B}(\mathbb{R}) &= \sigma\left(\{(c, +\infty) : c \in \mathbb{R}\}\right) = \sigma\left(\{[c, +\infty) : c \in \mathbb{R}\}\right) \\
&= \sigma\left(\{(-\infty, c) : c \in \mathbb{R}\}\right) = \sigma\left(\{(-\infty, c] : c \in \mathbb{R}\}\right), \\
\mathcal{B}(\overline{\mathbb{R}}) &= \sigma\left(\{(c, +\infty] : c \in \mathbb{R}\}\right) = \sigma\left(\{[c, +\infty] : c \in \mathbb{R}\}\right) \\
&= \sigma\left(\{[-\infty, c) : c \in \mathbb{R}\}\right) = \sigma\left(\{[-\infty, c] : c \in \mathbb{R}\}\right).
\end{aligned}$$

Exercise 2.19. (i) Show that the Borel σ-algebra on $\mathbb{R}^d$ contains all open and all closed subsets of $\mathbb{R}^d$.

(ii) Conclude that $\mathcal{B}(\mathbb{R}^d)$ contains all boxes of the form $\times_{i=1}^d J_i$, where $J_i \subseteq \mathbb{R}$ is an arbitrary interval.

(iii) Conclude that every singleton $\{x\}$, $x \in \mathbb{R}^d$, is a Borel set.

(iv) Show that the Borel σ-algebra is the smallest σ-algebra on $\mathbb{R}^d$ that contains all open sets (equivalently, all closed sets).

Remark 2.20. In a general topological space, the Borel σ-algebra is defined to be the smallest σ-algebra containing all open sets. It is easy to verify that in the cases in which we defined the Borel σ-algebras above, our definition coincides with this more general definition.

It is natural to ask if the Cartesian product of two Borel sets is again a Borel set. The answer is easily seen to be yes, as we show below. To formulate it, we introduce the notion of product σ-algebra.

Note that if $\mathcal{A}_i$ is a σ-algebra on $\mathcal{X}_i$ for $i \in [n]$ then the collection of all Cartesian products

$$\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n := \{A_1 \times \ldots \times A_n : A_i \in \mathcal{A}_i,\, i \in [n]\}$$

need not be a σ-algebra in general, but we can of course always take the generated σ-algebra

$$\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n := \sigma\left(\mathcal{A}_1 \left(\times\right) \ldots \left(\times\right) \mathcal{A}_n\right), \tag{2.5}$$

which we call the *product* of the σ-algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$. Note that (2.5) makes sense, and gives a σ-algebra for arbitrary $\mathcal{A}_i \subseteq \mathcal{X}_i$ that need not be σ-algebras. Moreover, we may define the product σ-ring exactly the same way. Note that

$$\mathrm{Box}(\mathbb{R}^d) = \mathrm{Box}(\mathbb{R}) \left(\times\right) \ldots \left(\times\right) \mathrm{Box}(\mathbb{R}). \tag{2.6}$$

Proposition 2.21. For any $d_1, \ldots, d_n \in \mathbb{N}$,

$$\mathcal{B}(\mathbb{R}^{d_1}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R}^{d_n}) = \mathcal{B}(\mathbb{R}^{d_1+\ldots+d_n}). \tag{2.7}$$

In particular,

$$\otimes_{i=1}^{d} \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d). \tag{2.8}$$

Proof. We only prove (2.8), as the proof of (2.7) goes exactly the same way. For $A \subseteq \mathbb{R}$ and $i \in [d]$, let

$$A \times \mathbb{R}^{[d]\setminus i} := \{\underline{x} \in \mathbb{R}^d : \, x_i \in A\}.$$

Then

$$\mathrm{Box}(\mathbb{R}) \subseteq \mathcal{B}_i := \{A \in \mathcal{B}(\mathbb{R}) : \, A \times \mathbb{R}^{[d]\setminus i} \in \mathcal{B}(\mathbb{R}^d)\} \subseteq \mathcal{B}(\mathbb{R}). \tag{2.9}$$

It is straightforward to verify that $\mathcal{B}_i$ is a σ-algebra, and hence, by (2.9), $\mathcal{B}(\mathbb{R}) = \sigma\left(\mathrm{Box}(\mathbb{R})\right) \subseteq \mathcal{B}_i \subseteq \mathcal{B}(\mathbb{R})$ yields that $\mathcal{B}_i = \mathcal{B}(\mathbb{R})$. Next, for any $A_i \in \mathcal{B}(\mathbb{R})$, the above yields

$$A_1 \times \ldots \times A_n = \cap_{i=1}^{n} \underbrace{\left(A_i \times \mathbb{R}^{[d]\setminus i}\right)}_{\in \mathcal{B}(\mathbb{R}^d)} \in \mathcal{B}(\mathbb{R}^d). \tag{2.10}$$

Hence, we get

$$\begin{aligned} \mathcal{B}(\mathbb{R}^d) = \sigma\left(\mathrm{Box}(\mathbb{R}^d)\right) &= \sigma\left(\mathrm{Box}(\mathbb{R}) \left(\times\right) \ldots \left(\times\right) \mathrm{Box}(\mathbb{R})\right) \\ &\subseteq \sigma\left(\mathcal{B}(\mathbb{R}) \left(\times\right) \ldots \left(\times\right) \mathcal{B}(\mathbb{R})\right) \subseteq \mathcal{B}(\mathbb{R}^d), \end{aligned} \tag{2.11}$$

where the first equality is by definition, the second equality is due to (2.6), the first containment is trivial, and the last one is due to (2.10). Hence, all containments in (2.11) are equalities, and we obtain (2.8). □

It is clear that the set of boxes, $\mathrm{Box}(\mathbb{R}^d)$ does not form a σ-ring; in general, the set difference of two boxes need not be a box, and already the union of two disjoint non-empty boxes is not a box. However, $\mathrm{Box}(\mathbb{R}^d)$ has the following weaker properties:

Lemma 2.22. Let $\mathcal{S} := \mathrm{Box}(\mathbb{R}^d)$. Then

- $\mathcal{S}$ is closed under finite intersections:

$$A_1, \dots, A_r \in \mathcal{S} \implies \cap_{i=1}^r A_i \in \mathcal{S};$$

- the set difference of any two elements in $\mathcal{A}$ can be written as the disjoint union of finitely many elements in $\mathcal{A}$:

$$A_1, A_2 \in \mathcal{S} \implies A_1 \setminus A_2 = \cup_{i=1}^n B_i \text{ for some } n \in \mathbb{N} \text{ and } B_1, \dots, B_n \in \mathcal{S}, \quad B_i \cap_{i \neq j} B_j = \emptyset.$$

Definition 2.23. A set system $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$ on an arbitrary set $\mathcal{X}$ is called a *semi-ring* if it satisfies the two properties given in Lemma reflemma:semi-ring.

Remark 2.24. It is easy to see that for the finite intersection property, it is sufficient to check that $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{A}$.

Example 2.25. $\mathrm{Box}(\mathbb{R}^d)$ is a semi-ring.

Although a semi-ring is a much "weaker" structure then a σ-ring, it is still rich enough to have many useful properties, and in fact, it is a very central concept in measure theory, as it is somehow the minimal structure that guarantees "good" properties.

One of the key features of semi-rings is that they allow for a similar disjunctization of set sequences as σ-rings (Lemma 2.12). We explore this in the following exercises.

Remark 2.26. While this disjunctization property may not seem terribly exciting at first sight, it will actually be the basis of many important features of the measure and integral theory that we develop; for instance, that even the most exotic measurable set can be arbitrarily well approximated by finitely many disjoint boxes, in the sense that the measure of their difference is negligible (Exercise 2.53), or that any integrable function can be arbitrarily well approximated by continuous functions.

Exercise 2.27. Let $A, A_1, \dots, A_r$ be elements of a semi-ring $\mathcal{S}$. Show that there exist pairwise disjoint elements $B_1, \dots, B_m \in \mathcal{S}$ such that

$$A \setminus (\cup_{i=1}^r A_i) = B_1 \cup \dots \cup B_m.$$

Solution: Hidden.

Exercise 2.28. Let $A_1, \ldots, A_r$ be finitely many elements in a semi-ring $\mathcal{S}$. Show that there exist pairwise disjoint elements $B_1, \ldots, B_m \in \mathcal{S}$ such that

$$A_1 \cup \ldots \cup A_r = B_1 \uplus \ldots \uplus B_m.$$

Solution: Hidden.

Exercise 2.29. Let $\{A_i\}_{i\in\mathcal{I}} \subseteq \mathcal{S}$ be a countable collection of elements in a semi-ring $\mathcal{S}$. Then there exists a countable collection $\{B_j\}_{j\in\mathcal{J}} \subseteq \mathcal{S}$ of pairwise disjoint elements in $\mathcal{S}$ such that

$$\cup_{i\in\mathcal{I}} A_i = \uplus_{j\in\mathcal{J}} B_j.$$

Solution: Hidden.

Specializing the above to boxes, we get the following:

Corollary 2.30. Let $\{A_i\}_{i\in\mathcal{I}} \subseteq \mathrm{Box}(\mathbb{R}^d)$ be a countable collection of boxes. Then there exists a countable collection $\{B_j\}_{j\in\mathcal{J}} \subseteq \mathrm{Box}(\mathbb{R}^d)$ of pairwise disjoint boxes such that

$$\cup_{i\in\mathcal{I}} A_i = \uplus_{j\in\mathcal{J}} B_j.$$

2.3 Measures

Let us now continue our project of assigning a volume to subsets of $\mathbb{R}^d$. In the previous section we have discussed the properties that the collection of sets to which we assign a volume should satisfy, and we came to the conclusion that they should form a σ-algebra on $\mathbb{R}^d$. Hence, our aim now is to show that a volume function may be defined on the smallest σ-algebra containing all boxes, namely, the Borel σ-algebra, satisfying the intuitive requirements that we postulated in the previous section.

We start our discussion by having a more detailed look at those properties, and also in a more abstract setting, as that will be very useful, e.g., for defining probabilistic models of physical systems. In this general setting, instead of the geometric idea of a volume, we will want to assign a number to every element of a σ-algebra in a way that imitates the requirements that we postulated for the volume function. Such an assignment of numbers to sets will be called a *measure*, and the d-dimensional volume, as well as the probability of sets in a probability space, will be special cases of it.

Note that we want to assign a measure to elements of a σ-algebra, which motivates the following terminology:

Definition 2.31. A pair $(\mathcal{X}, \mathcal{A})$, where $\mathcal{A}$ is a σ-algebra on the set $\mathcal{X}$, is called a *measurable space*. For a given measurable space, the elements of $\mathcal{A}$ are called *measurable sets*.

Definition 2.32. A function $\mu : \mathcal{A} \to [0, +\infty]$ on a σ-algebra $\mathcal{A}$ is a (positive) *measure* if $\mu(\emptyset) = 0$, and it is *countably additive* (or *σ-additive*), i.e.,

$$(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}, \quad A_n \cap_{n\neq m} A_m = \emptyset \implies \mu\left(\cup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

A triple $(\mathcal{X}, \mathcal{A}, \mu)$ is called a *measure space* if $\mathcal{A}$ is a σ-algebra on $\mathcal{X}$, and μ is a measure on $\mathcal{A}$. We denote the set of all measures on $(\mathcal{X}, \mathcal{A})$ by $\mathcal{M}(\mathcal{X}, \mathcal{A})$.

Definition 2.33. We say that a measure $\mu \in \mathcal{M}(\mathcal{X}, \mathcal{A})$ is *finite* if $\mu(\mathcal{X}) < +\infty$, and it is a *probability measure* if $\mu(\mathcal{X}) = 1$. We denote the set of probability measures on $(\mathcal{X}, \mathcal{A})$ by $\mathcal{S}(\mathcal{X}, \mathcal{A})$, and call it the *state space* of $(\mathcal{X}, \mathcal{A})$.

Remark 2.34. It is customary in probability theory to use the notation Ω instead of $\mathcal{X}$ for the basis set, and P to denote a probability measure on some σ-algebra on Ω. A triple $(\Omega, \mathcal{F}, P)$ is called a *probability space* if $\mathcal{F}$ is a σ-algebra on Ω, and P a probability measure on $\mathcal{F}$. Elements of $\mathcal{F}$ are called *events*, and $P(E)$ gives the probability of an event $E \in \mathcal{F}$ occuring. For instance, to describe the outcome of rolling a die, one may choose $\Omega = [6] = \{1, 2, \ldots, 6\}$, $\mathcal{F} = \mathcal{P}([6])$, and $P(\{i\}) := 1/6$ for every $i \in [6]$.

Example 2.35. Here we give a few simple examples of measures. We will encounter more complicated ones later on.

(i) $\mathcal{A}$ is an arbitrary σ-algebra, and $\mu \equiv 0$.

(ii) $\mathcal{A}$ is an arbitrary σ-algebra, $\mu(\emptyset) = 0$, and $\mu(A) = +\infty$ for all $A \in \mathcal{A} \setminus \{\emptyset\}$.

(iii) $\mathcal{A}$ is an arbitrary σ-algebra, and $\mu(A) := |A|$ is the cardinality of $A \in \mathcal{A}$ when it is finite, and $+\infty$ otherwise. This is called the *counting measure*.

(iv) $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ is an arbitrary σ-algebra, and for a fixed $x \in \mathcal{X}$,

$$\mu(A) := \delta_x(A) := \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

This is called the *point measure* or *Dirac measure* concentrated at point x.

(v) $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ is an arbitrary σ-algebra, and for fixed $x_1, \ldots, x_r \in \mathcal{X}$, $c_1, \ldots, c_r \in [0, +\infty]$,

$$\mu := \sum_{i=1}^{r} c_i \delta_{x_i} : \; A \mapsto \sum_{i:\, x_i \in A} c_i, \qquad A \in \mathcal{A}.$$

This is called a *finitely supported measure*, and is a generalization of the Dirac measure.

Remark 2.36. We can take generalized positive linear combinations of measures in a natural way, and obtain again a measure. That is, if $\mu_i \in \mathcal{M}(\mathcal{X}, \mathcal{A})$, and $c_i \in [0, +\infty]$, $i \in [n]$, then

$$(c_1\mu_1 + \ldots + c_n\mu_n)(A) := c_1\mu_1(A) + \ldots + c_n\mu_n(A), \qquad A \in \mathcal{A},$$

defines a measure on $\mathcal{A}$. The measure in the last example above is obtained from the Dirac measures by this construction.

Note that we allow some of the coefficients to be $+\infty$, and if $c_i = +\infty$ and $\mu_i(A) = 0$ then we need to evaluate an expression of the form $(+\infty) \cdot 0$. The standard convention in measure theory, that we will use throughout the text, is to define

$$(\pm\infty) \cdot 0 := 0.$$

Remark 2.37. We may define the sum of arbitrarily many (e.g., continuum many) non-negative numbers $\lambda_i \in [0, +\infty]$, $i \in I$, as

$$\sum_{i \in I} \lambda_i := \sup \left\{ \sum_{i \in J} \lambda_i : \; J \subseteq I \text{ finite} \right\}.$$

With this convention, we may generalize (iv) in Example 2.35 by allowing the positive linear combination of arbitrarily many measures, i.e., if $\mu_i \in \mathcal{M}(\mathcal{X}, \mathcal{A})$, and $c_i \in [0, +\infty]$, $i \in I$, where I may be any index set, then we set

$$\left(\sum_{i \in I} c_i \mu_i \right)(A) := \sum_{i \in I} c_i \mu_i(A).$$

For instance, the counting measure in (iii) of Example 2.35 is of this form, with $I = \mathcal{X}$, $c_x = 1$, $\mu_x = \delta_x$, $x \in \mathcal{X}$.

Definition 2.38. We say that a measure $\mu \in \mathcal{M}(\mathcal{X}, \mathcal{A})$ is *σ-finite* if $\mathcal{X}$ is the union of countably many sets in $\mathcal{A}$ with finite measure.

Exercise 2.39. Which of the measures in Example 2.35 are a) finite, b) probability measures, c) σ-finite measures?

Next, we briefly discuss some general properties of measures.

It is clear that a measure is also *finitely additive*, i.e.,

$$(A_i)_{i=1}^n \subseteq \mathcal{A}, \quad A_i \cap_{i \neq j} A_j = \emptyset \quad \Longrightarrow \quad \mu\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i),$$

since we can choose arbitrarily many of the A_n in Definition 2.32 to be the empty set. It is also clear that a measure is *monotone*, in the sense that for $A, B \in \mathcal{A}$, $A \subseteq B$, we have

$$\mu(B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \geq \mu(A).$$

Lemma 2.40. (Monotone continuity of measures) Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, and $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ be an increasing sequence, i.e., $A_n \subseteq A_{n+1}$, $n \in \mathbb{N}$. Then

$$\mu\left(\cup_{n\in\mathbb{N}} A_n\right) = \sup_{n\in\mathbb{N}} \mu(A_n) = \lim_{n\to+\infty} \mu(A_n).$$

Proof. As in Lemma 2.12, we write $\tilde{A}_n := A_n \setminus \left(\cup_{i=1}^{n-1} A_i\right)$, so that $\tilde{A}_n \in \mathcal{A}$, $n \in \mathbb{N}$, and for any $N \in \mathbb{N}$, $A_N = \cup_{n=1}^N \tilde{A}_n$. Then

$$\begin{aligned}\mu\left(\cup_{n\in\mathbb{N}} A_n\right) &= \mu\left(\cup_{n\in\mathbb{N}} \tilde{A}_n\right) = \sum_{n=1}^{+\infty} \mu(\tilde{A}_n) = \lim_{N\to+\infty} \sum_{n=1}^{N} \mu(\tilde{A}_n) \\ &= \lim_{N\to+\infty} \mu(A_N) = \sup_{n\in\mathbb{N}} \mu(A_N),\end{aligned}$$

where in the second and the fourth identity we used the σ-additivity of the measure, and the last identity is due to the monotonicity. □

Exercise 2.41. Show the following complement of the above statement: If $(\mathcal{X}, \mathcal{A}, \mu)$ is a measure space, and $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ is a decreasing sequence, i.e., $A_n \supseteq A_{n+1}$, $n \in \mathbb{N}$, such that $\mu(A_1) < +\infty$, then

$$\mu\left(\cap_{n\in\mathbb{N}} A_n\right) = \inf_{n\in\mathbb{N}} \mu(A_n) = \lim_{n\to+\infty} \mu(A_n).$$

Show an example where $\mu(A_1) = +\infty$ and the above continuity relation does not hold.

Lemma 2.42. (σ-subadditivity of measures) In any measure space $(\mathcal{X}, \mathcal{A}, \mu)$,

$$(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A} \quad \Longrightarrow \quad \mu\left(\cup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

Proof. Let $\tilde{A}_n := A_n \setminus \left(\cup_{i=1}^{n-1} A_i\right)$, $n \in \mathbb{N}$. Then, as we have seen in Lemma 2.40, $\tilde{A}_n \cap_{n \neq m} \tilde{A}_m = \emptyset$, and $\uplus_{n\in\mathbb{N}} \tilde{A}_n = \cup_{n\in\mathbb{N}} A_n$, and hence

$$\mu\left(\cup_{n\in\mathbb{N}} A_n\right) = \mu\left(\uplus_{n\in\mathbb{N}} \tilde{A}_n\right) = \sum_{n\in\mathbb{N}} \underbrace{\mu(\tilde{A}_n)}_{\leq \mu(A_n)} \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

□

Exercise 2.43. Show that the countable union of zero measure sets is again of zero measure. That is, if $A_n \in \mathcal{A}$, $\mu(A_n) = 0$, $n \in \mathbb{N}$, then $\mu\left(\cup_{n\in\mathbb{N}} A_n\right) = 0$.
(Hint: Use Lemma 2.42.)

Let us now return to our original goal of defining a volume function on the Borel sets of $\mathbb{R}^d$. A natural idea is to approximate more general sets by boxes. One possible approach is to use approximations by finitely many boxes from the inside and from the outside, and define the volume of a set as a limit of these approximations, provided they coincide. This approach leads to the concept of the Jordan measure, which has many useful properties and applications, but it does not satisfy all the requirements we postulated at the beginning of our discussion, and hence we follow a slightly different approach below, due to Lebesgue. (Readers interested in the Jordan measure in more detail may consult Appendix A.)

The fruitful approach will turn out to be using approximations by boxes only from the outside, but in the same time allow coverings with a countably infinite number of boxes. This leads to the concept of the *outer Lebesgue measure* Vol^*, defined as

$$\mathrm{Vol}^*(A) := \inf\left\{\sum_{n\in\mathbb{N}} \mathrm{Vol}(B_n) : \ B_n \in \mathrm{Box}(\mathbb{R}^d),\, n \in \mathbb{N},\, A \subseteq \cup_{n\in\mathbb{N}} B_n\right\} \tag{2.12}$$

for any $A \subseteq \mathbb{R}^d$.

Theorem 2.44. The Lebesgue outer measure Vol^* has the following properties:

(i) It is a measure on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ in the sense of Definition 2.32.

(ii) It is an extension of Vol on $\mathrm{Box}(\mathbb{R}^d)$ in the sense that

$$\mathrm{Vol}^*(B) = \mathrm{Vol}(B) = \prod_{i=1}^{d}(b_i - a_i), \qquad B = \times_{i=1}^{d}[a_i, b_i) \in \mathrm{Box}(\mathbb{R}^d).$$

(iii) It is translation-invariant, i.e., for any $A \subseteq \mathbb{R}^d$ and $y \in \mathbb{R}^d$,

$$\mathrm{Vol}^*\left(\{x + y : \ x \in A\}\right) = \mathrm{Vol}^*(A).$$

(iv) For any linear transformation T on $\mathbb{R}^d$, we have

$$\mathrm{Vol}^*(T(A)) = |\det(T)| \cdot \mathrm{Vol}^*(A), \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

Definition 2.45. The restriction of Vol^* to $\mathcal{B}(\mathbb{R}^d)$ is called the *Lebesgue measure* on the Borel sets of $\mathbb{R}^d$, and is denoted by λ_d.

Remark 2.46. We will use the simpler notation λ for the Lebesgue measure if the dimension is obvious from the context or if it is irrelevant.

Remark 2.47. If $A \in \mathcal{B}(\mathbb{R}^d)$ is a Borel set then we can talk about the *Lebesgue measure on* A, which simply means that we change the set $\mathcal{X}$ from $\mathbb{R}^d$ to A, the σ-algebra from $\mathcal{B}(\mathbb{R}^d)$ to $\mathcal{B}(A) = \{B \in \mathcal{B}(\mathbb{R}^d) : B \subseteq A\}$, and define the measure of each $B \in \mathcal{B}(A)$ to be $\lambda_d(A)$.

Remark 2.48. Note that rotational invariance of λ_d follows as a special case of (iv) in Theorem 2.44, since for any rotation $\mathcal{R}$, we have $|\det \mathcal{R}| = 1$.

We omit the proof of the above theorem, as it is beyond the scope of these notes, and would not add much to the understanding of the rest of the material, anyway. We only mention that (i) and (ii) are special cases of the Carathéodory extension theorem, which gives a general method of extending measures from a collection of sets to their generated σ-algebra. The proof of the Carathéodory extension theorem is completely elementary, and fits into a few pages; we refer the interested reader to Sections 1.4 and 1.5 of [**?**], or Section 2.2 of [**?**]. The transformation property (iv) can be proved using the Fubini-Tonelli theorem about the interchangeability of the order of integrals; see Theorem 2.44 in [**?**]. Property (iii) is the simplest, and we leave it as an exercise:

Exercise 2.49. Prove the translation-invariance of the Lebesgue outer measure.

Note that Theorem 2.44 only claims that Vol_d^* is a measure on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$, and it is natural to ask whether a measure satisfying (ii)–(iv) can be defined on a larger σ-algebra, or in fact on the whole of $\mathcal{P}(\mathbb{R}^d)$. The answer to the first question is positive; it turns out that Vol^* is in fact a measure on a σ-algebra that is strictly larger than the Borel σ-algebra. This is called the *Lebesgue σ-algebra*, and its cardinality is the same as the cardinality of all the subsets of $\mathbb{R}^d$ (i.e., 2 to the power continuum), while the cardinality of the Borel σ-algebra is "only" continuum. However, every Lebesgue-measurable set differs from a Borel set only by a set of zero Lebesgue outer measure, i.e., any Lebesgue-measurable set is essentially a Borel-measurable set plus something negligible in the measure-theoretic sense.

On the other hand, it turns out that there exists no translation-invariant measure on $\mathbb{R}$ that coincides with the usual length on intervals, i.e., satisfies (ii) and (iii). This

can be shown by a very simple argument, (see , e.g., Section 1.1 in [**?**]), provided that we accept the axiom of choice. In fact, it can be shown that the non-existence of such a measure is equivalent to the axiom of choice.

Remark 2.50. Note that the Lebesgue measure is in some sense a "uniform distribution" on the real line, but it is not a probability measure, and cannot be normalized to one. However, the same example showing the impossibility to define a translation-invariant extension of the length to all subsets of $\mathbb{R}$ shows also that it is impossible to define a uniform distribution on all possible subsets of $[0,1]$.

We defined the Borel σ-algebra in a rather abstract way, and it is indeed not really possible to give a constructive description of a general Borel set. However, as we will see in the following exercises, Borel sets are in fact very simple in a measure-theoretic sense: namely, any Borel set can be arbitrarily well approximated by finite disjoint unions of boxes.

We start with the following simple reformulation of the definition of the outer Lebesgue measure, showing that we may restrict the coverings in 2.12 without loss of generality to countable collections of *disjoint* boxes.

Exercise 2.51. Show that the Lebesgue measure $\lambda(A)$ of any Borel set $A \subseteq \mathbb{R}^d$ can be written as

$$\lambda(A) = \inf\left\{ \sum_{n\in\mathbb{N}} \lambda(B_n) : \, B_n \in \mathrm{Box}(\mathbb{R}^d),\, n \in \mathbb{N},\, B_n \cap_{n\neq m} B_m = \emptyset, \quad A \subseteq \cup_{n\in\mathbb{N}} B_n \right\}. \tag{2.13}$$

Solution: Hidden.

Definition 2.52. The *symmetric difference* of two sets $A, B \subseteq \mathcal{X}$ is

$$A \triangle B := (A \setminus B) \cup (B \setminus A).$$

Exercise 2.53. Let $A \subseteq \mathbb{R}^d$ be a Borel set of finite Lebesgue measure. Show that for any $\varepsilon > 0$, there exist $n_\varepsilon \in \mathbb{N}$ and finitely many disjoint boxes $B_{\varepsilon,1}, \ldots, B_{\varepsilon,n_\varepsilon}$ such that

$$\lambda(A \triangle B_\varepsilon) < \varepsilon, \qquad \text{where} \qquad B_\varepsilon := \cup_{k=1}^{n_\varepsilon} B_{\varepsilon,k}. \tag{2.14}$$

(Hint: Use Exercise 2.51.)

Solution: Hidden.

Remark 2.54. In words, (2.14) means that the part of A that is not covered by the boxes, as well as the parts of the boxes that do not cover some part of A, have small measure.

It is easy to see that appoximation by finitely many boxes in the above sense may not be possible $\lambda(A) = +\infty$; a simple example is given by $A := \cup_{n\in\mathbb{N}}[2n, 2n+1)$. However, we still have the following:

Exercise 2.55. Show that for any Borel set $A \in \mathcal{B}(\mathbb{R}^d)$, and any $\varepsilon > 0$, there exist countably many disjoint boxes $(B_n)_{n\in\mathbb{N}} \subseteq \mathrm{Box}(\mathbb{R}^d)$ such that

$$A \subseteq \cup_{n\in\mathbb{N}} B_n, \quad \text{and} \quad \lambda\left((\cup_{n\in\mathbb{N}} B_n) \setminus A\right) < \varepsilon.$$

Solution: Hidden.

Remark 2.56. See Exercises **??** and **??** for a generalizations of Exercises 2.53 and 2.55.

Sets of zero measure play an important role in the theory of integrals. The following is immediate from the definition Vol*:

Lemma 2.57. A set $A \in \mathcal{B}(\mathbb{R}^d)$ has zero outer Lebesgue measure if and only if for every $\varepsilon > 0$ there exists a sequence of boxes $B_k \in \mathrm{Box}(\mathbb{R}^{d_i})$, $k \in \mathbb{N}$, such that $A_i \subseteq \cup_{k\in\mathbb{N}} B_k$, and $\sum_{k\in\mathbb{N}} \lambda(B_k) < \varepsilon$.

Exercise 2.58. Use the definition of the Lebesgue measure (but not Theorem 2.59 or 2.62 below) to show that if $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$, $i \in [n]$, and there exists an i such that $\lambda(A_i) = 0$ then $\lambda(A_1 \times \ldots \times A_n) = 0$.

Solution: Hidden.

Finally, we also mention without proof that the product property in (ii) of Theorem 2.44 holds more generally, not only for the product of boxes, but of arbitrary Borel sets:

Theorem 2.59. For any $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$, $i \in [n]$,

$$\lambda_{d_1+\ldots+d_n}(A_1 \times \ldots \times A_n) = \lambda_{d_1}(A_1) \cdot \ldots \cdot \lambda_{d_n}(A_n).$$

The above can be reformulated using the notion of the product measure:

Definition 2.60. Let $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$, $i \in [n]$, be measure spaces. We say that a measure μ on $\otimes_{i=1}^n \mathcal{A}_i = \sigma\left(\{A_1 \times \ldots \times A_n : A_i \in \mathcal{A}_i,\, i \in [n]\}\right)$ *factorizes to the product of the* μ_i if

$$\mu(A_1 \times \ldots \times A_n) = \mu_1(A_1) \cdot \ldots \cdot \mu_n(A_n), \qquad A_i \in \mathcal{A}_i,\, i \in [n].$$

If there exists a unique such measure on $\otimes_{i=1}^n \mathcal{A}_i$ then we call it the *product* of the μ_i, and denote it by $\otimes_{i=1}^n \mu_i$.

The following is also a consequence of the general Carathéodory extension theorem:

Theorem 2.61. For any measure spaces $(\mathcal{X}_i, \mathcal{A}_i, \mu)$, $i \in [n]$, there exists a measure μ on $\otimes_{i=1}^n \mathcal{A}_i$ that factorizes to the product of the μ_i. If all μ_i are σ-finite then there is a exactly one such measure, i.e., $\otimes_{i=1}^n \mu_i$ exists.

The factorization property of the Lebesgue measure in Theorem 2.59 can be expressed in the following stronger form:

Theorem 2.62. For any $d_1, \ldots, d_n \in \mathbb{N}$,

$$\lambda_{d_1+\ldots+d_n} = \otimes_{i=1}^n \lambda_{d_i}.$$

2.4 Classical models of physical systems

We are now ready to define classical probabilistic models of physical systems.

Any such model is specified by a measurable space $(\Omega, \mathcal{F})$, where Ω is the set of *elementary events*, also called the set of *physical states* or *phase space* in physics. Ω may be a finite set, e.g., for modeling the roll of a die, we may choose $\Omega = [6] = \{1, 2, \ldots, 6\}$, or for the modeling of the toss of a coin, $\Omega = \{\text{heads}, \text{tails}\}$. It may also be countably infinite; e.g., when modeling the random walk of a particle on a d-dimensional square lattice, we would choose $\Omega = \mathbb{Z}^d$. In all of these examples, the natural choice for the σ-algebra $\mathcal{F}$ is the trivial $\mathcal{F} = \mathcal{P}(\Omega)$, i.e., all subsets of Ω are measurable. A more involved example is the description of the continuous-time random walk of a particle in a d-dimensional space, in which case the natural choice for the phase space is $\Omega = \mathbb{R}^d$, and the usual choice for $\mathcal{F}$ is the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$.

Possible states of the system are described by probability measures on $\mathcal{F}$, and we call the set of all probability measures on $\mathcal{F}$ the *state space* of the model, and denote it by $\mathcal{S}(\Omega, \mathcal{F})$. In this picture, a Dirac measure concentrated at a point ω represents a well-defined physical state of the system, while more general probability measures model our uncertainty about the physical state of the system. For instance, consider the task of describing the number on the top of a die after it has been rolled, but is covered by a cup so that we cannot see it. The die may have a well-defined physical state, i.e., one of its sides being on the top, but this is inaccessible to us. Hence, our best description of the state of the system (i.e., the number on the top of the die) may be some probability distribution on the possible physical states. For instance, if we expect the die to be fair, then this would be the uniform distribution on [6]. However, if we know that there is a small piece of lead inside the die that changes its chances of falling on some of its faces, then we may instead describe its state by a

non-uniform probability distribution, depending on the size and location of the piece of lead.

As another example, one may imagine a particle moving on a one-dimensional lattice described by $\Omega = \mathbb{Z}$, at each time instance moving one step to the left with probability p, and one step to the right with probability $1-p$. Even if we knew that the particle was at the origin at time $t = 0$, if we let the particle wander for, say, 3 time steps without looking at it, our best description of the particle's position will be a probability distribution on $\mathbb{Z}$ that does not correspond to a physical state, i.e., a Dirac distribution concentrated at some node $k \in \mathbb{Z}$.

Exercise 2.63. Calculate the probability distribution describing the particle's position after 3 time steps in the above example.

Remark 2.64. We use the terms "proability measure" and "probability distribution" more or less as as synonyms, and while some distinctions could be made between the two, there are no general rules for it. In principle, "probability measure" is the general notion, as introduced in the previous section. One context where "distribution" is preferred is when one talks about the probability measure induced by a random variable (or push-forward measure); see the next section. Another case where "probability distribution" may be used is when a probability measure is specified by a function on $\mathcal{X}$ instead of defining it as a function on the subsets of $\mathcal{X}$.

For instance, in the above two examples (the roll of a die, and the random walk on $\mathbb{Z}$), the natural σ-algebra to work with is $\mathcal{F} = \mathcal{P}(\mathcal{X})$ (where $\mathcal{X} = [6]$ or $\mathcal{X} = \mathbb{Z}$), and any probability measure ϱ on $\mathcal{P}(\mathcal{X})$ determines a function $\hat{\varrho} : \mathcal{X} \to [0,1]$ by $\hat{\varrho}(x) := \varrho(\{x\})$, such that

$$\sum_{x\in\mathcal{X}} \hat{\varrho}(x) = 1, \qquad \text{and} \qquad \varrho(A) = \sum_{x\in A} \hat{\varrho}(x), \quad A \in \mathcal{P}(\mathcal{X}).$$

Vice versa, any function $\hat{\varrho} : \mathcal{X} \to [0,1]$ with the property $\sum_{x\in\mathcal{X}} \hat{\varrho}(x) = 1$ determines a probability measure ϱ on $\mathcal{P}(\mathcal{X})$ via $\varrho(A) := \sum_{x\in A} \hat{\varrho}(x)$, $A \in \mathcal{P}(\mathcal{X})$. Such a function ϱ may be called a "probability density function", or a "weight function". We remark that in this case the correspondence between probability measures (functions on subsets with certain properties) and probability density functions (functions on points with certain properties) is one-to-one, but this is so only when the σ-algebra is the full power set. Probability measures may also be identified with density functions also in the general case, but this correspondence is more limited and more complicated; we will briefly touch upon this in the next section.

Finally, yet another example of defining a probability measure by a function is by a so-called cumulative distribution function, mainly for real-valued random variables, which is very common in probability theory, but we are not going to use this concept in these notes.

To complete the mathematical description of classical models, we need to find a description of measurements on the system, and specify a rule describing the probabilities of the various possible measurement outcomes when the measurement is performed in a given state of the system. For instance, imagine that in the above example, the particle and the origin are connected by a rubber band that stretches more and more as the particle gets further and further away from the origin, storing an amount of energy $E(k) = ck^2$ if the particle is at position k. What will we find if we measure the energy of the particle after the particle has wandered for three steps without us observing its position? If ϱ is the probability distribution describing its position, then we will observe an energy value

$$\begin{aligned} 0 \quad &\text{with probability} \quad \varrho(0), \\ c \quad &\text{with probability} \quad \varrho(1) + \varrho(-1) = \varrho(\{k : \, E(k) = c\}, \\ 4c \quad &\text{with probability} \quad \varrho(2) + \varrho(-2) = \varrho(\{k : \, E(k) = 4c\}, \end{aligned}$$

etc. If we are only interested in whether the energy is above a certain threshold E_0, then the probability of this can be computed by

$$\sum_{k \in \mathbb{Z} \colon ck^2 \geq E_0} \varrho(k) = \varrho\left(\{k \in \mathbb{Z} : \, ck^2 \geq E_0\}\right).$$

This suggests that physical quantities should be described by functions on the phase space Ω, and if the system is in state ϱ, then the probability that the result of the measurement of the physical quantity described by f falls in a set A should be computed by the formula

$$\mathbb{P}_{\varrho,f}(A) := \varrho\left(\{\omega \in \Omega : \, f(\omega) \in A\}\right). \tag{2.15}$$

While physical quantities like energy, position, momentum, etc., are usually described by real-valued functions, we may consider more general physical quantities, e.g., when rolling two dice together, our quantity of interest may be the parity of the sum on their top faces, that is modeled by a function $f : [6] \times [6] \to \{\text{odd}, \text{even}\}$. In the most general scenario, a physical quantity may be any function $f : \Omega \to \mathcal{X}$, where $\mathcal{X}$ is some arbitrary set.

Note that, in order for the RHS above to make sense,

$$\{\omega \in \Omega : \, f(\omega) \in A\} =: f^{-1}(A) \tag{2.16}$$

should be a measurable set in our model, i.e., an element of $\mathcal{F}$. (Here, f^{-1} does not denote the inverse (we do not assume f to be invertible), but the inverse image or preimage function that maps from $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\Omega)$.) This may be too much to require for every subset of $\mathcal{X}$, and it also does not seem necessary from a physical point

of view. Indeed, while we may be interested in whether the value of some physical quantity falls into an interval, it makes no physical sense to ask if its value falls into a set that we cannot describe, but only prove its existence using the axiom of choice, as we have seen in the previous section.

Thus, we will always assume that the set $\mathcal{X}$ of possible values of some physical quantity is also equipped with a collection of subsets, for which we may want to know the probability of the measurement outcome falling into any one of those subsets, and for the elements of which $f^{-1}(A) \in \mathcal{F}$ has to hold, in order to be able to compute probabilities as given in (2.16). As we will see in the next section, we may assume without loss of generality that this collection of subsets of $\mathcal{X}$ forms a σ-algebra. We will elaborate further on these concepts in the next section.

2.5 Measurable functions

In this section we discuss the concept of measurability of a function between two measurable spaces. This is motivated by the considerations in Section 2.4 related to building (classical) probabilistic models of physical phenomena, but it is also instrumental in building a theory of integrals that will allow us to define the function spaces that will play a central role in building quantum models.

As we have already mentioned in the previous section, a function $f : \mathcal{X} \to \mathcal{Y}$ induces a map $f^{-1} : \mathcal{P}(\mathcal{Y}) \to \mathcal{P}(\mathcal{X})$, where f^{-1} here does not denote the inverse (we do not assume f to be invertible), but the *inverse image* or *preimage* function, defined as

$$f^{-1}(B) := \{x \in \mathcal{X} : f(x) \in B\}, \qquad B \in \mathcal{P}(\mathcal{Y}).$$

Exercise 2.65. Show that the preimage function is compatible with the set operations in the sense that

$$\begin{aligned} &(i) \quad f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2), \\ &(ii) \quad f^{-1}\left(\cup_{i\in\mathcal{I}} B_i\right) = \cup_{i\in\mathcal{I}} f^{-1}(B_i), \\ &(iii) \quad f^{-1}\left(\cap_{i\in\mathcal{I}} B_i\right) = \cap_{i\in\mathcal{I}} f^{-1}(B_i), \end{aligned}$$

where $\mathcal{I}$ is an arbitary index set, and $B_1, B_2, B_i \in \mathcal{P}(\mathcal{Y})$.

Definition 2.66. Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces. We say that a function $f : \mathcal{X} \to \mathcal{Y}$ is *measurable* if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$, i.e., the preimage of all measurable subsets in $\mathcal{Y}$ is a measurable subset in $\mathcal{X}$.

Remark 2.67. In probability theory, measurable functions are called *random variables*, in statistics they are called *tests*.

Definition 2.68. In a classical probabilistic model $(\Omega, \mathcal{F})$ of a physical system, measurable functions $f : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ will be called $\mathcal{X}$-valued *sharp measurements.*

Remark 2.69. As the terminology suggests, one may also consider unsharp measurement; we will discuss these later.

Remark 2.70. Note that the concept of measurability of a function $f : \mathcal{X} \to \mathcal{Y}$ depends on the σ-algebras given on $\mathcal{X}$ and $\mathcal{Y}$. When we want to emphasize this, we may write the function as $f : (\mathcal{X}, A) \to (\mathcal{Y}, \mathcal{B})$, which still means that the function maps from $\mathcal{X}$ to $\mathcal{Y}$, but it also indicates that measurability is defined with respect to the σ-algebras $\mathcal{A}$ and $\mathcal{B}$. Alternatively, we may say that the function is $(\mathcal{A}, \mathcal{B})$-measurable.

When one of the σ-algebras are canonical (e.g., the Borel σ-algebra on $\mathbb{R}$, or $\mathcal{P}(\mathcal{X})$ for a finite $\mathcal{X}$) then we may omit that from the notation, and write $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{R}$, for instance.

Likewise, by measurability of a $\mathbb{K}$-valued function defined on a Borel subset A of $\mathbb{R}^d$, by measurability we always mean measurability with respect to the Borel σ-algebras $(\mathcal{B}(A), \mathcal{B}(\mathbb{K}))$, unless otherwise stated.

Exercise 2.71. Show that every function $f : \mathcal{X} \to \mathcal{Y}$ is measurable if $\mathcal{Y}$ is equipped with the smallest σ-algebra $\mathcal{B} = \{\emptyset, \mathcal{Y}\}$, or $\mathcal{X}$ is equipped with the largest σ-algebra $\mathcal{A} = \mathcal{P}(\mathcal{X})$.

The notion of measurability plays an analogous role in measure theory to the notion of continuity in topology. Recall that a set $G \subseteq \mathbb{K}^d$ is *open*, if for every $x \in G$ there exists an $\varepsilon_x > 0$ such that the ε_x-ball $B(x, \varepsilon_x) := \{y \in \mathbb{K}^d : \|y - x\| < \varepsilon\} \subseteq G$. The topology $\tau_{\mathbb{K}^d}$ of $\mathbb{K}^d$ is, by definition, the collection of all open sets in $\mathbb{K}^d$. The *relative topology* of a subset $A \subseteq \mathbb{K}^d$ is $\tau_A := \{G \cap A : G \in \tau_{\mathbb{R}^d}\}$.

Exercise 2.72. Let $A \subseteq \mathbb{K}^d$, and $f : A \to \mathbb{K}^m$ be a function.

(i) Show that f is continuous on A if and only if for all $B \in \tau_{\mathbb{R}^m}$, $f^{-1}(B) \in \tau_A$.

(ii) Show that if f is continuous and A is a Borel set then f is $(\mathcal{B}(A), \mathcal{B}(\mathbb{K}^d))$-measurable.

Remark 2.73. It is easy to see that the above holds true if $\mathbb{K}^d$ and $\mathbb{K}^m$ are replaced with arbitrary topological spaces.

Proposition 2.74. Every measurable function $f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ defines a map $f_* : \mathcal{M}(\mathcal{X}, \mathcal{A}) \to \mathcal{M}(\mathcal{Y}, \mathcal{B})$ by

$$f_*\mu := \mu \circ f^{-1}.$$

This map is positive linear, i.e., for every $\mu_1, \mu_2 \in \mathcal{M}(\mathcal{X}, \mathcal{A})$, $c_1, c_2 \in [0, +\infty]$,

$$f_*(c_1\mu_1 + c_2\mu_2) = c_1 f_*\mu_1 + c_2 f_*\mu_2,$$

and it maps probability measures into probability measures.

Proof. Trivial from the definition, exercise. □

Definition 2.75. Given a state $\varrho \in \mathcal{S}(\Omega, \mathcal{F})$ in a classical probabilistic model, and a sharp measurement $f : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$, we call

$$\mathbb{P}_{\varrho,f} := f_*\varrho \in \mathcal{S}(\mathcal{X}, \mathcal{A}) \tag{2.17}$$

the *distribution of measurement outcomes*, or *post-measurement distribution* defined by the state ϱ and the measurement f.

Remark 2.76. Note that (2.17) is the same as (2.15). The probability $\mathbb{P}_{\varrho,f}(A)$ should give a good approximation of the frequency with which the outcome of the measurement falls in the set A when the same measurement f is performed independently on many identical copies of the system, all prepared in state ϱ.

Remark 2.77. In measure theory, $f_*\mu$ is called the *push-forward* of the measure μ by the measurable function f. In probability theory, $\mathbb{P}_{\varrho,f}$ is called the *distribution* of the random variable f under the probability measure ϱ.

We will most often consider real-valued measurements, i.e., real-valued measurable functions, while in defining the relevant function spaces for quantum theory, we will need to work with complex-valued functions. Hence, we study measurability in these cases in more detail below.

We start with the following simple observation, that shows that it is enough to verify measurability on a generating system of the image σ-algebra.

Lemma 2.78. Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces, and $\mathcal{B}_0 \subseteq \mathcal{B}$ be a generator for $\mathcal{B}$, i.e., $\sigma(\mathcal{B}_0) = \mathcal{B}$. Then $f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ is measurable if and only if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}_0$.

Proof. Let $\widehat{\mathcal{B}} := \{B \subseteq \mathcal{Y} : f^{-1}(B) \in \mathcal{A}\}$. Then it is easy to verify that $\widehat{\mathcal{B}}$ is a σ-algebra. By assumption, it contains $\mathcal{B}_0$, and hence it also contains $\sigma(\mathcal{B}_0) = \mathcal{B}$. □

For an extended real-valued function $f : X \to \overline{\mathbb{R}}$, let us introduce the notations

$$\{f \geq c\} := \{x \in \mathcal{X} : f(x) \geq c\}, \qquad \{f > c\} := \{x \in \mathcal{X} : f(x) > c\},$$
$$\{f \leq c\} := \{x \in \mathcal{X} : f(x) \leq c\}, \qquad \{f < c\} := \{x \in \mathcal{X} : f(x) < c\},$$

for every $c \in \mathbb{R}$.

Remark 2.79. Note that every real-valued function is also an extended real-valued function, and hence in what follows, we always consider the more general case of extended real-valued functions, whenever possible.

By Exercise 2.18 and Lemma 2.78, we immediately have the following:

Corollary 2.80. A real-valued or extended real-valued function f on $(\mathcal{X}, \mathcal{A})$ is measurable if and only if any (and hence all) of the following holds:

$$(i)\ \ \{f \geq c\} \in \mathcal{A},\ \ c \in \mathbb{R}, \qquad (ii)\ \ \{f > c\} \in \mathcal{A},\ \ c \in \mathbb{R},$$
$$(iiii)\ \ \{f \leq c\} \in \mathcal{A},\ \ c \in \mathbb{R}, \qquad (iv)\ \ \{f < c\} \in \mathcal{A},\ \ c \in \mathbb{R}.$$

Exercise 2.81. Let $f_i : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}$, $i \in \mathcal{I}$ be functions, where $\mathcal{I}$ is a finite or countably infinite index set. Show that if all f_i is measurable, then so is $\inf_i f_i$ and $\sup_i f_i$, where the infimum and the supremum are taken in the pointwise sense, i.e., $(\inf_i f_i)(x) := \inf_i f_i(x)$, $x \in \mathcal{X}$, and similarly for the supremum.

Solution: Hidden.For every $c \in \mathbb{R}$,

$$\{\inf_i f_i \geq c\} = \cap_{i\in\mathcal{I}}\{f_i \geq c\}, \qquad \{\sup_i f_i \leq c\} = \cap_{i\in\mathcal{I}}\{f_i \leq c\},$$

and the assertion follows from Corollary 2.80 and the fact that a σ-algebra is closed under countable unions and products.

Exercise 2.82. Show that an extended real-valued function $f : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}$ is measurable if and only if

$$f_+ := \max\{0, f\} \qquad \text{and} \qquad f_- := -\min\{0, f\}$$

are also measurable.

Definition 2.83. For an extended real-valued sequence $(a_n)_{n\in\mathbb{N}} \subset \overline{\mathbb{R}}$, let

$$\liminf_n a_n := \sup_{n\in\mathbb{N}} \inf_{k\geq n} a_k, \qquad \limsup_n a_n := \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k,$$

be the *limit inferior* and the *limit superior* of the sequence, respectively. For a sequence of functions $f_n : \mathcal{X} \to \overline{\mathbb{R}}$, we define

$$(\liminf_n f_n)(x) := \liminf_n f_n(x), \qquad x \in \mathcal{X},$$
$$(\limsup_n f_n)(x) := \limsup_n f_n(x), \qquad x \in \mathcal{X}.$$

Remark 2.84. It is easy to see that $\liminf_n a_n$ is the smallest, and $\limsup_n a_n$ is the largest accumulation point of the sequence $(a_n)_{n\in\mathbb{N}}$. In particular, the sequence has a limit (possibly $\pm\infty$) if and only if $\liminf_n a_n = \limsup_n a_n$, and in this case this common value is equal to $\lim_n a_n$. We leave the verification of this as an exercise.

Exercise 2.81 with the above definition immediately implies that the following is true:

Exercise 2.85. Let $f_n : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be a sequence of measurable functions. Show that $\liminf_n f_n$ and $\limsup_n f_n$ are both measurable. In particular, if the sequence of functions is pointwise convergent, then the limit function $f(x) := \lim_n f_n(x)$ is also measurable.

Lemma 2.86. A complex-valued function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ is measurable if and only if $\mathfrak{Re}\, f$ and $\mathfrak{Im}\, f$ are measurable.

Proof. The Borel σ-algebra of $\mathbb{C}$ is generated by boxes of the form $J_1 + iJ_2 := \{a+ib : a \in J_1,\, b \in J_2\}$, where $J_1, J_2 \subseteq \mathbb{R}$ are intervals. By Lemma 2.78, f is measurable if and only if $\mathcal{A} \ni \{x \in \mathcal{X} : f(x) \in J_1 + iJ_2\} = (\mathfrak{Re}\, f)^{-1}(J_1) \cap (\mathfrak{Im}\, f)^{-1}(J_2)$. Thus, if both $\mathfrak{Re}\, f$ and $\mathfrak{Im}\, f$ are measurable then so if f. Vice versa, if f is measurable then taking $J_2 := \mathbb{R}$ shows that $(\mathfrak{Re}\, f)^{-1}(J_1) = (\mathfrak{Re}\, f)^{-1}(J_1) \cap (\mathfrak{Im}\, f)^{-1}(\mathbb{R}) \in \mathcal{A}$, for any interval $J_1 \subseteq \mathbb{R}$, and thus $\mathfrak{Re}\, f$ is measurable. The measurability of $\mathfrak{Im}\, f$ follows the same way, by taking $J_1 = \mathbb{R}$. □

Just as a general measurable set may be difficult to describe, so is a general measurable function. However, we can approximate every extended real-valued or complex-valued measurable function by simpler measurable functions, as we show below.

Definition 2.87. For a set $A \subseteq \mathcal{X}$, its *characteristic function* (or *indicator function*) $\mathbf{1}_A$ is defined as

$$\mathbf{1}_A(x) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Exercise 2.88. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space. Show that $A \subseteq \mathcal{X}$ is measurable if and only if its characteristic function $\mathbf{1}_A : \mathcal{X} \to \mathbb{R}$ is measurable.

Solution: The assertion follows immediately from the fact that

$$\{\mathbf{1}_A \geq c\} = \begin{cases} \emptyset \in \mathcal{A}, & c > 1, \\ A, & c \leq 1. \end{cases}$$

Definition 2.89. A function $f : \mathcal{X} \to \mathcal{Y}$ is called a *simple function* if it only takes finitely many different values, i.e., $|\operatorname{ran} f| < +\infty$.

Remark 2.90. Recall the previously introduced convention

$$(\pm\infty) \cdot 0 := 0.$$

Exercise 2.91. Let f be an extended real-valued or complex-valued simple function on $(\mathcal{X}, \mathcal{A})$.

(i) Show that f is simple if and only if it can be written as $f = \sum_{i=1}^{r} c_i \mathbf{1}_{A_i}$, where $A_i \cap_{i \neq j} A_j = \emptyset$, and $c_i \in \overline{\mathbb{R}}$ (extended real-valued) or $c_i \in \mathbb{C}$ (complex), $i \in [r]$.

(ii) Show that f is a measurable simple function if and only if all A_i in the above decomposition are measurable.

Proposition 2.92. Let f be an extended real-valued or complex-valued function on $(\mathcal{X}, \mathcal{A})$. Then f is measurable if and only if it is the pointwise limit of a sequence of measurable simple functions. Moreover, an approximating sequence $(f_n)_{n \in \mathbb{N}}$ of such functions can be taken so that $|f_n(x)| \leq |f(x)|$ for all $x \in \mathcal{X}$ and $n \in \mathbb{N}$, and if f is bounded then the convergence is uniform in x.

Proof. We have seen that the limit of measurable functions is again measurable, so we have to prove the converse direction. Let f be a non-negative extended real-valued measurable function, and for every $n \in \mathbb{N}$, define

$$f_n := \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mathbf{1}_{f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right)}.$$

Then clearly f_n is measurable if f is measurable, $f_n(x) \leq f(x)$ for all $x \in \mathcal{X}$, and $|f_n(x) - f(x)| \leq \frac{1}{2^n}$ for all $x \in \mathcal{X}$ such that $|f(x)| < n$, showing that $\lim_n f_n(x) = f(x)$ for all $x \in \mathcal{X}$ and that the convergence is uniform if f is bounded.

For an extended real-valued function, choose separate approximations for the positive part $f_+ := \max\{0, f\}$ and for the negative part $f_- := -\min\{0, f\}$, and for a complex-valued function approximate separately $\mathfrak{Re}\, f$ and $\mathfrak{Im}\, f$. □

Use the above proposition to prove the following properties of measurable functions:

Exercise 2.93. Let $f, g : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}$ be measurable functions, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Show that fg is measurable, for any $\lambda, \eta \in \mathbb{K}$, $\lambda f + \eta g$ is measurable, and if $g(x) \neq 0$ for all x then f/g is measurable, too.

Remark 2.94. Note that the above algebraic operations on functions that preserve measurablity also preserve continuity, another analogy between measure theory and topology.

Exercise 2.95. Let $f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ and $g : (\mathcal{Y}, \mathcal{B}) \to (\mathcal{Z}, \mathcal{C})$ be measurable. Show that $g \circ f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Z}, \mathcal{C})$ is measurable.

Conclude that if $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}^d$ is measurable and $g : \operatorname{ran} f \to \mathbb{K}^m$ is continuous then $g \circ f$ is measurable.

Definition 2.96. For functions $f_i \in \mathbb{K}^{\mathcal{X}_i}$, $i \in [n]$, their *tensor product* is the n-variable function

$$(f_1 \otimes \ldots \otimes f_n)(x_1, \ldots, x_n) := f_1(x_1) \cdot \ldots \cdot f_n(x_n), \qquad x_i \in \mathcal{X}_i,\ i \in [n].$$

Exercise 2.97. Show that if $f_i : (\mathcal{X}_i, \mathcal{A}_i) \to \mathbb{K}$, $i \in [n]$, are measurable then $f_1 \otimes \ldots \otimes f_n$ is measurable, where $\times_{i=1}^n \mathcal{X}_i$ is equipped with the product σ-algebra $\otimes_{i=1}^n \mathcal{A}_i = \sigma\left(\{A_1 \times \ldots \times A_n : A_i \in \mathcal{A}_i,\ i \in [n]\}\right)$.

2.6 Integral

Imagine that we conduct the same real-valued sharp measurement $f : (\Omega, \mathcal{F}) \to \mathbb{R}$ on many identical copies of a system, all prepared in the state $\varrho \in \mathcal{S}(\Omega, \mathcal{F})$. The average value and the variance of the measurement outcomes are two important characteristics of the measurement. It is intuitively clear that the values of these quantities should be predicted by our model to be

$$\mathbb{E}_\varrho(f) := \int_{\mathbb{R}} t \, d\,\mathbb{P}_{\varrho,f}(t) = \int_{\mathbb{R}} t \, d(\varrho \circ f^{-1})(t) = \int_\Omega f(\omega)\, d\varrho(\omega), \tag{2.18}$$

$$\mathbb{V}_\varrho(f) := \int_{\mathbb{R}} (t - \mathbb{E}_\varrho(f))^2 \, d\,\mathbb{P}_{\varrho,f}(t) = \int_\Omega (f(\omega) - \mathbb{E}_\varrho(f))^2 \, d\varrho(\omega). \tag{2.19}$$

Note, however, that the integrals above are not Riemann integrals in general, and hence their interpretation and the way to compute them is not covered by usual calculus. Below we discuss how to make sense of these expressions, and why the identities above hold.

Hence, our goal now is to develop a concept of integral that a) reduces to the usual Riemann integral for continuous real-valued functions on a compact interval (or, more generally, on a compact box in $\mathbb{R}^d$), and b) it is compatible with the general view that the integral measures the signed area (or volume) below the graph of a function. This latter leads to the following:

Definition 2.98. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and $f : \mathcal{X} \to \overline{\mathbb{R}}_+$ be a non-negative simple measurable function, given as $f = \sum_{i=1}^r c_i \mathbf{1}_{A_i}$, with all $A_i \in \mathcal{A}$ and $c_i \in [0, +\infty]$ (see Exercise 2.91). Then the *integral* of f with respect to the measure μ (or the μ-integral of f) is defined as

$$\int f \, d\mu := \int_{\mathcal{X}} f \, d\mu := \int_{\mathcal{X}} f(x)\, d\mu(x) := \sum_{i=1}^r c_i \mu(A_i).$$

For a measurable set $A \in \mathcal{A}$, we define

$$\int_A f \, d\mu := \int_{\mathcal{X}} f \mathbf{1}_A \, d\mu.$$

For convenience, let us introduce the notation

$$\mathcal{L}_{\text{simp}}(\mathcal{X}, \mathcal{A}, \overline{\mathbb{R}}_+) := \{f : \mathcal{X} \to \overline{\mathbb{R}}_+ : f \text{ simple and measurable}\}.$$

The following properties of the integral are easy to verify, and hence we leave them as an exercise:

Exercise 2.99. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space. The integral of non-negative simple measurable functions have the following properties:

(i) *positive linearity:* For any $f_1, f_2 \in \mathcal{L}_{\text{simp}}(\mathcal{X}, \mathcal{A}, \overline{\mathbb{R}}_+)$, any $c_1, c_2 \in [0, +\infty]$, and any $A \in \mathcal{A}$,

$$\int_A (c_1 f_1 + c_2 f_2)\, d\mu = c_1 \int_A f_1\, d\mu + c_2 \int_A f_2\, d\mu;$$

(ii) *monotonicity:* For any $f, g \in \mathcal{L}_{\text{simp}}(\mathcal{X}, \mathcal{A}, \overline{\mathbb{R}}_+)$, and any $A \in \mathcal{A}$,

$$f \le g \implies \int_A f\, d\mu \le \int_A g\, d\mu.$$

Next, we want to extend the concept of integral to more complicated functions. We start with non-negative measurable functions:

Definition 2.100. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, and μ be a measure on $\mathcal{A}$. For a non-negative measurable function $f : \mathcal{X} \to \overline{\mathbb{R}}_+$, the *integral* of f with respect to μ (or the μ-integral of f) is defined as

$$\int f\, d\mu := \sup \left\{ \int h\, d\mu : h \in \mathcal{L}_{\text{simp}}(\mathcal{X}, \mathcal{A}, \overline{\mathbb{R}}_+),\, h \le f \right\}.$$

Definition 2.101. When $\mathcal{X} = A$ is a Borel set in $\mathbb{R}^d$, $\mathcal{A} = \{A \cap B : B \in \mathcal{B}(\mathbb{R}^d)\}$ are the Borel sets in A, and $\mu = \lambda$ is the Lebesgue measure, then λ-integral of a Borel measurable function $f : A \to \overline{\mathbb{R}}$ is called the *Lebesgue integral* of f.

Exercise 2.102. Show that the above definition is consistent with the previous one, in the sense that if f is a simple measurable function then its integrals given in Definition 2.98 and in Definition 2.100 are the same.

Let us now recall that for a real-valued function $f : [a, b] \to \mathbb{R}$ on an interval $[a, b] \subseteq \mathbb{R}$, its *Riemann integral* is defined in the following way: For every partition

$\mathcal{P} : a = a_0 < a_1 < \ldots < a_n = b$ of the interval, one considers the lower and upper Riemann sums

$$I_*(f, \mathcal{P}) := \sum_{i=0}^{n-1} \left(\inf_{x \in [a_i, a_{i+1})} f(x) \right) (a_{i+1} - a_i),$$
$$I^*(f, \mathcal{P}) := \sum_{i=0}^{n-1} \left(\sup_{x \in [a_i, a_{i+1})} f(x) \right) (a_{i+1} - a_i).$$

Clearly, $I_*(f) \leq I^*(f)$ for any partition. The function f is called *Riemann integrable*, if

$$\sup_{\mathcal{P}} I_*(f, \mathcal{P}) = \inf_{\mathcal{P}} I^*(f, \mathcal{P}),$$

where the supremum and the infimum are over all partitions of $[a, b]$, and this common value is called the *Riemann integral* of f, denoted by $\int_a^b f(x)\, dx$.

Now, let $f : [a, b] \to \overline{\mathbb{R}}_+$ be a non-negative extended real-valued function. For any partition $\mathcal{P} : a = a_0 < a_1 < \ldots < a_n = b$, we have

$$\sum_{i=0}^{n-1} \left(\inf_{x \in [a_i, a_{i+1})} f(x) \right) \mathbf{1}_{[a_i, a_{i+1})} \leq f \leq \sum_{i=0}^{n-1} \left(\sup_{x \in [a_i, a_{i+1})} f(x) \right) \mathbf{1}_{[a_i, a_{i+1}]}.$$

By the monotonicity of the Lebesgue integral and Exercise 2.102, we have

$$I_*(f, \mathcal{P}) \leq \int_{[a,b]} f \, d\lambda \leq I^*(f, \mathcal{P}).$$

Hence, if f is Riemann integrable then

$$\int_a^b f(x)\, dx = \int_{[a,b]} f \, d\lambda.$$

On the other hand, consider $f : [0, 1] \to \mathbb{R}_+$, $f := \mathbf{1}_{\mathbb{Q} \cap [0,1]}$. Since every interval $[a_i, a_{i+1})$ in a partition $\mathcal{P}$ of $[0, 1]$ contains both a rational and an irrational number,

$$I_*(\mathbf{1}_{\mathbb{Q} \cap [0,1]}, \mathcal{P}) = 0, \qquad I^*(\mathbf{1}_{\mathbb{Q} \cap [0,1]}, \mathcal{P}) = 1$$

for every partition, and hence $\mathbf{1}_{\mathbb{Q} \cap [0,1]}$ is not Riemann integrable on $[0, 1]$. On the on the other hand, it is a non-negative measurable function, and hence it has a Lebesgue integral according to Definition 2.100.

It is also not too difficult to show that every Riemann integrable function is measurable, and hence we can conclude that the Lebesgue integral is a proper extension of the Riemann integral for functions on a compact interval.

Exercise 2.103. Compute the Lebesgue integral $\int_{[0,1]} \mathbf{1}_{\mathbb{Q}\cap[0,1]}\, d\lambda$ based on Definition 2.100.

The following important inequalities in probability theory follow immediately from the above definition of the integral:

Exercise 2.104. Let f be a $\mathbb{K}$-valued measurable function on a measure space $(\mathcal{X}, \mathcal{A}, \mu)$, and $g : \mathbb{R}_+ \to \mathbb{R}_+$ be monotone. Show that for any $c \in \mathbb{R}_+$ such that $g(c) > 0$,

$$\mu\left(\{|f| \geq c\}\right) \leq \frac{1}{g(c)} \int_{\mathcal{X}} (g \circ |f|)\, d\mu. \qquad \text{(Genralized Markov inequality)} \tag{2.20}$$

Remark 2.105. When $(\mathcal{X}, \mathcal{A}, \mu) = (\Omega, \mathcal{F}, P)$ for some probability space, and $X : \Omega \to \mathbb{R}$ is a real-valued random variable, then (2.20) yields

$$P\left(\{|X| \geq c\}\right) \leq \frac{1}{c}\, \mathbb{E}_P(|X|), \qquad \text{(Markov inequality)}, \tag{2.21}$$

$$P\left(\{X \geq c\}\right) \leq \inf_{t>0} e^{-tc}\, \mathbb{E}_P(e^{tX}), \qquad \text{(exponential Markov inequality)}. \tag{2.22}$$

In the above, $\mathbb{E}_P(e^{tX})$ is called the *moment generating function* of X.

The following simple concept is crucial in all applications of measure theory:

Definition 2.106. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space. We say that some statement about the points of $\mathcal{X}$ holds *μ-almost everywhere* (μ-a.e. in short), if those points of $\mathcal{X}$ where it does not hold form a measurable set of measure 0. When the measure μ is fixed, we simply say *almost everywhere (a.e.)*.

Exercise 2.107. Let $f : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}_+$ be a non-negative measurable function.

(i) Show that $\int f\, d\mu = 0 \iff f = 0 \quad \mu$-a.e.

(ii) Show that $\int f\, d\mu < +\infty \implies f < +\infty \quad \mu$-a.e.

While the definition of the integral of non-negative functions given above is intuitively very clear, it is not the most useful one to work with. A simplified approach is enabled by the following fundamental theorem of integral theory:

Theorem 2.108. (Monotone convergence theorem)
Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and $f_n : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}_+$, $f_n \leq f_{n+1}$, $n \in \mathbb{N}$, be a monotone increasing sequence of non-negative measurable functions. Then

$$\int \lim_n f_n\, d\mu = \lim_n \int f_n\, d\mu = \sup_n \int f_n\, d\mu$$

for every measure μ on $(\mathcal{X}, \mathcal{A})$.

The proof of the above theorem is not complicated at all, but it is not really necessary for us, so we refer the interested reader to [**?**, Theorem 2.14].

Corollary 2.109. Let $f : \mathcal{X} \to \overline{\mathbb{R}}_+$ be a non-negative measurable function, and $0 \le f_1 \le f_2 \le \dots$ be a sequence of simple measurable functions converging pointwise to f. Then

$$\int f \, d\mu = \lim_{n \to +\infty} \int f_n \, d\mu. \tag{2.23}$$

In particular,

$$\int f \, d\mu = \lim_{n \to +\infty} \sum_{k=0}^{2^n - 1} \frac{k}{2^n} \mu \left(f^{-1} \left(\left[\frac{k}{2^n}, \frac{k+1}{2^n} \right) \right) \right).$$

Proof. The first assertion is immediate from the monotone convergence theorem, and the second one follows by applying (2.23) to the approximating sequence constructed in the proof of 2.92. □

Remark 2.110. Let $(r_n)_{n \in \mathbb{N}}$ be an enumeration of the rational numbers in $[0, 1]$, and define $f_n := \mathbf{1}_{\{r_1, \dots, r_n\}}$, $n \in \mathbb{N}$. Obviously, $(f_n)_{n \in \mathbb{N}}$ is a monotone increasing sequence of functions, f_n is Riemann integrable for every $n \in \mathbb{N}$, and it is easy to see that

$$\int_0^1 f_n(x) \, dx = 0, \qquad n \in \mathbb{N}.$$

However, as we have seen above,

$$\lim_{n \to +\infty} f_n = \mathbf{1}_{\mathbb{Q} \cap [0,1]}$$

is not Riemann integrable. Hence, the dominated convergence theorem may fail to hold for the Riemann integral because already the Riemann integral of the limit function is undefined.

Another important corollary of the monotone convergence theorem is the following:

Corollary 2.111. Let $f_n : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}_+$ be measurable functions for every $n \in \mathbb{N}$. Then

$$\int_{\mathcal{X}} \sum_{n \in \mathbb{N}} f_n(x) \, d\mu(x) = \sum_{n \in \mathbb{N}} \int_{\mathcal{X}} f_n(x) \, d\mu(x).$$

In particular, the integral is additive on non-negative measurable functins.

Proof. Let us first consider the case $f_n \equiv 0$ for every $n \geq 3$, and let $f_{n,k} \leq f_{n,k+1} \leq \ldots \to f_i$, $n = 1, 2$. Then

$$\begin{aligned}
\int_{\mathcal{X}} (f_1(x) + f_2(x))\, d\mu(x) &= \lim_{n\to+\infty} \int_{\mathcal{X}} (f_{1,k}(x) + f_{2,k}(x))\ d\mu(x) \\
&= \lim_{n\to+\infty} \left[\int_{\mathcal{X}} f_{1,k}(x)\, d\mu(x) + \int_{\mathcal{X}} f_{2,k}(x)\, d\mu(x) \right] \\
&= \lim_{n\to+\infty} \int_{\mathcal{X}} f_{1,k}(x)\, d\mu(x) + \lim_{n\to+\infty} \int_{\mathcal{X}} f_{2,k}(x)\, d\mu(x) \\
&= \int_{\mathcal{X}} f_1(x)\, d\mu(x) + \int_{\mathcal{X}} f_2(x)\, d\mu(x),
\end{aligned}$$

where the first equality is due to the monotone convergence theorem, the second equality is due to Exercise 2.99, the third equality is trivial, and the last equality is again due to the monotone convergence theorem.

Iterating the above, we obtain that the integral is additive on finite sums of non-negative measurable functions. Finally, for an infinite sum we obtain

$$\begin{aligned}
\int_{\mathcal{X}} \sum_{n\in\mathbb{N}} f_n(x)\, d\mu(x) &= \int_{\mathcal{X}} \lim_{N\to+\infty} \sum_{n=1}^{N} f_n(x)\, d\mu(x) \\
&= \lim_{N\to+\infty} \int_{\mathcal{X}} \sum_{i=1}^{n} f_n(x)\, d\mu(x) \\
&= \lim_{N\to+\infty} \sum_{i=1}^{n} \int_{\mathcal{X}} f_n(x)\, d\mu(x) \\
&= \sum_{n\in\mathbb{N}} \int_{\mathcal{X}} f_n(x)\, d\mu(x),
\end{aligned}$$

where the first equality is by definition, the second equality is by the monotone convergence theorem, the third equality follows from finite additivity, and the last equality is again by definition. □

Corollary 2.112. Let $f : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}_+$ be a non-negative measurable function. Then

$$(f\mu)(A) := \int_A f\, d\mu = \int_{\mathcal{X}} (f\mathbf{1}_A)\, d\mu, \qquad A \in \mathcal{A},$$

defines a measure on $\mathcal{A}$, that we call the product of f and μ.

Proof. For any $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$, $A_n \cap_{n\neq m} A_m = \emptyset$, let $f_n := f\mathbf{1}_{A_n}$. Then

$$\begin{aligned}(f\mu)\left(\cup_{n\in\mathbb{N}} A_n\right) &= \int_{\mathcal{X}} f\mathbf{1}_{\cup_{n\in\mathbb{N}} A_n}\, d\mu = \int_{\mathcal{X}} \sum_{n\in\mathbb{N}} f\mathbf{1}_{A_n}\, d\mu = \int_{\mathcal{X}} \sum_{n\in\mathbb{N}} f_n\, d\mu \\ &= \sum_{n\in\mathbb{N}} \int_{\mathcal{X}} f_n\, d\mu = \sum_{n\in\mathbb{N}} \int_{\mathcal{X}} f\mathbf{1}_{A_n}\, d\mu = \sum_{n\in\mathbb{N}} (f\mu)(A_n),\end{aligned}$$

where the fourth equality is due to Corollary 2.111, and the rest are obvious. □

Example 2.113. Let $(\mathcal{X}, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, and $f(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. This is the density function of the standard normal distribution. The probability that the value of a random variable X with this distribution falls into an interval $[a, b]$ is given by

$$\mathbb{P}(X \in [a, b]) = \int_{[a,b]} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, d\lambda(x) = (f\lambda)([a, b]).$$

Likewise, the probability that the value falls into an arbitrary Borel set A is given by

$$\mathbb{P}(X \in A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, d\lambda(x) = (f\lambda)(A).$$

Definition 2.114. Let $X : (\Omega, \mathcal{F}) \to \mathbb{R}^d$ be a measurable function, and P be a probability measure on $\mathcal{F}$. We say that X is a continuous (more precisely: absolutely continuous w.r.t. the Lebesgue measure) random variable if there exists a non-negative measurable function $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \to \overline{\mathbb{R}}_+$ such that

$$X_* P = P \circ X^{-1} = f\lambda_d.$$

The function f is called the *density function* of X.

Finally, we define the integral of extended real-valued functions of arbitrary sign, and of complex-valued functions.

Definition 2.115. Let $f : (\mathcal{X}, A) \to \overline{\mathbb{R}}$ be an extended real-valued measurable function, and μ be a measure on $\mathcal{A}$. We say that the μ-integral of f exists, if at least one of $\int f_+\, d\mu$ and $\int f_-\, d\mu$ is finite, and define

$$\int f\, d\mu := \int f_+\, d\mu - \int f_-\, d\mu.$$

We say that f is *integrable*, if its integral exists and is finite.

Remark 2.116. Note that the integral of a non-negative measurable function always exists, but it is not necessarily finite. Likewise, the integral of an extended real-valued measurable function exists if $\int f_+ \, d\mu < +\infty$ and $\int f_- \, d\mu = +\infty$ or the other way around, but the integral of f in these cases is infinite.

Hence, f being *integrable* is a stronger property than the *existence of the integral* of f. This might make the terminology slightly confusing at first, but one gets used to it by time.

Definition 2.117. Let $f : (\mathcal{X}, A) \to \mathbb{C}$ be a complex measurable function. We say that f is *integrable* with respect to a measure μ on $\mathcal{A}$, if both $\mathfrak{Re}\, f$ and $\mathfrak{Im}\, f$ are integrable, and define the μ-integral of f as

$$\int f \, d\mu := \int \mathfrak{Re}\, f \, d\mu + i \int \mathfrak{Im}\, f \, d\mu.$$

Definition 2.118. If in Definitions 2.115 or 2.117 $(\mathcal{X}, \mathcal{A}, \mu) = (A, \mathcal{B}(A), \lambda)$ for some Borel measurable set $A \subseteq \mathbb{R}^d$ then we call the corresponding integral of a function f its *Lebesgue integral*.

Exercise 2.119. Let f, g be extended real-valued or complex-valued functions on a measurable space $(\mathcal{X}, \mathcal{A})$, and μ be a measure on $\mathcal{A}$. Show that

$$f = g \ \ \mu\text{-a.e.} \implies \Big[f \text{ is integrable} \iff g \text{ is integrable} \Big],$$

and if both are integrable then $\int f \, d\mu = \int g \, d\mu$.

Exercise 2.120. Let f be an extended real-valued or complex-valued measurable function on a measure space $(\mathcal{X}, \mathcal{A}, \mu)$.

(i) Show that if the integral of f exists then

$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu =: \|f\|_1 .$$

(ii) Show that

$$f \ \text{ is integrable} \iff \|f\|_1 = \int |f| \, d\mu < +\infty.$$

(iii) Show that if f is extended real-valued and it is integrable then it is finite μ-a.e.

(iv) Show that $f = 0$ μ-a.e. $\iff \|f\|_1 = 0$.

Remark 2.121. The integral of the absolute value of a function is called its 1-*norm*; we will discuss this quantity in much more detail later.

Exercise 2.122. Let f, g be measurable $\mathbb{K}$-valued functions on a measure space $(\mathcal{X}, \mathcal{A}, \mu)$, and $c \in \mathbb{K}$. Show that

$$\|cf\|_1 = |c| \, \|f\|_1 \,, \qquad \|f + g\|_1 \leq \|f\|_1 + \|g\|_1$$

(This means that $\|.\|_1$ is a semi-norm; see Section **??**.)

Exercise 2.123. Show that if $f_1, f_2 : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}$ are μ-integrable then so is $c_1 f_1 + c_2 f_2$ for any $c_1, c_2 \in \mathbb{K}$, and

$$\int (c_1 f_1 + c_2 f_2) = c_1 \int f_1 \, d\mu + c_2 \int f_2 \, d\mu.$$

The findings of Exercises 2.122 and 2.123 can be summarized as follows:

Corollary 2.124. The integrable $\mathbb{K}$-valued functions on a measure space form a vector space on which the integral is a linear functional, and the 1-norm is a semi-norm.

Exercise 2.125. Show that (2.18) holds for any measurable real-valued function $f : (\Omega, \mathcal{F}) \to \mathbb{R}$ that is integrable w.r.t. ϱ.
(Hint: Use approximation by simple measurable functions.)

Exercise 2.126. Let $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$, $i \in [n]$, be measure spaces, and μ be a measure on $\otimes_{i=1}^n \mathcal{A}_i$ that factorizes to the product of the μ_i (see Definition 2.60). Show that if $f_i \in \mathbb{K}^{\mathcal{X}_i}$, $i \in [n]$, are integrable then so is $\otimes_{i=1}^n f_i$, and

$$\int_{\times_{i=1}^n \mathcal{X}_i} (\otimes_{i=1}^n f_i) \; d\mu = \prod_{i=1}^n \int_{\mathcal{X}_i} f_i \, d\mu_i.$$

Conclude that for any integrable Borel measurable functions $f_i \in \mathbb{K}^{\mathbb{R}^{d_i}}$, $i \in \mathbb{N}$,

$$\int_{\mathbb{R}^{d_1 + \ldots + d_n}} (\otimes_{i=1}^n f_i) \; d\lambda = \prod_{i=1}^n \int_{\mathbb{R}^{d_i}} f_i \, d\lambda.$$

(Hint: Consider simple functions first.)

We close this section with the following convergence theorem for integrals, which we will use extensively in the discussion of quantum observables. Its proof follows easily from the monotone convergence theorem; see Section 2.3 in [**?**].

Theorem 2.127. (Dominated convergence theorem)
Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and $f_n : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$, $n \in \mathbb{N}$, be a sequence of measurable functions that is pointwise convergent, and has an integrable dominating function g, i.e.,

$$\exists \lim_{n\to+\infty} f_n(x), \quad x \in \mathcal{X}, \qquad \text{and} \qquad |f_n(x)| \leq g(x), \quad x \in \mathcal{X},\, n \in \mathbb{N},$$

where $g : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ is measurable and $\int_{\mathcal{X}} |g(x)|\, d\mu(x) < +\infty$. Then

$$\exists \lim_{n\to+\infty} \int f_n(x)\, d\mu(x) = \int \left(\lim_{n\to+\infty} f_n(x) \right) d\mu(x).$$

The most important applications of the dominated convergence theorem that we will use are the continuity and differentiability of parametric integrals:

Theorem 2.128. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, and $f : \mathcal{X} \times [a, b] \to \mathbb{C}$ be such that for all $t \in [a, b]$, $f(., t) : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ is measurable, and $\int_{\mathcal{X}} |f(x, t)|\, d\mu(x) < +\infty$.

(i) Assume that there exists an integrable function $g : (\mathcal{X}, \mathcal{A}) \to \overline{\mathbb{R}}_+$, $\int_{\mathcal{X}} |g(x)|\, d\mu(x) < +\infty$, such that

$$\sup_{t\in[a,b]} |f(x, t)| \leq g(x), \quad x \in \mathcal{X}, \qquad \text{and} \qquad \exists \lim_{t\to t_0} f(x, t), \qquad x \in \mathcal{X}.$$

Then

$$\exists \lim_{t\to t_0} \int_{\mathcal{X}} f(x, t)\, d\mu(x) = \int_{\mathcal{X}} \left(\lim_{t\to t_0} f(x, t) \right) d\mu(x). \tag{2.24}$$

In particular, if $f(x, .)$ is continuous at t_0 for every $x \in \mathcal{X}$, then so is $\int_{\mathcal{X}} f(x, .)\, d\mu(x)$.

(ii) Assume that

$$\exists\, \partial_2 f(x, t), \quad x \in \mathcal{X},\, t \in (a, b), \qquad \text{and} \qquad \sup_{t\in(a,b)} |\partial_2 f(x, t)| \leq g(x), \quad x \in \mathcal{X},$$

for some measurable function g with $\int_{\mathcal{X}} g(x)\, d\mu(x) < +\infty$. Then

$$\exists\ \partial_2 \int_{\mathcal{X}} f(x, t)\, d\mu(x) = \int_{\mathcal{X}} \partial_2 f(x, t)\, d\mu(x), \qquad t \in (a, b).$$

Proof. (i) Let $(t_n)_{n\in\mathbb{N}} \subseteq [a,b]$ be such that $\lim_{n\to+\infty} t_n = t_0$, and $f_n(x) := f(x,t_n)$. By assumption, $|f_n(x)| \le g(x)$, $x \in \mathcal{X}$, $n \in \mathbb{N}$, and hence we can apply the dominated convergence theorem, which yields

$$\begin{aligned}\exists \lim_{n\to+\infty} \int_{\mathcal{X}} f_n(x)\, d\mu(x) &= \int \left(\lim_{n\to+\infty} f_n(x) \right) d\mu(x) \\ &= \int \left(\lim_{n\to+\infty} f(x,t_n) \right) d\mu(x) \\ &= \int_{\mathcal{X}} \left(\lim_{t\to t_0} f(x,t) \right) d\mu(x).\end{aligned}$$

Since this holds for any sequence $(t_n)_{n\in\mathbb{N}}$ converging to t_0, the limit of the integrals on the LHS in (2.24) exists, and the equality in (2.24) holds.

(ii) Consider a $t_0 \in (a,b)$, and a sequence $(t_n)_{n\in\mathbb{N}} \subseteq (a,b) \setminus \{t_0\}$ converging to t_0. Define

$$h_n(x) := \frac{f(x,t_n) - f(x,t_0)}{t_n - t_0}, \qquad x \in \mathcal{X},\ n \in \mathbb{N}.$$

Then h_n is measurable, and

$$|h_n(x)| \le \sup_{t\in(a,b)} |\partial_2 f(x,t)| \le g(x), \quad x \in \mathcal{X},\ n \in \mathbb{N},$$

where the first inequality is due to the mean value theorem, and the second inequality is by assumption. Hence, we can apply the dominated convergence theorem to the sequence $(h_n)_{n\in\mathbb{N}}$ to obtain

$$\exists \lim_{n\to+\infty} \int_{\mathcal{X}} h_n(x)\, d\mu(x) = \int_{\mathcal{X}} \left(\lim_{n\to+\infty} h_n(x) \right) d\mu(x) = \int_{\mathcal{X}} \partial_2 f(x,t_0)\, d\mu(x).$$

Note that the first limit above is equal to

$$\begin{aligned}&\lim_{n\to+\infty} \int_{\mathcal{X}} \frac{f(x,t_n) - f(x,t_0)}{t_n - t_0}\, d\mu(x) \\ &= \lim_{n\to+\infty} \frac{1}{t_n - t_0} \left(\int_{\mathcal{X}} f(x,t_n)\, d\mu(x) - \int_{\mathcal{X}} f(x,t_0)\, d\mu(x) \right) \\ &= \partial_2 \left(\int_{\mathcal{X}} f(x,t)\, d\mu(x) \right) \Bigg|_{t=t_0}.\end{aligned}$$

Since this is true for any sequence $(t_n)_{n\in\mathbb{N}}$ as above, the assertion follows.

□

3 Measure theory proper

3.1 Set systems

For a set $\mathcal{X}$, we denote by $\mathcal{P}(\mathcal{X}) := \{H \subseteq \mathcal{X}\}$ the *power set* of $\mathcal{X}$, which is the collection of all subsets of $\mathcal{X}$. A *set system* $\mathcal{A}$ on $\mathcal{X}$ is simply a subset of $\mathcal{P}(\mathcal{X})$; for convenience, we include in the definition that it contains the empty set $\emptyset$. We have already encountered various important classes of subsystems: σ-rings, σ-algebras, and semi-rings in Section 2.2, and topologies in Section 1. For completeness, we collect their definitions below, together with that of a few further classes of set systems.

Definition 3.1. Let $\mathcal{X} \neq \emptyset$ be a non-empty set, and $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$, We say that $\mathcal{A}$ is a

(i) *semi-ring*, if

- $A, B \in \mathcal{A} \Longrightarrow A \cap B \in \mathcal{A}$, (closed under finite intersection)
- for all $A, B \in \mathcal{A}$, there exist finitely many pairwise disjoint sets $A_i \in \mathcal{A}$, $i = 1, \ldots, r$, such that $A \setminus B = \cup_{i=1}^{r} A_i$;

(ii) *ring*, if

- $A, B \in \mathcal{A} \Longrightarrow A \setminus B \in A$ (closed under set difference)
- $A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A}$; (closed under finite union)

(iii) σ-*ring*, if

- $A, B \in \mathcal{A} \Longrightarrow A \setminus B \in A$ (closed under set difference)
- $\{A_i\}_{i \in \mathcal{I}} \subseteq \mathcal{A}$, $\mathcal{I}$ countable $\Longrightarrow \cup_{i \in \mathcal{I}} A_i \in \mathcal{A}$; (closed under countable union)

(iv) σ-*algebra*, if it is a σ-ring and $\mathcal{X} \in \mathcal{A}$, which is equivalent to

- $\mathcal{X} \in \mathcal{A}$,
- $A \in \mathcal{A} \Longrightarrow \mathcal{X} \setminus A \in A$ (closed under complement)
- $\{A_i\}_{i \in \mathcal{I}} \subseteq \mathcal{A}$, $\mathcal{I}$ countable $\Longrightarrow \cup_{i \in \mathcal{I}} A_i \in \mathcal{A}$; (closed under countable union)

(v) *topology*, if

- $\emptyset, \mathcal{X} \in \mathcal{A}$,
- $\{G_i\}_{i \in \mathcal{I}} \subseteq \mathcal{A} \Longrightarrow \cup_{i \in \mathcal{I}} G_i \in \mathcal{A}$, where $\mathcal{I}$ can be any index set, (closedness under arbitrary union)

- $\{G_i\}_{i\in\mathcal{I}} \subseteq \mathcal{A} \Longrightarrow \cap_{i\in\mathcal{I}} G_i \in \mathcal{A}$ if $\mathcal{I}$ is finite.
  (closedness under finite intersection)

The first four classes are naturally related to measure theory, while the concept of topology is á priori somewhat disjoint. However, we will often be interested in σ-algebras that are generated by a topology, e.g., the Borel σ-algebra on a topological space, and will study the interrelation of measures and the topology.

Remark 3.2. Note that we only required in the definition that a semi-ring contains the union of *two* of its elements, but it is easy to see (by induction) that this implies that it contains the union of any finite collection of its elements. Likewise, a ring contains the union of an arbitrary finite collection of its elements.

Note also that the identity $A \cap B = A \setminus (A \setminus B)$ implies that a ring is closed under fnite intersections, and a semi-ring is closed under countable intersections.

It is easy to see that for any $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$, we have the implications

$$\mathcal{A} \text{ semi-ring} \implies \text{ring} \implies \sigma\text{-ring} \implies \sigma\text{-algebra}.$$

The implications are not true in general in the converse direction, as the following examples show.

Example 3.3. Let $\mathrm{Box}(\mathbb{R}^d) := \{\times_{i=1}^d [a_i, b_i) : a_i, b_i \in \mathbb{R}, i \in [r]\}$ be the set of boxes in $\mathbb{R}^d$. Then $\mathrm{Box}(\mathbb{R}^d)$ is a semi-ring, as we will show in ..... It is trivial to verify that $\mathrm{Box}(\mathbb{R}^d)$ is not a ring.

On the other hand, for any semi-ring $\mathcal{S}$, the generated ring is

$$\{\cup_{i=1} S_i : S_i \in \mathcal{S},\, i \in [r],\, r \in \mathbb{N}\}.$$

In particular, the collection of unions of finitely many boxes

$$\widetilde{\mathrm{Box}}(\mathbb{R}^d) := \left\{\cup_{i=1} T_i : T_i \in \mathrm{Box}(\mathbb{R}^d),\, i \in [r],\, r \in \mathbb{N}\right\}$$

is a ring, and it is easy to see that it is not a σ-ring (e.g., $\mathbb{R}^d$ is the union of countably many boxes, but not of finitely many boxes, so $\widetilde{\mathrm{Box}}(\mathbb{R}^d)$ is not closed under countable union).

Consider now an uncountable set $\mathcal{X}$, and let $\mathcal{R} := \{H \subseteq \mathcal{X} : H \text{ is countable}\}$. Then $\mathcal{R}$ is a σ-ring, but not a σ-algebra.

Semi-rings play a particularly important role in measure theory, for the following reasons:

- They have good disjunctization properties; see Section 2.2.

- They are closed under product, as we will show below. This is not true for any of the other classes of set systems; in general, the product of σ-algebras need not have stronger structure than being a semi-ring. This property of semi-rings is fundamental for the definition of product measures; see Section **??**.
- The left-closed right-open intervals (as well as the set of all intervals) in $\mathbb{R}$ form a semi-ring; more generally (due to the above closedness property), the boxes in $\mathbb{R}^d$ form a semi-ring.
- According to the Carathédory extension theorem, the semi-ring structure is sufficient to guarantee that any measure on a semi-ring has an extension to a measure on the generated σ-algebra; in particular, the volume function on the boxes extends to a measure (the Lebesgue measure) on the Borel sets of $\mathcal{B}(\mathbb{R}^d)$.

Definition 3.4. Let $\mathcal{A}_i \subseteq \mathcal{P}(\mathcal{X}_i)$ for all $i \in \mathcal{I}$, where $\mathcal{I}$ is an arbitrary index set. The *element-wise product* of the $\mathcal{A}_i$ is defined as

$$(\times)_{i\in\mathcal{I}}\,\mathcal{A}_i := \{\times_{i\in\mathcal{I}} A_i : \, A_i \in \mathcal{A}_i,\, i \in \mathcal{I}\}\,.$$

Example 3.5. $\mathrm{Box}(\mathbb{R}^d) = (\times)_{i=1}^{d}\,\mathcal{T}(\mathbb{R})$, i.e., the set of d-dimensional boxes is the d-fold product of the set of 1-dimensional boxes (intervals) with itself.

Proposition 3.6. Let $\mathcal{S}_i \subseteq \mathcal{P}(\mathcal{X}_i)$, $i = 1, \ldots, r$ be a finite collection of semi-rings. Then their element-wise product $\mathcal{S}_1\,(\times)\ldots(\times)\,\mathcal{S}_r$ is a semi-ring on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_r$.

Proof. We prove by induction on r. Let $r = 2$, and $A_i \times B_i \in \mathcal{S}_i$, $i = 1, 2$. Then

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2) \in \mathcal{S}_1\,(\times)\,\mathcal{S}_2,$$

so $\mathcal{S}_1\,(\times)\,\mathcal{S}_2$ is closed under intersections. Next, note that

$$(A_1 \times B_1) \setminus (A_2 \times B_2) = [(A_1 \setminus A_2) \times B_1] \cup [(A_1 \cap A_2) \times (B_1 \setminus B_2)]\,.$$

Since both $\mathcal{S}_1$ and $\mathcal{S}_2$ are semi-rings, we have decompositions $A_1 \setminus A_2 = \cup_{i=1}^{m} C_i$, $B_1 \setminus B_2 = \cup_{i=1}^{n} D_i$ for some $C_1, \ldots, C_m \in \mathcal{S}_1$ and $D_1, \ldots, D_n \in \mathcal{S}_2$. Hence,

$$(A_1 \times B_1) \setminus (A_2 \times B_2) = (\cup_{i=1}^{m} C_i \times B_1) \cup \left(\cup_{j=1}^{n} (A_1 \cap A_2) \times D_j\right)$$

is a decomposition of $(A_1 \times B_1) \setminus (A_2 \times B_2)$ into the disjoint union of finitely many elements in $\mathcal{S}_1\,(\times)\,\mathcal{S}_2$. Thus, $\mathcal{S}_1\,(\times)\,\mathcal{S}_2$ is a semi-ring.

Now assume that the claim of the proposition is true for $r = 1, \ldots, n$, and let $\mathcal{S}_i \in \mathcal{P}(\mathcal{X}_i)$ be semi-rings for $i = 1, \ldots, n+1$. Then we have the trivial identification $\mathcal{S}_1\,(\times)\ldots(\times)\,\mathcal{S}_n\,(\times)\,\mathcal{S}_{n+1} = (\mathcal{S}_1\,(\times)\ldots(\times)\,\mathcal{S}_n)\,(\times)\,\mathcal{S}_{n+1}$. By the induction hypothesis, $\tilde{S} := \mathcal{S}_1\,(\times)\ldots(\times)\,\mathcal{S}_n$ is a semi-ring, and hence, by the above proof, $\tilde{\mathcal{S}}\,(\times)\,\mathcal{S}_{n+1}$ is a semi-ring, too. □

Exercise 3.7. Show that $\mathrm{Box}(\mathbb{R})$ is a semi-ring.

Corollary 3.8. $\mathrm{Box}(\mathbb{R}^d)$ is a semi-ring.

Proof. We have $\mathrm{Box}(\mathbb{R}^d) = (\times)_{i=1}^{d} \mathrm{Box}(\mathbb{R})$, and the assertion follows immediately from Exercise 3.7 and Proposition 3.6. □

The following is trivial to verify:

Proposition 3.9. Let $\{\mathcal{A}_i\}_{i\in\mathcal{I}} \subseteq \mathcal{P}(\mathcal{X})$ be a collection of set systems for an arbitrary index set $\mathcal{I}$, such that for every $i \in \mathcal{I}$, $\mathcal{A}_i$ is a semi-ring/ring/σ-ring/σ-algebra. Then $\cap_{i\in\mathcal{I}}\mathcal{A}_i$ is also a semi-ring/ring/σ-ring/σ-algebra.

As an immediate consequence, we can see that for any $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$, there is a smallest semi-ring/ring/σ-ring/σ-algebra containing $\mathcal{A}$, that we call the semi-ring/ring/σ-ring/σ-algebra *generated by* $\mathcal{A}$.

Definition 3.10. Let $\mathcal{A} \subseteq \mathcal{X}$. We will use the notations

$$\sigma(\mathcal{A}) := \cap\{\mathcal{C} \subseteq \mathcal{P}(\mathcal{X}) : \mathcal{A} \subseteq \mathcal{C}, \mathcal{C} \text{ is a } \sigma\text{-algebra}\}$$

for the σ-algebra generated by $\mathcal{A}$.

For a non-empty set $\mathcal{X}$, let $\Sigma(\mathcal{X})$ denote the set of all σ-algebras on $\mathcal{X}$, i.e.,

$$\Sigma(\mathcal{X}) := \{\mathcal{A} \subseteq \mathcal{P}(\mathcal{X}) : \mathcal{A} \text{ is a } \sigma\text{-algebra}\}.$$

There is a natural partial order on $\Sigma(\mathcal{X})$, given by the set-theoretic inclusion. For any collection $\{\mathcal{A}_i\}_{i\in\mathcal{I}} \subseteq \Sigma(\mathcal{X})$,

$$\wedge_{i\in\mathcal{I}}\mathcal{A}_i := \cap_{i\in\mathcal{I}}\mathcal{A}_i$$
$$\vee_{i\in\mathcal{I}}\mathcal{A}_i := \sigma\left(\cup_{i\in\mathcal{I}}\mathcal{A}_i\right)$$

are the largest lower bound and the smallest upper bound of $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$, respectively. This shows that $\Sigma(\mathcal{X})$ is a *complete lattice*. Moreover, $\Sigma(\mathcal{X})$ has a smallest element $\{\emptyset, \mathcal{X}\}$, and largest element $\mathcal{P}(\mathcal{X})$.

Definition 3.11. Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces. We say that a map $f : \mathcal{X} \to \mathcal{Y}$ is $\mathcal{A} \to \mathcal{B}$ *measurable*, if for every measurable set in $(\mathcal{Y}, \mathcal{B})$, its inverse image under f is measurable in $(\mathcal{X}, \mathcal{A})$, i.e.,

$$\forall B \in \mathcal{B} : \quad f^{-1}(B) := \{x \in \mathcal{X} : f(x) \in B\} \in \mathcal{A}.$$

We will also use the terminology "$f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ is measurable". When we consider $\mathcal{A}$ and $\mathcal{B}$ fixed, we will simply call such an f *measurable*.

Example 3.12.

- Let $\mathcal{X}$ be equipped with the largest possible σ-algebra, $\mathcal{A} := \mathcal{P}(\mathcal{X})$. Then for any $(\mathcal{Y}, \mathcal{B})$, any map $f : \mathcal{X} \to \mathcal{Y}$ is measurable. That is, any map is $\mathcal{P}(\mathcal{X}) \to \mathcal{B}$ measurable, irrespective of what $\mathcal{B}$ is.

- Similarly, if $\mathcal{Y}$ is equipped with the smallest possible σ-algebra $\mathcal{B} = \{\emptyset, \mathcal{X}\}$ then for any $(\mathcal{X}, \mathcal{A})$, any map $f : \mathcal{X} \to \mathcal{Y}$ is measurable. That is, any map is $\mathcal{A} \to \{\emptyset, \mathcal{Y}\}$ measurable, irrespective of what $\mathcal{A}$ is.

- Let $\mathcal{X}$ be equipped with the smallest possible σ-algebra, $\mathcal{A} := \{\emptyset, \mathcal{X}\}$. Then only the constant maps will be measurable for any $(\mathcal{Y}, \mathcal{B})$. That is, only the constant maps are $\{\emptyset, \mathcal{X}\} \to \mathcal{B}$ measurable, irrespective of what $\mathcal{B}$ is.

- Let $f : \mathcal{X} \to \mathcal{Y}$ be $\mathcal{A} \to \mathcal{B}$ measurable. If we replace $\mathcal{A}$ with a larger σ-algebra $\mathcal{A}' \supseteq \mathcal{A}$, and $\mathcal{B}$ with a smaller σ-algebra $\mathcal{B}' \subseteq \mathcal{B}$, then f will also be $\mathcal{A}' \to \mathcal{B}'$ measurable.

Example 3.13. Consider the constructions in Section **??**.

- Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces, and $f : \mathcal{X} \to \mathcal{Y}$ be $\mathcal{A} \to \mathcal{B}$ measurable. Then for any $E \in \mathcal{A}$, $f\big|_E$ is $\mathcal{A}\big|_E \to \mathcal{B}$ measurable.

- Let $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i \in \mathcal{I}}$ be disjoint measurable spaces, and $f_i : \mathcal{X}_i \to \mathcal{Y}_i$ be maps for all $i \in \mathcal{I}$. Then

  $$f : \cup_{i \in \mathcal{I}} \mathcal{X}_i \to \mathcal{Y}, \qquad f(x) := f_i(x) \text{ if } x \in \mathcal{X}_i,$$

  is $\cup_{i \in \mathcal{I}} \mathcal{A}_i \to \mathcal{B}$ measurable if and only if f_i is $\mathcal{A}_i \to \mathcal{B}$ measurable for all $i \in \mathcal{I}$.

- Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and $\mathcal{F} \subseteq \mathcal{Y}^{\mathcal{X}}$ be a family of functions from $\mathcal{X}$ to $\mathcal{Y}$. Then the push-forward σ-algebra $\overrightarrow{\mathcal{F}}(\mathcal{A})$ is the largest σ-algebra $\mathcal{B}$ on $\mathcal{Y}$ such that all $f \in \mathcal{F}$ is $\mathcal{A} \to \mathcal{B}$ measurable.

- Let $(\mathcal{Y}, \mathcal{B})$ be a measurable space, and $\mathcal{F} \subseteq \mathcal{Y}^{\mathcal{X}}$ be a family of functions from $\mathcal{X}$ to $\mathcal{Y}$. Then the pull-back σ-algebra $\overleftarrow{\mathcal{F}}(\mathcal{A})$ is the smallest σ-algebra $\mathcal{A}$ on $\mathcal{X}$ such that all $f \in \mathcal{F}$ is $\mathcal{A} \to \mathcal{B}$ measurable.

- Let $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i \in \mathcal{I}}$ be an arbitrary collection of measurable spaces. Then $\otimes^{\text{cyl}}_{i \in \mathcal{I}} \mathcal{A}_i$ is the smallest σ-algebra $\mathcal{B}$ such that all coordinate functions

  $$\pi_i : \times_{i \in \mathcal{I}} \mathcal{X}_i \to \mathcal{X}_i, \qquad \pi_i(x) := x_i, \qquad x \in \times_{i \in \mathcal{I}} \mathcal{X}_i,$$

  are $\mathcal{B} \to \mathcal{A}_i$ measurable.

The last example above shows that the cylinder product is in some sense more natural than the full product of the σ- algebras (note again that for a countable index set the two coincide).

1. Restriction of a measurable space

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and $E \in \mathcal{A}$. Then

$$\mathcal{A}\big|_E := \{A \cap E : \, A \in \mathcal{A}\}$$

is a σ-algebra on $\mathcal{E}$, that we call the *restriction* of $\mathcal{A}$ onto E. We call the measurable space $(\mathcal{X}, \mathcal{A})\big|_E := (E \, \mathcal{A}\big|_E)$ the restriction of $(\mathcal{X}, \mathcal{A})$ onto E.

2. Disjoint union of measurable spaces

Let $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i \in \mathcal{I}}$ be an arbitrary collection of measurable spaces, where the $\mathcal{X}_i$ are pairwise disjoint. Then

$$\uplus_{i \in \mathcal{I}} \mathcal{A}_i := \{\uplus_{i \in \mathcal{I}} A_i : \, A_i \in \mathcal{A}_i, \, i \in \mathcal{I}\}$$

is a σ-algebra on $\uplus_{i \in \mathcal{I}} \mathcal{X}_i$, that we call the *disjoint union* of of the σ-algebras $\{\mathcal{A}_i\}_{i \in \mathcal{I}}$. The measurable space $\uplus_{i \in \mathcal{I}} (\mathcal{X}_i, \mathcal{A}_i) := (\uplus_{i \in \mathcal{I}} \mathcal{X}_i, \uplus_{i \in \mathcal{I}} \mathcal{A}_i)$ is called the disjoint union of the measurable spaces $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i \in \mathcal{I}}$.

3. Push-forward of a σ-algebra by a function family

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and $\mathcal{F} := \{f_i : \, \mathcal{X} \to \mathcal{Y}\}$ be a collection of functions from $\mathcal{X}$ to $\mathcal{Y}$. Then

$$\vec{\mathcal{F}}(\mathcal{A}) := \left\{B \subseteq \mathcal{Y} : \, f_i^{-1}(B) \in \mathcal{A}, \, i \in \mathcal{I}\right\} = \cap_{i \in \mathcal{I}} \left\{B \subseteq \mathcal{Y} : \, f_i^{-1}(B) \in \mathcal{A}\right\}$$

is a σ-algebra in $\mathcal{Y}$, that we call the *push-forward* of the σ-algebra $\mathcal{A}$ by the function family $\mathcal{F}$.

4. Pull-back of a σ-algebra by a function family

Let $\mathcal{X}$ be a set, $(\mathcal{Y}, \mathcal{B})$ be a measurable space, and $\mathcal{F} := \{f_i : \, \mathcal{X} \to \mathcal{Y}\}$ be a collection of functions from $\mathcal{X}$ to $\mathcal{Y}$. Then

$$\overleftarrow{\mathcal{F}}(\mathcal{A}) := \sigma\left(\left\{f_i^{-1}(B) : \, B \in \mathcal{B}, \, i \in \mathcal{I}\right\}\right) = \sigma\left(\cup_{i \in \mathcal{I}} \left\{f_i^{-1}(B) : \, B \in \mathcal{B}\right\}\right)$$

is a σ-algebra in $\mathcal{X}$, that we call the *pull-back* of the σ-algebra $\mathcal{B}$ by the function family $\mathcal{F}$.

Before the next construction, we introduce the following notation: If $\mathcal{A}_i \subseteq \mathcal{P}(\mathcal{X}_i)$ for all $i \in \mathcal{I}$, where $\mathcal{I}$ is an arbitrary index set, then

$$(\times)_{i \in \mathcal{I}} \, \mathcal{A}_i := \{\times_{i \in \mathcal{I}} A_i : \, A_i \in \mathcal{A}_i, \, i \in \mathcal{I}\} \, .$$

5. The product of measurable spaces

Let $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i\in\mathcal{I}}$ be an arbitrary collection of measurable spaces. A product set $\times_{i\in\mathcal{I}} A_i$, where all $A_i \in \mathcal{A}_i$ are measurable, is called a *measurable box*. The σ-algebra generated by all measurable boxes is the *product σ-algebra*

$$\begin{aligned}\otimes_{i\in\mathcal{I}}\mathcal{A}_i &:= \sigma\left((\times)_{i\in\mathcal{I}}\,\mathcal{A}_i\right)\\ &= \sigma\left(\{\times_{i\in\mathcal{I}} A_i :\, A_i \in \mathcal{A}_i,\, i \in \mathcal{I}\}\right).\end{aligned}$$

We define the product $\otimes_{i\in\mathcal{I}}(\mathcal{X}_i, \mathcal{A}_i)$ of the measurable spaces $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i\in\mathcal{I}}$ as $\times_{i\in\mathcal{I}}\mathcal{X}_i$ equipped with the product σ-algebra, i.e.,

$$\otimes_{i\in\mathcal{I}}(\mathcal{X}_i, \mathcal{A}_i) := (\times_{i\in\mathcal{I}}\mathcal{X}_i, \otimes_{i\in\mathcal{I}}\mathcal{A}_i).$$

6. The cylinder product of measurable spaces

Let $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i\in\mathcal{I}}$ be an arbitrary collection of measurable spaces.

Definition 3.14. Let $A_{i_1} \in \mathcal{A}_{i_1}, \ldots, A_{i_m} \in \mathcal{A}_{i_m}$ be a finite collection of measurable sets for some $i_1, \ldots, i_m \in \mathcal{I}$. The *cylinder set* $(A_{i_1}, \ldots, A_{i_m})^{\text{cyl}}$ is defined as

$$(A_{i_1}, \ldots, A_{i_m})^{\text{cyl}} := \left\{x \in \times_{i\in\mathcal{I}}\mathcal{X}_i :\, x_{i_j} \in A_{i_j},\, j \in [m]\right\}.$$

The σ-algebra generated by all cylinder sets is the *cylinder product σ-algebra*

$$\otimes^{\text{cyl}}_{i\in\mathcal{I}}\mathcal{A}_i = \sigma\left(\left\{(A_{i_1}, \ldots, A_{i_m})^{\text{cyl}} :\, A_{i_j} \in \mathcal{A}_{i_j},\, j \in [m],\, \{i_1, \ldots, i_m\} \subseteq \mathcal{I},\, m \in \mathbb{N}\right\}\right).$$

We define the cylinder product $\otimes^{\text{cyl}}_{i\in\mathcal{I}}(\mathcal{X}_i, \mathcal{A}_i)$ of the measurable spaces $\{(\mathcal{X}_i, \mathcal{A}_i)\}_{i\in\mathcal{I}}$ as $\times_{i\in\mathcal{I}}\mathcal{X}_i$ equipped with the cylinder product σ-algebra, i.e.,

$$\otimes^{\text{cyl}}_{i\in\mathcal{I}}(\mathcal{X}_i, \mathcal{A}_i) := \left(\times_{i\in\mathcal{I}}\mathcal{X}_i, \otimes^{\text{cyl}}_{i\in\mathcal{I}}\mathcal{A}_i\right).$$

Exercise 3.15. (i) Show that the collection of the cylinder sets forms a semi-ring.

(ii) Show that when $\mathcal{I}$ is countable, the two notions of product coincide, i.e.,

$$\otimes_{i\in\mathcal{I}}\mathcal{A}_i = \otimes^{\text{cyl}}_{i\in\mathcal{I}}\mathcal{A}_i.$$

In particular, the cylinder sets generate the product σ-algebra.

3.2 The Borel σ-algebra

Example 3.16. What are the differences between the definitions of a σ-algebra and a topology? Show an example of a set $\mathcal{X}$ and some $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ that is a σ-algebra but not a topology. Show an example of a topology $\tau \subseteq \mathcal{P}(\mathcal{X})$ that is not a σ-algebra.

The following notion is crucial in the study of the connection of topology and measure theory.

Definition 3.17. Let $(X, \tau_{\mathcal{X}})$ be a topological space, where $\tau_{\mathcal{X}}$ denotes the set of open sets in $\mathcal{X}$. The *Borel σ-algebra* $\mathcal{B}(\mathcal{X})$ on $\mathcal{X}$ is the σ-algebra generated by the collection of open sets in $\mathcal{X}$, i.e., $\mathcal{B}(\mathcal{X}) := \sigma(\tau_{\mathcal{X}})$.

Note that this depends on the topology on $\mathcal{X}$, and different topologies might lead to different Borel σ-algebras. In general, we will have one topology on $\mathcal{X}$ fixed, and we suppress the dependence of the Borel σ-algebra on the topology in the notation $\mathcal{B}(\mathcal{X})$. If we want to make this dependence explicit, we will use the notation $\mathcal{B}(\mathcal{X}, \tau)$.

Example 3.18. It is easy to see that $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathrm{Box}(\mathbb{R}^d))$, i.e., the set of boxes generate the same σ-algebra as the set of all open sets. Moreover, by lemma **??**,

$$\mathcal{B}(\mathbb{R}^d) = \sigma(\mathrm{Box}(\mathbb{R}^d)) = \sigma_r(\mathrm{Box}(\mathbb{R}^d)).$$

Assume now that, instead of the Euclidean topology, we equip $\mathbb{R}^d$ with the discrete topology (generated by the discrete metric $d(x, y) := 1 \ \forall x \neq y$). Then $\mathcal{B}(\mathbb{R}^d, \tau_{\mathrm{disc}}) = \mathcal{P}(\mathbb{R}^d)$, that is strictly larger than the Borel σ-algebra corresponding to the Euclidean topology.

Our aim in the rest of this section is to explore the relation between the notion of product for topological spaces and for σ-algebras. More precisely, we will show that for a finite collection of separable topological spaces, the Borel σ-algebra of the product space is exactly the product of the Borel σ-algebras of the components. In particular, the Borel σ-algebra of $\mathbb{R}^d$ is the product of the Borel σ-algebras of d copies of $\mathbb{R}$.

To explore the connection between the notion of product for topological and for measure spaces, it will be useful to define the product of σ-rings, analogously to the product of σ-algebras.

For a collection $\{(\mathcal{X}_i, \mathcal{R}_i)\}_{i \in \mathcal{I}}$ of σ-rings, we define their product σ-ring $\otimes^{(r)}_{i \in \mathcal{I}} \mathcal{A}_i$ as

$$\begin{aligned} \otimes^{(r)}_{i \in \mathcal{I}} \mathcal{A}_i &:= \sigma_r \left((\times)_{i \in \mathcal{I}} \mathcal{R}_i \right) \\ &= \sigma_r \left(\{ \times_{i \in \mathcal{I}} A_i : \ A_i \in \mathcal{R}_i, \, i \in \mathcal{I} \} \right). \end{aligned}$$

Lemma 3.19. Let Γ_i be a set system in $\mathcal{X}_i$ for $i = 1, \dots, n$. Then

$$\sigma_r(\Gamma_1) \otimes^{(r)} \dots \otimes^{(r)} \sigma_r(\Gamma_n) = \sigma_r\left(\sigma_r(\Gamma_1)\,(\times)\dots(\times)\,\sigma_r(\Gamma_n)\right) \tag{3.25}$$

$$= \sigma_r\left(\Gamma_1\,(\times)\dots(\times)\,\Gamma_n\right). \tag{3.26}$$

Proof. The first equality is by definition, and the inclusion $\sigma_r\left(\sigma_r(\Gamma_1)\,(\times)\dots(\times)\,\sigma_r(\Gamma_n)\right) \supseteq \sigma_r\left(\Gamma_1\,(\times)\dots(\times)\,\Gamma_n\right)$ is trivial, hence we only have to prove the converse inclusion

$$\sigma_r\left(\sigma_r(\Gamma_1)\,(\times)\dots(\times)\,\sigma_r(\Gamma_n)\right) \subseteq \sigma_r\left(\Gamma_1\,(\times)\dots(\times)\,\Gamma_n\right). \tag{3.27}$$

We do this by induction on n.

Let $n = 2$. For a fixed $B \in \Gamma_2$, let

$$\{A \subseteq \mathcal{X}_1 :\ A \times B \in \sigma_r\left(\Gamma_1(\times)\Gamma_2\right)\}. \tag{3.28}$$

Since $(\cup_{j\in\mathcal{J}} A_j) \times B = \cup_{j\in\mathcal{J}}(A_j \times B)$, and $(A_1 \setminus A_2) \times B = (A_1 \times B) \setminus (A_2 \times B)$, we see that the set system in (7.188) is a σ-ring. Since it contains Γ_1, we get

$$\sigma_r(\Gamma_1)\,(\times)\{B\} \subseteq \sigma_r\left(\Gamma_1(\times)\Gamma_2\right).$$

Since this holds for every $B \in \Gamma_2$, we can further conclude that

$$\sigma_r(\Gamma_1)\,(\times)\,\Gamma_2 \subseteq \sigma_r\left(\Gamma_1(\times)\Gamma_2\right). \tag{3.29}$$

As above, we can see that

$$\{B \subseteq \mathcal{X}_2 :\ A \times B \in \sigma_r\left(\Gamma_1(\times)\Gamma_2\right) \text{ for all } A \in \sigma_r(\Gamma_1)\}$$
$$= \bigcap_{A\in\sigma_r(\Gamma_1)} \{B \subseteq \mathcal{X}_2 :\ A \times B \in \sigma_r\left(\Gamma_1(\times)\Gamma_2\right)\}$$

is a σ-ring. By (3.29) it contains Γ_2, and hence

$$\sigma_r(\Gamma_1)\,(\times)\,\sigma_r(\Gamma_2) \subseteq \sigma_r\left(\Gamma_1(\times)\Gamma_2\right),$$

from which (3.25) follows.

Assume that we have proved the assertion for all $n = 1, \dots, m$, and let $n = m+1$. Then

$$\begin{aligned}
&\sigma_r\left(\Gamma_1\,(\times)\dots(\times)\,\Gamma_m\,(\times)\,\Gamma_{m+1}\right)\\
&\quad\supseteq \sigma_r\left(\sigma_r\left(\Gamma_1\,(\times)\dots(\times)\,\Gamma_m\right)(\times)\,\sigma_r(\Gamma_{m+1})\right)\\
&\quad\supseteq \sigma_r\left(\sigma_r\left(\sigma_r(\Gamma_1)\,(\times)\dots(\times)\,\sigma_r(\Gamma_m)\right)(\times)\,\sigma_r(\Gamma_{m+1})\right)\\
&\quad\supseteq \sigma_r\left(\sigma_r(\Gamma_1)\,(\times)\dots(\times)\,\sigma_r(\Gamma_m)\,(\times)\,\sigma_r(\Gamma_{m+1})\right),
\end{aligned}$$

where the first inclusion follows by applying (3.27) with $n = 2$, the second by applying (3.27) with $n = m$, and the last inclusion is trivial. □

Corollary 3.20. Let Γ_i be a set system in $\mathcal{X}_i$ for $i = 1, \ldots, n$. Assume that for all i, there exists a countable family $\{A_{i,j}\}_{j \in \mathcal{J}} \subseteq \Gamma_i$ such that $\cup_{j \in \mathcal{J}} A_{i,j} = \mathcal{X}_i$. Then

$$\sigma(\Gamma_1) \otimes \ldots \otimes \sigma(\Gamma_n) = \sigma\left(\Gamma_1 (\times) \ldots (\times) \Gamma_n\right).$$

Proof. Immediate from Lemma **??** and Lemma 3.19. □

Corollary 3.21. Let $\{(\mathcal{X}_i, d_i)\}_{i \in \mathcal{I}}$ be a finite collection of separable metric spaces. Then

$$\mathcal{B}\left(\otimes_{i \in \mathcal{I}} (\mathcal{X}_i, d_i)\right) = \otimes_{i \in \mathcal{I}} \mathcal{B}\left(\mathcal{X}_i, d_i\right).$$

Proof. Due to the assumption of separability, every open set in $\otimes_{i \in \mathcal{I}} \tau_{d_i}$ is a countable union of open boxes, and therefore

$$\mathcal{B}\left(\otimes_{i \in \mathcal{I}} (\mathcal{X}_i, d_i)\right) := \sigma\left(\otimes_{i \in \mathcal{I}} \tau_{d_i}\right) = \sigma\left((\times)_{i \in \mathcal{I}} \tau_{d_i}\right)$$

(see Exercise 1.8). The assumption of Corollary 3.20 is trivially satisfied (as $\mathcal{X}_i \in \tau_{d_i}$), and therefore we have

$$\sigma\left((\times)_{i \in \mathcal{I}} \tau_{d_i}\right) = \otimes_{i \in \mathcal{I}} \sigma\left(\tau_{d_i}\right) = \otimes_{i \in \mathcal{I}} \mathcal{B}\left(\mathcal{X}_i, d_i\right).$$

□

As a special case, we get

Corollary 3.22. For every $d \in \mathbb{N}$, $\mathcal{B}(\mathbb{K}^d) = \otimes_{i=1}^{d} \mathcal{B}(\mathbb{K})$. More generally, for any $d_1, \ldots, d_r \in \mathbb{N}$,

$$\mathcal{B}\left(\mathbb{K}^{d_1 + \ldots + d_r}\right) = \otimes_{i=1}^{r} \mathcal{B}(\mathbb{K}^{d_i}).$$

Proof. Immediate from Corollary 3.21 and Exercise 1.8. □

3.3 Measurable maps

The continuity of a map between topological spaces is defined analogously to the notion of measurability:

Definition 3.23. Let $(\mathcal{X}, \tau_{\mathcal{X}})$ and $(\mathcal{Y}, \tau_{\mathcal{Y}})$ be topological spaces and $f : \mathcal{X} \to \mathcal{Y}$ be a map. We say that f is *continuous*, if the inverse image of any open set in $\mathcal{Y}$ is an open set in $\mathcal{X}$, i.e.,

$$\forall G \in \tau_{\mathcal{Y}} : \quad f^{-1}(G) \in \tau_{\mathcal{X}}.$$

Exercise 3.24. Show that when $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ are metric spaces then $f : \mathcal{X} \to \mathcal{Y}$ is continuous in the above sense if and only if

$$\forall x \in \mathcal{X} \;\; \forall \varepsilon > 0 \;\; \exists \delta > 0 : \;\; d(x, y) < \delta \implies d(f(x), f(y)) < \varepsilon,$$

the usual description of continuity in metric spaces.

Exercise 3.25. Show that the product topology is the smallest topolgy on $\times_{i \in \mathcal{I}} \mathcal{X}_i$ such that all the coordinate functions

$$\pi_i : \times_{i \in \mathcal{I}} \mathcal{X}_i \to \mathcal{X}_i, \qquad \pi_i(x) := x_i, \qquad x \in \times_{i \in \mathcal{I}} \mathcal{X}_i$$

are continuous.

Remark 3.26. When a function f is defined on a measurable space $(\mathcal{X}, \mathcal{A})$ and maps into a topological space $(\mathcal{Y}, \tau)$ then, unless otherwise specified, by its measurability we mean its $\mathcal{A} \to \mathcal{B}(\mathcal{Y})$ measurability, where $\mathcal{B}(\mathcal{Y})$ is the Borel σ- algebra on $\mathcal{Y}$.

This applies to functions mapping into $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and, more generally, to functions mapping into normed vector spaces over $\mathbb{K}$.

Similarly, for functions taking extended real values, i.e., mapping into $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, measurability means measurability when $\overline{\mathbb{R}}$ is equipped with its Borel σ-algebra, generated by all sets $\{(-\infty, c) : c \in \mathbb{R}\}$.

Exercise 3.27. Let $(\mathcal{X}, \tau_{\mathcal{X}})$ and $(\mathcal{Y}, \tau_{\mathcal{Y}})$ be topological spaces. Show that if $f : \mathcal{X} \to \mathcal{Y}$ is continuous then it is $\mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{Y})$ measurable.

Next, we explore various operations that preserve measurability.

Proposition 3.28. The composition of measurable functions is measurable, i.e., if $f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ and $g : (\mathcal{Y}, \mathcal{B}) \to (\mathcal{Z}, \mathcal{C})$ are measurable then $g \circ f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Z}, \mathcal{C})$ is also measurable.

Proof. Trivial. □

We will often use the following lemma, without extra notice.

Lemma 3.29. Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces, and let $\Gamma \subseteq \mathcal{P}(\mathcal{Y})$ be a generator system for $\mathcal{B}$, i.e., $\sigma(\Gamma) = \mathcal{B}$. Then a function $f : \mathcal{X} \to \mathcal{Y}$ is measurable if and only if $f^{-1}(B) \in \mathcal{A}$ for all $A \in \Gamma$.

Proof. Obviously, measurability of f implies that $f^{-1}(B) \in \mathcal{A}$ for any $B \in \Gamma$. To prove the converse, note that $\{B \subseteq \mathcal{Y} : f^{-1}(B) \in \mathcal{A}\}$ is a σ-algebra. By assumption, this contains Γ, and hence it contains $\mathcal{B}$, too, proving the measurability of f. □

Example 3.30.

- A function $f : \mathcal{X}, \to \overline{\mathbb{R}}$ is measurable if and only if $\{f > c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$, if and only if $\{f \geq c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$, if and only if $\{f < c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$, if and only if $\{f \leq c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$.
- Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and $(\mathcal{Y}, d)$ be a metric space. Then $f : \mathcal{X} \to \mathcal{Y}$ is measurable if and only if $f^{-1}(B(y, 1/m)) \in \mathcal{A}$ for all $y \in \mathcal{Y}$ and $m \in \mathbb{N}$, where $B(y, \varepsilon) := \{z \in \mathcal{Y} : d(z, y) < \varepsilon\}$ for all $y \in \mathcal{Y}$, $\varepsilon > 0$.

Lemma 3.31. Let $\{(\mathcal{Y}_i, \mathcal{B}_i)\}_{\in \mathcal{I}}$ be an arbitrary collection of measurable spaces, and $(\mathcal{X}, \mathcal{A})$ be a measurable space. A function $f : \mathcal{X} \to \times_{i=1}^r \mathcal{Y}_i$ is $\mathcal{A} \to \otimes_{i \in \mathcal{I}}^{\text{cyl}} \mathcal{B}_i$ measurable if and only if all of its coordinates $f_i := \pi_i \circ f$ are measurable.

Proof. Since all projections π_i are $\otimes_{i \in \mathcal{I}}^{\text{cyl}} \mathcal{B}_i \to \mathcal{B}_i$ measurable, measurability of f implies the measurability of all f_i. Conversely, if all f_i are measurable then for all finite collection $B_{i_j} \in \mathcal{B}_{i_j}$, $j = 1, \ldots, m$, $f^{-1}((B_{i_1}, \ldots, B_{i_m})^{\text{cyl}}) = \cap_{j=1}^m f_{i_j}^{-1}(B_{i_j}) \in \mathcal{A}$. Since the cylinder sets generate the cylinder product σ-algebra on $\times_{i=1}^r \mathcal{Y}_i$ by definition, measurability of f follows by Lemma 3.29. □

The following corollary is immediate:

Corollary 3.32. Let $f_i : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B}_i)$ be measurable functions for all i, where $\{(\mathcal{Y}_i, \mathcal{B}_i)\}_{i \in \mathcal{I}}$ is an arbitrary collection of measurable spaces. Then $f(x) := (f_i(x))_{i \in \mathcal{I}} \in \times_{i \in \mathcal{I}} \mathcal{X}_i$, $x \in \mathcal{X}$ is measurable.

Corollary 3.33. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and $f(x) := (f_1(x), \ldots, f_d(x)) \in \mathbb{K}^d$, $x \in \mathcal{X}$, be a function. Then the following are equivalent:

(i) f is measurable

(ii) for all linear functionals $\varphi \in \mathbb{K}^*$, $\varphi \circ f : (\mathcal{X}, \mathcal{A}) \to (\mathbb{K}, \mathcal{B}(\mathbb{K}))$ is measurable,

(iii) for all $i \in [d]$, f_i is measurable.

Proof. (i)$\Longrightarrow$(ii) due to 3.27, as φ is continuous. (ii)$\Longrightarrow$(iii) is trivial, as taking the i-th coordinate is a linear functional on $\mathbb{K}^d$. Finally, (iii)$\Longrightarrow$(i) follows by Lemma 3.31 and Corollary 3.22. □

Lemma 3.34. Let $\{(\mathcal{Y}_i, d_i)\}_{i=1}^r$ be a finite collection of separable metric spaces, let $(\mathcal{Z}, \tau_{\mathcal{Z}})$ be an arbitrary topological space, and $\Phi : \times_{i=1}^r \mathcal{Y}_i \to \mathcal{Z}$ be a continuous map. Then for any measurable space $(\mathcal{X}, \mathcal{A})$, and measurable functions $f_i : \mathcal{X} \to \mathcal{Y}_i$, the function $x \mapsto \Phi(f_1(x), \ldots, f_n(x))$ is measurable.

Proof. □

Definition 3.35. We say that a map $f : \mathcal{X} \to \mathcal{Y}$ is *simple* if its range $\{f(x) : x \in \mathcal{X}\}$ is finite. Obviously, a function f is simple if and only if there exists a finite partition $\mathcal{X} = \uplus_{i=1}^{r} A_i$ such that f is constant on every A_i.

Assume that $\mathcal{X}$ and $\mathcal{Y}$ are equipped with σ-algebras $\mathcal{A}$ and $\mathcal{B}$. Then we say that $f : \mathcal{X} \to \mathcal{Y}$ is a *measurable simple function* or a *simple measurable function* if it is simple and measurable. This is equivalent to the existence of a finite partition $\mathcal{X} = \uplus_{i=1}^{r} A_i$ such that f is constant on all A_i, and all A_i are measurable. We denote the set of measurable simple functions from $(\mathcal{X}, A)$ to $(\mathcal{Y}, B)$ by $\mathcal{E}(\mathcal{X}, \mathcal{A}, \mathcal{Y}, \mathcal{B})$. When $\mathcal{Y}$ is a topological space and $\mathcal{B}$ is its Borel σ-algebra, we use the shorter notation $\mathcal{E}(\mathcal{X}, \mathcal{A}, \mathcal{Y})$.

When $\mathcal{Y}$ is a vector space, or $Y = \overline{\mathbb{R}}$, we can write every simple function as

$$f = \sum_{i=1}^{r} c_i \mathbf{1}_{A_i},$$

where every $c_k \in \mathcal{Y}$, $\mathcal{X} = \uplus_{i=1}^{r} A_i$, and f is measurable if and only if it can be written in the above form with all A_i measurable.

As it turns out, any measurable function mapping into a separable metric space can be well approximated by simple measurable functions. We start with the following lemma, important on its own right.

Lemma 3.36. Let $f_n : (\mathcal{X}, A) \to (\mathcal{Y}, d)$ be functions mapping from a measurable space $(\mathcal{X}, A)$ to a separable metric space $(\mathcal{Y}, d)$ such that $(f_n)_{n\in\mathbb{N}}$ is pointwise convergent. If all f_n are measurable then so is $\lim_n f_n$.

Proof. Let $f := \lim_n f_n$. The following is easy to see: For any open set $U \in \tau_{\mathcal{Y}}$, and any $x \in \mathcal{X}$,

$$f(x) \in U \iff \exists m \in \mathbb{N} \text{ s.t. for all large enough } n, \ \ d(f_n(x), \mathcal{Y} \setminus U) > 1/m.$$

This can be rewritten as

$$f^{-1}(U) = \cup_{m\in\mathbb{N}} \cup_{N\in\mathbb{N}} \cap_{n=N}^{+\infty} f_n^{-1} \left(\{y \in \mathcal{Y} : d(y, \mathcal{Y} \setminus U) > 1/m\}\right). \tag{3.30}$$

Note that $\{y \in \mathcal{Y} : d(y, \mathcal{Y} \setminus U) > 1/m\}$ is an open set for every $m \in \mathbb{N}$, and hence if all f_n are measurable then $f_n^{-1}\left(\{y \in \mathcal{Y} : d(y, \mathcal{Y} \setminus U) > 1/m\}\right)$ is measurable for all n and m. Thus, by (3.30), $f^{-1}(U)$ is measurable as well. □

Lemma 3.37. (Approximation lemma) Let f be a function mapping from a measurable space $(\mathcal{X}, \mathcal{A})$ into a separable metric space $(\mathcal{Y}, d)$. Then f is measurable if and only if there exists a sequence $(f_n)_{n\in\mathbb{N}}$ if simple measurable functions that pointwise converges to f.

Moreover, if $\mathcal{Y}$ is a separable normed space, then we can choose the sequence $(f_n)_{n\in\mathbb{N}}$ such that $\|f_n(x)\| \le \|f(x)\|$ at all $x \in \mathcal{X}$, and if $y = \overline{\mathbb{R}}$ then we can further assume that $f_n \le f_{n+1} \le f$ for all $n \in \mathbb{N}$.

Proof. By the separability assumption, we can find countably many open sets $\{B_n\}_{n\in\mathbb{N}} \subseteq \tau_{\mathcal{Y}}$ such that for any $y \in \mathcal{Y}$ and any $U \in \tau_{\mathcal{Y}}$ containing Y, there exists a B_n such that $y \in B_n \subseteq U$. (For instance, the collection of open balls with rational radius around points of a countable dense set will do.)

For every $n \in \mathbb{N}$, let y_n be a point in the closure of B_n. Let $f_1 \equiv y_1$, and for every $n \geq 2$, define f_n recursively as

$$f_n(x) := \begin{cases} y_n, & \text{if } f(x) \in B_n, \text{ and } d(f(x), y_n) < d(f(x), f_{n-1}(x)), \\ f_{n-1}(x), & \text{otherwise.} \end{cases}$$

Obviously, all f_n are simple functions, as $|\operatorname{ran} f_n| \leq n$. Moreover, f_1 is measurable, and if f_{n-1} is measurable then so is f_n. Indeed, the value of f_n can only change, as compared to the value of f_{n-1}, to become the constant y_n on the set $f^{-1}\left(B_n \cap \{y \in \mathcal{Y} : d(y, y_n) < d\}\right)$ □

3.4 Set functions

Next, we explore the properties of the volume function on the set of boxes. Again, we follow a more general approach, where we introduce various notions for later purposes.

Definition 3.38. Let $\mathcal{A}$ be a set system in some set $\mathcal{X}$. A *set function* on $\mathcal{A}$ is a function $\alpha : \mathcal{A} \to [0, +\infty]$ such that $\alpha(\emptyset) = 0$.

It is clear again that the volume function Vol_d on the set of d-dimensional boxes is a set function.

Definition 3.39. Let α be a set function on a set system $\mathcal{A}$. We say that α is

- *monotone*, if $\quad A, B \in \mathcal{A},\ A \subseteq B \implies \alpha(A) \leq \alpha(B)$.
- *finitely superadditive*/σ*-superadditive*, if for any $A \in \mathcal{A}$, and any finite/countable collection $\{A_i\}_{i\in\mathcal{I}} \subseteq \mathcal{A}$ of pairwise disjoint sets,

$$\cup_{i\in\mathcal{I}} A_i \subseteq A \implies \sum_{i\in\mathcal{I}} \alpha(A_i) \leq \alpha(A). \tag{3.31}$$

- *finitely subadditive*/σ*-subadditive* if for any $A \in \mathcal{A}$, and any finite/countable collection $\{A_i\}_{i\in\mathcal{I}} \subseteq \mathcal{A}$ such that

$$A \subseteq \cup_{i\in\mathcal{I}} A_i, \qquad \text{we have} \qquad \alpha(A) \leq \sum_{i\in\mathcal{I}} \alpha(A_i).$$

- *finitely additive/σ-additive* if for every finite/countable family $\{A_i\}_{i\in\mathcal{I}} \subseteq \mathcal{A}$ of pairwise disjoint sets such that $\uplus_{i\in I} A_i \in \mathcal{A}$, we have

$$\alpha\left(\uplus_{i\in I} A_i\right) = \sum_{i\in\mathcal{I}} \alpha(A_i).$$

A σ-additive set function on $\mathcal{A}$ is called a *measure* on $\mathcal{A}$.

Remark 3.40. Note that $\cup_{i\in\mathcal{I}} A_i$ is not required to be an element of $\mathcal{A}$ in any of the above definitions. In particular, we only require additivity/σ-additivity of α on a disjoint union when the union itself is again an element of $\mathcal{A}$.

Exercise 3.41. Show that a set function α is finitely superadditive if and only if it is σ-superadditive, and either property implies monotonicity.

Remark 3.42. Note that in the definitions of subadditivity, we do not require the A_i to be pairwise disjoint.

According to the following proposition, a set function on a semi-ring that is finitely additive and σ-subadditive is also σ-additive, i.e., a measure. This is an important observation, as the first two properties are in general easier to verify than σ-additivity.

Proposition 3.43. A finitely additive set function on a semi-ring is monotone, σ-superadditive and finitely subadditive. Moreover, it is σ-subadditive if and only if it is σ-additive.

Proof. Let $\mathcal{S}$ be a semi-ring and $\alpha : \mathcal{S} \to [0, +\infty]$ be additive. Let $A, A_1, \ldots, A_r \in \mathcal{S}$ such that the A_i are pairwise disjoint and $\uplus_{i=1}^r A_i \subseteq A$. By Exercise 2.27, we have $A \setminus (\uplus_{i=1}^r A_i) = \uplus_{j=1}^m B_j$, with all $B_j \in \mathcal{S}$. Thus, $A = (\uplus_{i=1}^r A_i) \uplus \left(\uplus_{j=1}^m B_j\right)$, and by the additivity of α,

$$\alpha(A) = \sum_{i=1}^r \alpha(A_i) + \sum_{j=1}^m \alpha(B_j) \geq \sum_{i=1}^r \alpha(A_i).$$

Thus, α is finitely superadditive, and therefore also σ-superadditive and monotone.

Assume now that $A, A_1, \ldots, A_r \in \mathcal{S}$ are such that $A \subseteq \cup_{i=1}^r A_i$. Let $A_i' := A \cap A_i$, so that $A_i' \in \mathcal{S}$ and $A = \cup_{i=1}^r A_i'$. By Exercise 2.27, $\tilde{A}_i' := A_i' \setminus \left(\cup_{j=1}^{r-1} A_j'\right) = \uplus_{j=1}^{m_i} B_{i,j}$, where the $B_{i,j}$ are pairwise disjoint elements of $\mathcal{S}$. Then $A = \uplus_{i=1}^r \uplus_{j=1}^{m_i} B_{i,j}$, and

$$\alpha(A) = \sum_{i=1}^r \sum_{j=1}^{m_i} \alpha(B_{i,j}) \leq \sum_{i=1}^r \alpha(A_i') \leq \sum_{i=1}^r \alpha(A_i),$$

where the first inequality is due to finite superadditivity applied to $\uplus_{j=1}^{m_i} B_{i,j} \subseteq A'_i$, and the second one is due to monotonicity applied to $A'_i \subseteq A_i$.

If α is σ-subadditive and $\mathcal{A} \ni A = \uplus_{i\in\mathcal{I}} A_i$ for some countable family $\{A_i\}_{i\in\mathcal{I}} \subseteq \mathcal{A}$ then we have $\alpha(A) \leq \sum_{i\in\mathcal{I}} \alpha(A_i)$, and the converse inequality also holds due to the previously established σ-superadditivity. Thus, α is σ-additive.

Conversely, assume that α is σ-additive, and let $\mathcal{A} \ni A = \uplus_{i\in\mathcal{I}} A_i$ for some countable family $\{A_i\}_{i\in\mathcal{I}} \subseteq \mathcal{A}$. We can assume without loss of generality that $\mathcal{I} = N$, and by Exercise 2.27, we have $A_n \setminus \left(\cup_{i=1}^{n-1} A_i\right) = \uplus_{j\in J_n} B_{n,j}$ for some finite collection $\{B_{n,j}\}_{j\in J_n} \subseteq \mathcal{A}$ for every $n \geq 2$. For $n = 1$ we define $J_1 := \{1\}$, $B_{1,1} := A_1$. Thus, $A = \uplus_{n\in\mathbb{N}} \uplus_{j\in J_n} (A \cap B_{j,n})$, and $\alpha(A) = \sum_{n\in\mathbb{N}} \sum_{j\in J_n} \alpha(A \cap B_{j,n}) \leq \sum_{n\in\mathbb{N}} \alpha(A_n)$, where the equality is due to σ-additivity, and the inequality follows by superadditivity. □

Proposition 3.44. The volume function Vol is finitely additive on $\mathrm{Box}(\mathbb{R}^d)$.

Proof. Let $A, A_1, \ldots, A_n \in$
$boxx(\mathbb{R}^d)$ be such that $A = \cup_{k=1}^r A_k$, and $A_k \cap_{k\neq j} A_j = \emptyset$. For every k, we can write A_k as $A_k = \times_{i=1}^d [a_{k,i}, b_{k,i})$. Let us cut every box A_k into smaller boxes by the hyperplanes $H_{i,j} := \{\mathbf{x} \in \mathbb{R}^d : x_i = a_{j,i}\}$ for every $i \in [d]$ and $j \in [r]$. This way, we get new boxes $A_{k,i,j}$ (some of them may be empty) such that $A_k = \uplus_{i\in[d],j\in[r]} A_{k,i,j}$, and it is easy to see that $\mathrm{Vol}(A_k) = \sum_{i\in[d],j\in[r]} \mathrm{Vol}(A_{k,i,j})$ for every $k \in [r]$, and $\mathrm{Vol}(A) = \sum_{k\in[r],i\in[d],j\in[r]} \mathrm{Vol}(A_{k,i,j})$, from which the statement follows. □

The idea of the above proof can be used to show that the product of additive set functions on a semi-ring is again additive.

Definition 3.45. For every $i = 1, \ldots, r$, let $\mathcal{A}_i \subseteq \mathcal{P}(\mathcal{X}_i)$ be a set system, and $\alpha_i : \mathcal{A}_i \to [0, +\infty]$ be a set function. The *product* $\times_{i=1}^r \alpha_i : (\times)_{i=1}^r \mathcal{A}_i \to [0, +\infty]$ of these set functions is defined as

$$\left(\times_{i=1}^r \alpha_i\right)(A_1 \times \ldots \times A_r) := \alpha_1(A_1) \cdot \ldots \cdot \alpha_r(A_r), \qquad A_i \in \mathcal{A}_i,\ i \in [r].$$

Example 3.46. For every $d \in \mathbb{N}$, $\mathrm{Vol}_d = \times_{i=1}^d \mathrm{Vol}_1$, i.e., the volume function Vol_d on the d-dimensional boxes $\mathrm{Box}(\mathbb{R}^d)$ is the d-fold product of the volume function Vol_1 (on $\mathrm{Box}(\mathbb{R})$) with itself.

Proposition 3.47. Let α_i be a finitely additive set function on a semi-ring $\mathcal{S}_i \subseteq \mathcal{P}(\mathcal{X}_i)$ for every $i \in [r]$. Then $\times_{i=1}^r \alpha_i$ is finitely additive on $(\times)_{i=1}^d \mathcal{S}_i$.

Proof. Let $A_1 \times \ldots \times A_r = \uplus_{i=1}^m A_{i,1} \times \ldots \times A_{i,r}$, where $A_k, A_{i,k} \in \mathcal{S}_i$ for all $k \in [r]$, $i \in [m]$. By Exercise 2.28, for every $k \in [r]$, there exist $\{B_{j,k}\}_{j=1}^{m_k} \subseteq \mathcal{S}_k$ such that $\cup_{i=1}^m A_{i,k} = \uplus_{j=1}^{m_k} B_{j,k}$, and for every $k \in [r]$, $i \in [m]$, there exists a $J_{i,k} \subseteq [m_k]$ such

that $A_{i,k} = \cup_{j \in J_{i,k}} B_{j,k}$. We can assume without loss of generality that all the above sets are non-empty, since otherwise the claim is trivial. Then

$$A_{i,1} \times \ldots \times A_{i,r} = \times_{k=1}^{r} \cup_{j \in J_{i,k}} B_{j,k} = \cup_{j_1 \in J_{i,1}} \ldots \cup_{j_r \in J_{i,r}} B_{j_1,1} \times \ldots \times B_{j_r,r}$$

from which

$$\begin{aligned} \left(\times_{k=1}^{d} \alpha_k\right)(A_{i,1} \times \ldots \times A_{i,r}) &= \prod_{k=1}^{d} \alpha_k \left(\cup_{j \in J_{i,k}} B_{j,k}\right) = \prod_{k=1}^{d} \sum_{j \in J_{i,k}} \alpha_k \left(B_{j,k}\right) \\ &= \sum_{j_1 \in J_{i,1}} \ldots \sum_{j_r \in J_{i,r}} \prod_{k=1}^{d} \alpha_k(B_{j_k,k}). \end{aligned} \tag{3.32}$$

The assumption that the $A_{i,1} \times \ldots \times A_{i,r}$ are disjoint for different i's implies that the index sequences $J_{i,1} \times \ldots \times J_{i,r}$ are also disjoint for different i's. Hence,

$$\begin{aligned} A_1 \times \ldots \times A_r &= \cup_{i=1}^{m} \cup_{j_1 \in J_{i,1}} \ldots \cup_{j_r \in J_{i,r}} B_{j_1,1} \times \ldots \times B_{j_r,r} \\ &= \cup_{j_1=1}^{m_1} \ldots \cup_{j_r=1}^{m_r} B_{1,j_1} \times \ldots \times B_{r,j_r} \\ &= \left(\cup_{j_1=1}^{m_1} B_{1,j_1}\right) \times \ldots \times \left(\cup_{j_r=1}^{m_r} B_{r,j_r}\right). \end{aligned}$$

From this,

$$\begin{aligned} \left(\times_{k=1}^{r} \alpha_i\right)(A_1 \times \ldots \times A_r) &= \prod_{k=1}^{r} \alpha_k \left(\cup_{j_k=1}^{M} B_{k,j_k}\right) = \prod_{k=1}^{r} \sum_{j_k=1}^{m_k} \alpha_i(B_{k,j_k}) \\ &= \sum_{j_1=1}^{m_1} \ldots \sum_{j_r=1}^{m_r} \alpha_1(B_{1,j_1}) \cdot \ldots \cdot \alpha_r(B_{r,j_r}) \\ &= \sum_{i=1}^{m} \sum_{j_1 \in J_{i,1}} \ldots \sum_{j_r \in J_{i,r}} \prod_{k=1}^{d} \alpha_k(B_{j_k,k}). \end{aligned} \tag{3.33}$$

Comparing (3.32) and (3.33), we get the assertion. □

Remark 3.48. Combining Example 3.46 and Proposition 3.47 gives an alternative proof of the additivity of the d-dimensional volume function on boxes, Proposition 3.44.

Proposition 3.49. The volume function Vol is σ-subadditive on $\mathrm{Box}(\mathbb{R}^d)$.

Proof. Let $A \in \mathrm{Box}(\mathbb{R}^d)$ and $\{A_n\}_{n \in \mathbb{N}} \subset \mathrm{Box}(\mathbb{R}^d)$ be such that $A \subseteq \cup_{n \in \mathbb{N}} A_n$. We need to show that $\mathrm{Vol}(A) \leq \sum_{n \in \mathbb{N}} \mathrm{Vol}(A_n)$. Obviously, if $\sum_{n \in \mathbb{N}} \mathrm{Vol}(A_n) = +\infty$ then there is nothing to show, and hence for the rest we assume the contrary.

For every $\varepsilon > 0$, we can find $B_\varepsilon \in \mathrm{Box}(\mathbb{R}^d)$ such that $\overline{A}_\varepsilon \subseteq A$, and

$$\mathrm{Vol}(A_\varepsilon) \leq \mathrm{Vol}(A) \leq \mathrm{Vol}(A_\varepsilon) + \varepsilon. \tag{3.34}$$

Moreover, for every $n \in \mathbb{N}$, we can find $A_{n,\varepsilon} \in \mathrm{Box}(\mathbb{R}^d)$ such that $A_n \subset A^\circ_{n,\varepsilon}$, and

$$\mathrm{Vol}(A_n) \leq \mathrm{Vol}(A_{n,\varepsilon}) < \mathrm{Vol}(A_n) + \frac{\varepsilon}{2^n}. \tag{3.35}$$

Then we have $\overline{A}_\varepsilon \subseteq \cup_{n\in\mathbb{N}} A^\circ_{n,\varepsilon}$, i.e., the open sets $\{A^\circ_{n,\varepsilon}\}_{n\in\mathbb{N}}$ form an open cover of the compact set $\overline{A}_\varepsilon$. Thus, there exists a finite subcover, i.e., $n_1, \ldots, n_r \in \mathbb{N}$ such that $A_\varepsilon \subseteq \cup_{i=1}^r A^\circ_{n_i,\varepsilon} \subseteq \cup_{i=1}^r A_{n_i,\varepsilon}$. From this we obtain

$$\mathrm{Vol}(A) - \varepsilon \leq \mathrm{Vol}(A_\varepsilon) \leq \sum_{i=1}^r \mathrm{Vol}(A_{n_i,\varepsilon}) \leq \sum_{n\in\mathbb{N}} \mathrm{Vol}(A_{n,\varepsilon}) \leq \sum_{n\in\mathbb{N}} \left(\mathrm{Vol}(A_n) + \frac{\varepsilon}{2^n}\right),$$

where the first inequality is due to (3.34), the second inequality is due to the finite subadditivity of the volume function on $\mathrm{Box}(\mathbb{R}^d)$ (Corollary 3.50), the third inequality is trivial, and the last inequality is by (3.35). This gives

$$\mathrm{Vol}(A) - \varepsilon \leq \varepsilon + \sum_{n\in\mathbb{N}} \mathrm{Vol}(A_n),$$

and since this holds for any $\varepsilon > 0$, the assertion follows. □

Propositions 3.44, 3.49 and 3.43 yield immediately the following:

Corollary 3.50. The volume function Vol is a measure on the semi-ring $\mathrm{Box}(\mathbb{R}^d)$ of d-dimensional boxes.

3.5 Outer measures and the Carathéodory extension

Now, let us continue with our original problem of extending the volume function. Since it adds no extra difficulty, we will follow an abstract general approach, of which our original problem will be a special case. Hence, we will consider the general problem of extending a set function on some set system to a measure on a σ-algebra that contains the original set system.

Recall the definition of the Lebesgue outer measure in Definition **??**. The following is a generalization of that concept:

Definition 3.51. Let $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$ be a set system, and $\alpha : \mathcal{S} \to [0, +\infty]$ be a set function (recall that $\emptyset \in \mathcal{S}$ and $\alpha(\emptyset) = 0$ by definition). The *outer measure* generated by α is the set function $\alpha^* : \mathcal{P}(\mathcal{X}) \to [0, +\infty]$ given by

$$\alpha^*(A) := \inf\left\{ \sum_{n\in\mathbb{N}} \alpha(A_n) : (A_n)_{n\in\mathbb{N}} \subseteq \mathcal{S},\, n \in \mathbb{N},\, A \subseteq \cup_{n\in\mathbb{N}} A_n, \right\}, \qquad A \subseteq \mathcal{X}. \tag{3.36}$$

Remark 3.52. Note that the outer measure is defined for every subset of the basis set $\mathcal{X}$.

Example 3.53. The Lebesgue outer measure λ^* given in Definition **??** is the outer measure (in the sense of Definition 3.51) generated by the volume function λ on $\mathcal{S} = \mathrm{Box}(\mathbb{R}^d)$.

Exercise 3.54. Let α be an additive set function on a semi-ring $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$. Then

$$\alpha^*(A) = \inf\left\{\sum_{n\in\mathbb{N}} \alpha(A_n) : (A_n)_{n\in\mathbb{N}} \subseteq \mathcal{S},\, n \in \mathbb{N},\, A_n \cap_{n\neq m} A_m = \emptyset,\, A \subseteq \cup_{n\in\mathbb{N}} A_n,\right\} \tag{3.37}$$

for all $A \in \mathcal{P}(\mathcal{X})$, i.e., it is enough to consider disjoint covers in the definition of the outer measure.

Solution: Hidden.

It is easy to see that any outer measure is monotone:

Exercise 3.55. Show that the outer measure α^* corresponding to some set function α has the monotonicity property $A \subseteq B \Longrightarrow \alpha^*(A) \leq \alpha^*(B)$.

The outer measure in general need not be σ-additive or even finitely additive on any σ-algebra other than $\{\emptyset, \mathcal{X}\}$. However, it has the weaker property of σ-subadditivity:

Definition 3.56. A set function α is *σ-subadditive* on a set system $\mathcal{S}$ if

$$A \in \mathcal{S},\ A_n \in \mathcal{S},\, n \in \mathbb{N},\quad A \subseteq \cup_{n\in\mathbb{N}} A_n \quad\Longrightarrow\quad \alpha(A) \leq \sum_{n\in\mathbb{N}} \alpha(A_n).$$

Proposition 3.57. Let α^* be the outer measure corresponding to some set function α on a set system $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$. Then α^* is σ-subadditive on $\mathcal{P}(\mathcal{X})$.

Proof. Let $A \subseteq \mathcal{X}$, $A_n \subseteq \mathcal{X}$, $n \in \mathbb{N}$, be such that $A \subseteq \cup_{n\in\mathbb{N}} A_n$. We have to show that $\alpha^*(A) \leq \sum_{n\in\mathbb{N}} \alpha^*(A_n)$. It is obviously true if $\sum_{n\in\mathbb{N}} \alpha^*(A_n) = +\infty$, so for the rest we assume the contrary. Then $\alpha^*(A_n) < +\infty$ for all $n \in \mathbb{N}$. Hence, by definition, for every $\varepsilon > 0$ there exist $A_{n,\varepsilon,k} \in \mathcal{S}$, $k \in \mathbb{N}$, such that $A_n \subseteq \cup_{k\in\mathbb{N}} A_{n,\varepsilon,k}$, and

$$\sum_{k\in\mathbb{N}} \alpha(A_{n,\varepsilon,k}) < \alpha^*(A_n) + \frac{\varepsilon}{2^n}.$$

Then $A \subseteq \cup_{n\in\mathbb{N}} \cup_{k\in\mathbb{N}} A_{n,\varepsilon,k}$, and thus

$$\alpha^*(A) \leq \sum_{n,k\in\mathbb{N}} \alpha(A_{n,\varepsilon,k}) < \sum_{n\in\mathbb{N}} \left(\alpha^*(A_n) + \frac{\varepsilon}{2^n}\right) = \sum_{n\in\mathbb{N}} \alpha^*(A_n) + \varepsilon.$$

Since the inequality between the leftmost and the rightmost terms in the above line holds for all $\varepsilon > 0$, we get the desired inequality $\alpha^*(A) \leq \sum_{n\in\mathbb{N}} \alpha^*(A_n)$. □

The property of σ-subadditivity turns out to be so important in studying the extension of set functions into measures that we give σ-subadditive set functions a name.

Definition 3.58. A set σ-subadditive set function β on some set system is called an *(abstract) outer measure* or simply an outer measure.

Remark 3.59. Note that the extension process given in Definition 3.51 is one way of defining outer measures, but there are outer measures not given by this extension procedure. We use the terminology "abstract outer measure" when we want to emphasize this.

Remark 3.60. Note that an (abstract) outer measure is automatically monotone.

Note that by definition, if α is a set function on a set system $\mathcal{S}$ then

$$\alpha^*\big|_{\mathcal{S}} \leq \alpha, \quad \text{i.e., } \forall A \in \mathcal{S}: \quad \alpha^*(A) \leq \alpha(A). \tag{3.38}$$

(Indeed, this follows by taking the trivial cover $A \subseteq A \cup \emptyset \cup \emptyset \cup \ldots$) In general, $\alpha^*\big|_{\mathcal{S}} = \alpha$ need not hold, i.e., α^* need not be an extension of α. It is clear that if we want α^* to be an extension of α then they should have the same properties on the original set system $\mathcal{S}$. As it turns out, σ-subadditivity is the decisive property in this respect:

Proposition 3.61. Let α be a set function on a set system $\mathcal{S}$. Then α^* is an extension of α, i.e., $\alpha^*\big|_{\mathcal{S}} = \alpha$, if and only if α is σ-subadditive on $\mathcal{S}$.

Proof. Since α^* is σ-subadditive on $\mathcal{P}(\mathcal{X})$, it is also σ-subadditive on $\mathcal{S}$, and hence σ-subadditivity of α on $\mathcal{S}$ is a necessary condition for α^* to be an extension of α.

Assume now that α is σ-subadditive on $\mathcal{S}$ and let $A \in \mathcal{S}$. Then for any $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{S}$ such that $A \subseteq \cup_{n\in\mathbb{N}} A_n$, we have $\alpha(A) \leq \sum_{n\in\mathbb{N}} \alpha(A_n)$. Taking the infimum over all such covers, we get $\alpha(A) \leq \alpha^*(A)$. Since the converse inequality is trivial (see (3.38)), the assertion follows. □

Next, we want to find a concept of measurability that will eventually lead to a σ-additive extension of the volume function to a σ-algebra that contains $\mathrm{Box}(\mathbb{R}^d)$. The following definition may seem less intuitive at the first sight than the concept of Jordan measurability, but it turns out to be the right one to achieve our goal.

Definition 3.62. Let α be a set function on a set system $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$, and α^* the generated outer measure. We say that a set $A \subseteq \mathcal{X}$ is α^*-*measurable*, if

$$\forall T \subseteq \mathcal{X}: \quad \alpha^*(T) = \alpha^*(T \setminus A) + \alpha^*(T \cap A). \tag{3.39}$$

We denote the set of α^*-measurable sets by $\mathcal{M}(\alpha^*)$.

More generally, if $\beta : \mathcal{P}(\mathcal{X}) \to [0, +\infty]$ is an abstract outer measure, then we say that $A \subseteq \mathcal{X}$ is β-*measurable*, if

$$\forall T \subseteq \mathcal{X}: \quad \beta(T) = \beta(T \setminus A) + \beta(T \cap A).$$

We denote the set of β-measurable sets by $\mathcal{M}(\beta)$.

Remark 3.63. Note that we have now two different concepts of measurability of a set. In Definition **??**, it simply meant that the set is an element of a given σ-algebra, without any reference to a measure or an outer measure. On the contrary, in the above Definition 3.62, there is no mentioning of a σ-algebra, and measurability is defined with respect to an outer measure. It is important to keep in mind this difference. The meaning in which measurability is applied to a set should always be clear from the context.

Remark 3.64. Note that $\alpha^*(T) \leq \alpha^*(T \setminus A) + \alpha^*(T \cap A)$ holds for any $A, T \subseteq \mathcal{X}$ due to the σ-subadditivity of α^* (Proposition 3.57), and hence (3.39) is equivalent to

$$\forall T \subseteq \mathcal{X}: \quad \alpha^*(T) \geq \alpha^*(T \setminus A) + \alpha^*(T \cap A). \tag{3.40}$$

Remark 3.65. Note that we are looking for a σ-additive extension of α, and (3.39) expresses a certain additivity criterion that, however, seems much weaker at the first sight then the σ-additivity we are looking for.

On the other hand, it also seems stronger in the sense that it is required to hold for every $T \in \mathcal{P}(\mathcal{X})$, whereas we don't expect α to have a (σ-)additive extension to the whole of $\mathcal{P}(\mathcal{X})$. However, it turns out that it is enough to check (3.40) for sets T in the original set system $\mathcal{S}$.

Lemma 3.66. In the setting of Definition 3.62, $A \subseteq \mathcal{X}$ is α^*-measurable if and only if

$$\forall B \in \mathcal{S}: \quad \alpha(B) \geq \alpha^*(B \setminus A) + \alpha^*(B \cap A). \tag{3.41}$$

Proof. Necessity: If A is α^*-measurable then we have $\alpha^*(B \setminus A) + \alpha^*(B \cap A) = \alpha^*(B) \leq \alpha(B)$ for any $B \subseteq \mathcal{X}$ (not only for $B \in \mathcal{S}$), where the second inequality is trivial from the definition of the outer measure (see (3.38)).

To prove sufficiency, we have to show that (3.41) implies that $\alpha^*(T) \geq \alpha^*(T \setminus A) + \alpha^*(T \cap A)$ for any $T \subseteq \mathcal{X}$ (see Remark 3.64). Note that this inequality is trivial if $\alpha^*(T) = +\infty$, and hence for the rest we assume the contrary. Then, by the definition of α^*, for every $\varepsilon > 0$, there exist $B_{n,\varepsilon} \in \mathcal{S}$, $n \in \mathbb{N}$, such that $T \subseteq \cup_{n\in\mathbb{N}} B_{n,\varepsilon}$, and $\sum_{n\in\mathbb{N}} \alpha(B_{n,\varepsilon}) < \alpha^*(T) + \varepsilon$. Thus,

$$\begin{aligned}\varepsilon + \alpha^*(T) > \sum_{n\in\mathbb{N}} \alpha(B_{n,\varepsilon}) &\geq \sum_{n\in\mathbb{N}} (\alpha^*(B_{n,\varepsilon} \setminus A) + \alpha^*(B_{n,\varepsilon} \cap A)) \\ &\geq \alpha^*(T \setminus A) + \alpha^*(T \cap A),\end{aligned}$$

as required, where the second inequality follows from the assumption (3.41) applied to each $B_{n,\varepsilon}$ and A, the last inequality is due to the σ-subadditivity of α^*. □

Remark 3.67. Note that the criterion (3.41) is not only a weakening of (3.39) in the sense that we only require the inequality to hold for sets B in the original set system $\mathcal{S}$, but also that we replace $\alpha^*(B)$ with $\alpha(B)$ in the upper bound, and $\alpha(B)^* \leq \alpha(B)$ for any $B \in \mathcal{S}$ by definition.

Corollary 3.68. Let α be an additive set function on a semi-ring $\mathcal{S}$. Then $\mathcal{S} \subseteq \mathcal{M}(\alpha^*)$, i.e., all elements of $\mathcal{S}$ are α^*-measurable.

Proof. Let $A \in \mathcal{S}$. Then the semi-ring property implies that for every $B \in \mathcal{S}$, we have $B \cap A \in \mathcal{S}$, and $B \setminus A = \uplus_{i=1}^r A_i$ for some pairwise disjoint $A_1, \ldots, A_r \in \mathcal{S}$. Then

$$\begin{aligned}\alpha(B) = \alpha(B \cap A) + \sum_{i=1}^r \alpha(A_i) &\geq \alpha^*(B \cap A) + \sum_{i=1}^r \alpha^*(A_i) \\ &\geq \alpha^*(B \cap A) + \alpha^*(\uplus_{i=1}^r A_i),\end{aligned}$$

where equality follows by the additivity of α, the second inequality is due to $\alpha \geq \alpha^*$, and the last inequality is due to the (σ-)subadditivity of α^*. Hence, by Lemma 3.66, A is α^*-measurable. □

Next, we verify that $M(\alpha^*)$ and $\alpha^*\big|_{\mathcal{M}(\alpha^*)}$ have the desired properties, i.e., the former is a σ-algebra, and the latter is a measure on it. This is true in somewhat more generality, and we state it so:

Theorem 3.69. Let $\beta : \mathcal{P}(\mathcal{X}) \to [0, +\infty]$ be an abstract outer measure (i.e., a σ-subadditive set function). Then the set of *β-measurable sets*

$$\mathcal{M}(\beta) := \{A \in \mathcal{P}(\mathcal{X}) : \forall T \subseteq \mathcal{X} : \quad \beta(T) = \beta(T \setminus A) + \beta(T \cap A)\}$$

is a σ-algebra, and $\beta\big|_{M(\beta)}$ is a measure.

Proof. For any $A \subseteq \mathcal{X}$, $A \cap \emptyset = \emptyset$ and $A \setminus \emptyset = A$ implies

$$\beta(A \cap \emptyset) + \beta(A \setminus \emptyset) = \beta(A),$$

and thus $\emptyset \in M(\beta)$. Likewise, $A \cap \mathcal{X} = A$ and $A \setminus \mathcal{X} = \emptyset$ yields

$$\beta(A \cap \mathcal{X}) + \beta(A \setminus \mathcal{X}) = \beta(A),$$

and thus $\mathcal{X} \in \mathcal{M}(\beta)$.

Now let $A \in \mathcal{M}(\beta)$. Note that for any $B \subseteq \mathcal{X}$, $B \setminus (X \setminus A) = B \cap A$ and $B \cap (\mathcal{X} \setminus A) = B \setminus A$, and hence

$$\beta(B) = \beta(B \setminus A) + \beta(B \cap A) = \beta(B \cap (\mathcal{X} \setminus A)) + \beta(B \setminus (X \setminus A)),$$

where the first equality is due to $A \in \mathcal{M}(\beta)$. Hence, $\mathcal{X} \setminus A \in M(\beta)$.

We prove that $\mathcal{M}(\beta)$ is closed under countable unions in a series of steps. First, we show that it is closed under finite unions. Let $A_1, A_2 \in \mathcal{M}(\beta)$, and $B \subseteq \mathcal{X}$ arbitrary. Then

$$\begin{aligned}\beta(B) &= \beta(B \setminus A_1) + \beta(B \cap A_1) \\ &= \beta\left((B \setminus A_1) \setminus A_2\right) + \beta\left((B \setminus A_1) \cap A_2\right) + \beta(B \cap A_1),\end{aligned}$$

where the first equality is due to $A_1 \in \mathcal{M}(\beta)$, and the second one is due to $A_2 \in \mathcal{M}(\beta)$. Now, observe that $(B \setminus A_1) \setminus A_2 = B \setminus (A_1 \cup A_2)$, and $((B \setminus A_1) \cap A_2) \uplus (B \cap A_1) = B \cap (A_1 \cup A_2)$. Using the subadditivity of β, we get that the above can be further lower bounded as

$$\beta(B) \geq \beta\left(B \setminus (A_1 \cup A_2)\right) + \beta\left(B \cap (A_1 \cup A_2)\right).$$

Since this is true for every $B \subseteq \mathcal{X}$, and the converse inequality is trivial by subadditivity, we get $A_1 \cap A_2 \in \mathcal{M}(\beta)$.

Next, we show that for any pairwise disjoint $A_1, \ldots, A_r \in \mathcal{M}(\beta)$, and any $T \subseteq \mathcal{X}$, we have

$$\beta(T \cap (\uplus_{i=1}^{r} A_i)) = \sum_{i=1}^{r} \beta(T \cap A_i). \tag{3.42}$$

Note that it is enough to prove this for $r = 2$, and then use iteration for larger r. Moreover, we may assume that $T \subseteq \uplus_{i=1}^{r} A_i$, since otherwise we can replace T with $T \cap (\uplus_{i=1}^{r} A_i)$. Under these assumption, we have

$$\beta\left(T \cap (A_1 \uplus A_2)\right) = \beta(T) = \beta(T \setminus A_1) + \beta(T \cap A_1) = \beta(T \cap A_2) + \beta(T \cap A_1),$$

where the second inequality is due to $A_1 \in \mathcal{M}(\beta)$, and the last one is due to $T \setminus A_1 = T \cap A_2$.

Now we prove that $\mathcal{M}(\beta)$ is closed under countable unions. Indeed, let $A_n \in \mathcal{M}(\beta)$, $n \in \mathbb{N}$. For every $n \in \mathbb{N}$, let $\tilde{A}_n := A_n \setminus \left(\cup_{i=1}^{n-1} A_i\right)$, so that the $\tilde{A}_i$ are pairwise disjoint, and $\cup_{n\in\mathbb{N}} A_i = \uplus_{n\in\mathbb{N}} \tilde{A}_n$. Let $T \subseteq \mathcal{X}$ be arbitrary. For every $r \in \mathbb{N}$, we have

$$\begin{aligned}
\beta(T) &= \beta\left(T \setminus (\cup_{n=1}^{r} \tilde{A}_n)\right) + \beta\left(T \cap (\cup_{n=1}^{r} \tilde{A}_n))\right) \\
&\geq \beta\left(T \setminus (\cup_{n=1}^{+\infty} \tilde{A}_n)\right) + \beta\left(T \cap (\cup_{n=1}^{r} \tilde{A}_n))\right) \\
&= \beta\left(T \setminus (\cup_{n=1}^{+\infty} \tilde{A}_n)\right) + \sum_{n=1}^{r} \beta\left(T \cap A_n\right),
\end{aligned}$$

where in the first equality we used that $\mathcal{M}(\beta)$ is closed under finite unions, and this $\cup_{n=1}^{r} \tilde{A}_n \in \mathcal{M}(\beta)$; the inequality is due to the monotonicity of β, and the last equality follows by (3.42). Taking the limit $r \to +\infty$, we get

$$\begin{aligned}
\beta(T) &\geq \beta\left(T \setminus (\cup_{n=1}^{+\infty} \tilde{A}_n)\right) + \sum_{n=1}^{+\infty} \beta\left(T \cap \tilde{A}_n\right) \\
&\geq \beta\left(T \setminus (\cup_{n=1}^{+\infty} \tilde{A}_n)\right) + \beta\left(T \cap (\cup_{n=1}^{+\infty} \tilde{A}_n)\right),
\end{aligned} \tag{3.43}$$

where the last inequality follows by the σ-subadditivity of β, due to $T \cap (\cup_{n=1}^{+\infty} \tilde{A}_n) = \cup_{n=1}^{+\infty} (T \cap \tilde{A}_n)$.

Finally, we show that β is σ-additive on $\mathcal{M}(\beta)$. To this end, let $A_n \in \mathcal{M}(\beta)$, $n \in \mathbb{N}$, be pairwise disjoint sets. Writing $T := \cup_{n\in\mathbb{N}} A_n$ in (3.43), we get

$$\beta\left(\cup_{n\in\mathbb{N}} A_n\right) \geq \sum_{n=1}^{+\infty} \beta\left(A_n\right),$$

and the converse inequality is trivial by the σ-subadditivity of β. □

Combining the above, we arrive at the main result of this section:

Theorem 3.70. (Carathéodory extension theorem)
Let α be a measure on a semi-ring $\mathcal{S}$, and α^* the generated outer measure. Then $\mathcal{M}(\alpha^*)$ is a σ-algebra that contains $\mathcal{S}$, and

$$\alpha^*\Big|_{\mathcal{M}(\alpha^*)} \quad \text{is a measure, such that} \quad \alpha^*\Big|_{\mathcal{S}} = \alpha.$$

That is, α^* is an extension of α to a measure on the σ-algebra $\mathcal{M}(\alpha^*)$.

Now let us return to the special case where $\mathcal{S} = \mathrm{Box}(\mathbb{R}^d)$ is the semi-ring of boxes in $\mathbb{R}^d$, and $\alpha = \mathrm{Vol}$ is the volume function.

Definition 3.71. Measurable sets with respect to the Lebesgue outer measure Vol^*, in the sense of Definition 3.62, are called *Lebesgue measurable*. We use the notation $\Lambda(\mathbb{R}^d) := \mathcal{M}(\mathrm{Vol}^*)$ for the collection of Lebesgue measurable sets in $\mathbb{R}^d$. The Lebesgue outer measure restricted to $\Lambda(\mathbb{R}^d)$ is a measure by Theorem 3.69, that we call the *Lebesgue measure*.

It is natural to ask how the set of Lebesgue measurable sets is related to the set of boxes, and what properties Lebesgue measurable sets have. We will answer these and various related questions in the next section.

3.6 Properties of the Carathéodory extension

Definition 3.72. We say that a measure μ on a σ-algebra $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ is *complete*, if

$$A \in \mathcal{A} \text{ and } \mu(A) = 0 \implies \forall B \subseteq A: \quad B \in \mathcal{A},$$

i.e., any subset of a measurable set of measure zero is also measurable (and hence is of measure 0).

Remark 3.73. Note that in the above definition, measurability refers to being an element of a fixed σ-algebra.

Proposition 3.74. Let $\beta : \mathcal{P}(\mathcal{X}) \to [0, +\infty]$ be an abstract outer measure. Then every $A \subseteq \mathcal{X}$ with $\beta(A) = 0$ is β-measurable. In particular, $\beta\big|_{\mathcal{M}(\beta)}$ is a complete measure.

Proof. Let $\beta(A) = 0$. Then for any $T \subseteq \mathcal{X}$, monotonicity of β implies $\beta(T \setminus A) \leq \beta(T)$, and $\beta(T \cap A) \leq \beta(A) = 0$. Hence,

$$\beta(T) \geq \beta(T \setminus A) + \beta(A),$$

i.e., A is measurable. From this, the rest of the statement follows. □

As the following proposition shows, measurable sets in the Carathéodory extension of a measure on a semi-ring can be arbirarily well approximated by finite disjoint unions of elements of the semi-ring. In particular, it gives a generalization of Exercises 2.53 and 2.55.

Proposition 3.75. Let μ be a measure on a semi-ring $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$, and let $A \subseteq \mathcal{P}(\mathcal{X})$ be such that $\mu^*(A) < +\infty$. Then $A \in \mathcal{M}(\mu^*)$ if and only if for any $\varepsilon > 0$ there exist finitely many disjoint elements $B_{\varepsilon,1}, \ldots, B_{\varepsilon,n_\varepsilon} \in \mathcal{S}$ such that

$$\mu^*(A \triangle B_\varepsilon) < \varepsilon, \qquad \text{where} \qquad B_\varepsilon := \cup_{k=1}^{n_\varepsilon} B_{\varepsilon,k}. \tag{3.44}$$

Proof. "only if": By Exercise 3.54, for any $\varepsilon > 0$, there exists a sequence of disjoint elements $B_{\varepsilon,k} \in \mathcal{S}$, $k \in \mathbb{N}$, such that

$$A \subseteq \cup_{k=1}^{+\infty} B_{\varepsilon,k}, \qquad \text{and} \qquad \mu^*(A) \leq \sum_{k=1}^{+\infty} \mu^*(B_{\varepsilon,k}) < \mu^*(A) + \varepsilon.$$

Finiteness of the infinite sum implies that there exists an n_ε such that $\sum_{k=n_\varepsilon+1}^{+\infty} \mu^*(B_{\varepsilon,k}) < \varepsilon$. Now, with $B_\varepsilon := \cup_{k=1}^{n_\varepsilon} B_{\varepsilon,k}$, we have

$$\mu^*\left(A \setminus B_\varepsilon\right) \leq \mu^*\left(\cup_{k=n_\varepsilon+1}^{+\infty} B_{\varepsilon,k}\right) = \sum_{k=n_\varepsilon+1}^{+\infty} \mu^*(B_{\varepsilon,k}) < \varepsilon,$$

and

$$\begin{aligned} \mu^*\left(B_\varepsilon \setminus A\right) \leq \mu^*\left(\cup_{k=1}^{+\infty} B_{\varepsilon,k} \setminus A\right) &= \mu^*\left(\cup_{k=1}^{+\infty} B_{\varepsilon,k}\right) - \mu^*(A) \\ &= \sum_{k=1}^{+\infty} \mu^*(B_{\varepsilon,k}) - \mu^*(A) < \varepsilon. \end{aligned}$$

In the equalities above, we used that $A \in \mathcal{M}(\mu^*)$ and $B_{\varepsilon,k} \in \mathcal{M}(\mu^*)$, $k \in \mathbb{N}$, and that μ^* is a measure on $\mathcal{M}(\mu^*)$. Putting it together,

$$\mu^*\left(\left(A \setminus B_\varepsilon\right) \cup \left(B_\varepsilon \setminus A\right)\right) = \mu^*\left(B_\varepsilon \setminus A\right) + \mu^*\left(B_\varepsilon \setminus A\right) \leq 2\varepsilon.$$

Changing ε to $\varepsilon/2$ in the above argument yields the desired bound in (3.44).

"if": By Lemma 3.66, it is sufficient to prove that (3.41) holds for any $B \in \mathcal{S}$. Let $B_{\varepsilon,1}, \ldots, B_{\varepsilon,n_\varepsilon}$ be as in the assumption. We have

$$\begin{aligned} B \cap A &= \underbrace{\left[B \cap \left(A \setminus B_\varepsilon\right)\right]}_{\subseteq A \setminus B_\varepsilon} \cup \underbrace{\left[B \cap A \cap B_\varepsilon\right]}_{\subseteq B \cap B_\varepsilon} \\ B \setminus A &\subseteq \left[B \setminus B_\varepsilon\right] \cup \underbrace{\left[B \cap \left(B_\varepsilon \setminus A\right)\right]}_{\subseteq B_\varepsilon \setminus A}, \end{aligned}$$

whence

$$\begin{aligned} \mu^*(B \cap A) + \mu^*(B \setminus A) &\leq \underbrace{\mu^*(A \setminus B_\varepsilon)}_{<\varepsilon} + \mu^*(B \cap B_\varepsilon) + \mu^*(\underbrace{B \setminus B_\varepsilon}_{=\cup_{i=1}^m C_i}) + \underbrace{\mu^*(B_\varepsilon \setminus A)}_{<\varepsilon} \\ &< 2\varepsilon + \mu^*\left(\cup_{k=1}^{n_\varepsilon} B \cap B_{\varepsilon,k}\right) + \mu^*\left(\cup_{i=1}^m C_i\right) \\ &= 2\varepsilon + \mu(B), \end{aligned}$$

where $C_1, \ldots, C_m$ are disjoint elements in $\mathcal{S}$ such that $B \setminus B_\varepsilon = \cup_{i=1}^m C_i$; see Exercise 2.27. The first inequality above follows from the subadditivity and monotonicity of μ^*, the second inequality is by assumption, and the equality in the end follows from the fact that on $\mathcal{S}$, μ is a measure, and $\mu^* = \mu$. Since the above holds for all $\varepsilon > 0$, we can conclude that $\mu^*(B \cap A) + \mu^*(B \setminus A) \leq \mu(B)$. $\square$

It is easy to see that appoximation by finitely many boxes in the above sense may not be possible if $\mu(A) = +\infty$; a simple example is given by the volume function on $\mathcal{S} = \mathrm{Box}(\mathbb{R})$ and $A := \uplus_{n\in\mathbb{N}}[2n, 2n+1)$. However, we still have the following:

Exercise 3.76. Let μ be a completely σ-finite measure on a semi-ring $\mathcal{S}$, and let $A \in \mathcal{M}(\mu^*)$. Show that for any $\varepsilon > 0$, there exist countably many disjoint boxes $(B_n)_{n\in\mathbb{N}} \subseteq \mathcal{S}$ such that

$$A \subseteq \uplus_{n\in\mathbb{N}} B_n, \quad \text{and} \quad \mu^*\left((\uplus_{n\in\mathbb{N}} B_n) \setminus A\right) < \varepsilon.$$

Solution: Hidden.

Further approximation properties are given by the following:

Lemma 3.77. Let μ be a measure on a semi-ring $\mathcal{S}$.

1. For every $A \subseteq \mathcal{X}$ with $\mu^*(A) < +\infty$, there exists an $\bar{A} \in \mathcal{S}_{\sigma\delta}$ such that

$$A \subseteq \bar{A}, \quad \text{and} \quad \mu^*(A) = \mu^*(\bar{A}).$$

If $A \in \mathcal{M}(\mu^*)$ then also $\mu^*(\bar{A} \setminus A) = 0$.

2. If μ is completely σ-finite then for every measurable $A \in \mathcal{M}(\mu^*)$, there exist $\bar{A} \in \mathcal{S}_{\sigma\delta}$ and $\tilde{A} \in \mathcal{S}_{\delta\sigma}$ such that

$$\tilde{A} \subseteq A \subseteq \bar{A}, \quad \text{and} \quad \mu^*(\bar{A} \setminus \tilde{A}) = 0.$$

Proof. 1. By definition, for every $\varepsilon > 0$, there exist $\{A_{\varepsilon,n}\}_{n\in\mathbb{N}} \subseteq \mathcal{S}$ such that $A \subseteq \cup_{n\in\mathbb{N}} A_{\varepsilon,n}$, and $\sum_{n\in\mathbb{N}} \mu(A_{\varepsilon,n}) < \mu^*(A) + \varepsilon$. Then

$$A \subseteq \bar{A} := \cap_{m\in\mathbb{N}} \cup_{n\in\mathbb{N}} A_{1/m,n}, \quad \text{and}$$
$$\mu^*(\bar{A}) \leq \mu^*\left(\cup_{n\in\mathbb{N}} A_{1/m,n}\right) \leq \sum_{n\in\mathbb{N}} \mu(A_{1/m,n}) < \mu^*(A) + 1/m, \qquad m \in \mathbb{N},$$

due to the monotonicity and the σ-subadditivity of μ^*. From this, $\mu^*(\bar{A}) = \mu^*(A)$. Since $\bar{A} \in \mathcal{M}(\mu^*)$, if also $A \in \mathcal{M}(\mu^*)$ then the additivity of μ^* on $\mathcal{M}(\mu^*)$ implies $\mu^*(\bar{A} \setminus A) = \mu^*(\bar{A}) - \mu^*(A) = 0$.

2. By assumption, we have a decomposition $\mathcal{X} = \uplus_{k\in\mathbb{N}} \mathcal{X}_k$, where $\mathcal{X}_k \in \mathcal{S}$ and $\mu(\mathcal{X}_k) < +\infty$ for all k. Let $A \in \mathcal{M}(\mu^*)$ so that also $A_k := A \cap \mathcal{X}_k \in \mathcal{M}(\mu^*)$. By the same argument as in the previous point, for every $k \in \mathbb{N}$, there exists an $\bar{A}_k \in \mathcal{S}_{\sigma\delta}$ such that $\mu^*(\bar{A}_k \setminus A_k) = 0$, and it is easy to see that we can take $\bar{A}_k$ such that $\bar{A}_k \subseteq \mathcal{X}_k$. Let $\bar{A} := \uplus_{k\in\mathbb{N}} \bar{A}_k$; then

$$\mu^*(\bar{A} \setminus A) = \mu^*\left(\uplus_{k\in\mathbb{N}} (\bar{A}_k \setminus A_k)\right) \leq \sum_{k\in\mathbb{N}} \mu^*(\bar{A}_k \setminus A_k) = 0.$$

By the above, we have $B \in \mathcal{S}_{\sigma\delta}$ such that $\mathcal{X} \setminus A \subseteq B$ and $\mu^*(B \setminus (\mathcal{X} \setminus A)) = 0$. Let $\tilde{A} := \mathcal{X} \setminus B$; then

$$\tilde{A} = \mathcal{X} \setminus B \subseteq \mathcal{X} \setminus (\mathcal{X} \setminus A) = A, \text{ and}$$
$$\mu^*(A \setminus \tilde{A}) = \mu^*(A \cap B) = \mu^*(B \setminus (\mathcal{X} \setminus A)) = 0.$$

Finally, $\mu^*(\bar{A} \setminus \tilde{A}) \leq \mu^*(\bar{A} \setminus A) + \mu^*(A \setminus \tilde{A}) = 0$. □

Corollary 3.78. Let μ be a completely σ-finite measure on a semi-ring $\mathcal{S}$. Then for every $A \subseteq \mathcal{X}$,

$$A \in \mathcal{M}(\mu^*) \iff \exists\ \tilde{A} \in \sigma(\mathcal{S}), \text{ and } A_0 \text{ with } \mu^*(A_0) = 0 \text{ s.t. } A = \tilde{A} \cup A_0. \tag{3.45}$$

Moreover, for any such decomposition, $\mu^*(\tilde{A}) \leq \mu^*(A) \leq \mu^*(\tilde{A}) + \mu^*(A_0) = \mu^*(\tilde{A})$ yields $\mu^*(A) = \mu^*(\tilde{A})$.

Remark 3.79. Note that by possibly replacing A_0 with $A_0 \setminus \tilde{A}$, A_0 can be chosen to be disjoint from $\tilde{A}$ in (3.45).

Theorem 3.80. Let μ be a measure on a semi-ring $\mathcal{S}$, and ν be a measure on a σ-algebra $\mathcal{A}$ such that $\sigma(\mathcal{S}) \subseteq \mathcal{A} \subseteq \mathcal{M}(\mu^*)$, and $\nu\big|_{\mathcal{S}} = \mu$.

1. $\nu^* \leq \mu^*$, in particular, $\mu^*(A) = 0 \Longrightarrow \nu^*(A) = 0$.
2. For every $A \in \mathcal{M}(\mu^*) \cap \mathcal{M}(\nu^*)$ with $\mu^*(A) < +\infty$, $\mu^*(A) = \nu^*(A)$.
3. If μ is completely σ-finite on $\mathcal{S}$ then $\mathcal{M}(\mu^*) = \mathcal{M}(\nu^*)$. In particular, $\nu = \mu^*\big|_{\mathcal{A}}$.

Proof. 1. Let $A \subseteq \mathcal{X}$. Since any countable $\mathcal{S}$-cover of A is also a countable $\mathcal{A}$-cover of A, we get

$$\begin{aligned}\nu^*(A) &= \inf\left\{\sum_{n\in\mathbb{N}} \mu(A_n) :\ A \subseteq \cup_{n\in\mathbb{N}} A_n,\ \{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}\right\} \\ &\leq \inf\left\{\sum_{n\in\mathbb{N}} \mu(A_n) :\ A \subseteq \cup_{n\in\mathbb{N}} A_n,\ \{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}\right\} = \mu^*(A).\end{aligned}$$

2. Due to σ-additivity, μ^* and ν^* coincide on countable disjoint unions of elements in $\mathcal{S}$. Since any countable union of elements in $\mathcal{S}$ can be written as a disjoint union (Lemma **??**), we get that μ^* and ν^* coincide on $\mathcal{S}_\sigma$. By Lemma 3.77, for any $A \subseteq \mathcal{X}$ with $\mu^*(A) < +\infty$, there exists a decreasing sequence in $\mathcal{S}_\sigma$, $A_1 \supseteq A_2 \supseteq \ldots \supseteq A$

such that $\mu^*(A) = \mu^*(\bar{A})$, where $\bar{A} := \cap_{n\in\mathbb{N}} A_n$. By the above, $\nu^*(A_n) = \mu^*(A_n)$ for all n. Due to monotone continuity of measures,

$$\mu^*(A) = \mu^*(\bar{A}) = \lim_{n\to+\infty} \mu^*(A_n) = \lim_{n\to+\infty} \nu^*(A_n) = \nu^*(\bar{A}).$$

Now, if A is μ^*-measurable then we also have $\mu^*(\bar{A} \setminus A) = \mu^*(\bar{A}) - \mu^*(A) = 0$, and thus also $\nu^*(\bar{A} \setminus A) = 0$, according to the previous point. Assume that A is also ν^*-measurable. Then

$$\begin{aligned}\mu^*(A) &= \mu^*(\bar{A} \setminus (\bar{A} \setminus A)) = \mu^*(\bar{A}) - \mu^*(\bar{A} \setminus A) = \mu^*(\bar{A}) \\ &= \nu^*(\bar{A}) = \nu^*(\bar{A}) - \nu^*(\bar{A} \setminus A) = \nu^*(A).\end{aligned}$$

3. By assumption, there exists a decomposition $\mathcal{X} = \cup_{k\in\mathbb{N}} \mathcal{X}_k$ with $\mathcal{X}_k \in \mathcal{S}$, $\mu(\mathcal{X}_k) < +\infty$ for all k. Let $A \in \mathcal{M}(\mu^*) \cap \mathcal{M}(\nu^*)$ and for every $k \in \mathbb{N}$, let $A_k := A \cap \mathcal{X}_k$. Then $A_k \in \mathcal{M}(\mu^*) \cap \mathcal{M}(\nu^*)$ and $\mu^*(A_k) < +\infty$ for all k, and hence by the previous point, $\mu^*(A) = \sum_{k\in\mathbb{N}} \mu^*(A_k) = \sum_{k\in\mathbb{N}} \nu^*(A_k) = \nu^*(A)$. That is, μ^* and ν^* coincide on $\mathcal{M}(\mu^*) \cap \mathcal{M}(\nu^*)$, and in particular, μ^* and ν coincide on $\mathcal{A}$.

Let $A \subseteq \mathcal{X}$ be arbitrary. We have seen that $\nu^*(A) \leq \mu^*(A)$; in particular, if $\nu^*(A) = +\infty$ then $\nu^*(A) = \mu^*(A)$. Assume next that $\nu^*(A) < +\infty$. Then for every $\varepsilon > 0$ there exists a countable $\mathcal{A}$-cover $(A_{\varepsilon,n})_{n\in\mathbb{N}} \subseteq \mathcal{A}$ such that $A \subseteq \cup_{n\in\mathbb{N}} A_{\varepsilon,n}$ and $\sum_{n\in\mathbb{N}} \nu(A_{\varepsilon,n}) < \nu^*(A) + \varepsilon$. By the above, $\mu^*(A_{\varepsilon,n}) = \nu(A_{\varepsilon,n}) < +\infty$ for all n. Hence, there exist $(B_{\varepsilon,n,k})_{k\in\mathbb{N}} \subseteq \mathcal{P}$ such that $A_{\varepsilon,n} \subseteq \cup_{k\in\mathbb{N}} B_{\varepsilon,n,k}$, and $\sum_{k\in\mathbb{N}} \mu(B_{\varepsilon,n,k}) < \mu^*(A_{\varepsilon,n}) + \varepsilon/2^n = \nu(A_{\varepsilon,n}) + \varepsilon/2^n$. Finally, $A \subseteq \cup_{n\in\mathbb{N}} \cup_{k\in\mathbb{N}} B_{\varepsilon,n,k}$, and $\mu^*(A) \leq \sum_{n\in\mathbb{N}} \sum_{k\in\mathbb{N}} \mu(B_{\varepsilon,n,k}) < \sum_{n\in\mathbb{N}} (\nu(A_{\varepsilon,n}) + \varepsilon/2^n) < \nu^*(A) + 2\varepsilon$. Since this holds for all $\varepsilon > 0$, we get $\mu^*(A) \leq \nu^*(A)$. We have already established the converse inequality, and thus we obtain $\mu^* = \nu^*$ on $\mathcal{S}(\mathcal{X})$. □

Corollary 3.81. A completely σ-finite measure on a semi-ring extends uniquely to a σ-finite measure on the σ-algebra generated by the semi-ring.

Corollary 3.82. The above theorem implies that first extending a completely σ-finite measure μ from a semi-ring $\mathcal{S}$ to any sub-σ-algebra of $\mathcal{M}(\mu^*)$ using the Carathéodory method, and then extending the so obtained measure again with the Carathéodory method, we do not get anything new. Formally,

$$\left(\mu^*\big|_{\mathcal{A}}\right)^* = \mu^*$$

for any σ-algebra $\mathcal{A}$ such that $\mathcal{S} \subseteq \mathcal{A} \subseteq \mathcal{M}(\mu^*)$.

3.7 Product measure

For set systems $\mathcal{A}_i$ on some sets $\mathcal{X}_i$, and non-negative set functions μ_i on $\mathcal{A}_i$, let us use the notations

$$\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n := \{A_1 \times \ldots \times A_n : A_i \in \mathcal{A}_i,\, i \in [n]\},$$

and

$$\mu_1 \times \ldots \times \mu_n : \mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n \to [0, +\infty], \tag{3.46}$$

$$A_1 \times \ldots \times A_n \mapsto \mu_1(A_1) \cdot \ldots \cdot \mu_n(A_n). \tag{3.47}$$

By Proposition 3.6, if all the $\mathcal{A}_i$ are semi-rings then $\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n$ is a semi-ring on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$. Moreover, we have the following:

Proposition 3.83. In the above setting, if $\mathcal{A}_i$ is a semi-ring and μ_i is a pre-measure on $\mathcal{A}_i$ for all i then $\mu_1 \times \ldots \times \mu_n$ is a pre-measure on the semi-ring $\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n$.

Proof. We need to prove that if

$$A_1 \times \ldots \times A_n = \cup_{k \in \mathbb{N}} (A_1^{(k)} \times \ldots \times A_n^{(k)}), \tag{3.48}$$

where $A_i, A_i^{(k)} \in \mathcal{A}_i$ for all i and k then

$$(\mu_1 \times \ldots \times \mu_n)(A_1 \times \ldots \times A_n) = \sum_{k \in \mathbb{N}} (\mu_1 \times \ldots \times \mu_n)(A_1^{(k)} \times \ldots \times A_n^{(k)}). \tag{3.49}$$

(3.48) is equivalent to

$$\mathbf{1}_{A_1 \times \ldots \times A_n} = \sum_{k \in \mathbb{N}} \mathbf{1}_{A_1^{(k)} \times \ldots \times A_n^{(k)}},$$

which we can rewrite as

$$\mathbf{1}_{A_1}(x_1) \cdot \ldots \cdot \mathbf{1}_{A_n}(x_n) = \sum_{k \in \mathbb{N}} \mathbf{1}_{A_1^{(k)}}(x_1) \cdot \ldots \cdot \mathbf{1}_{A_n^{(k)}})(x_n), \qquad x_i \in \mathcal{X}_i,\ i \in [n].$$

Let $\overline{\mu}_i := \mu_i^*\big|_{\mathcal{M}(\mu_i^*)}$ be the Carathéodory extension of μ_i. Integrating with respect to $\overline{\mu}_1$, and using the σ-additivity of the integral, we get that

$$\mu_1(A_1) \cdot \mathbf{1}_{A_2}(x_2) \cdot \ldots \cdot \mathbf{1}_{A_n}(x_n) = \sum_{k \in \mathbb{N}} \mu_1(A_1^{(k)}) \cdot \mathbf{1}_{A_2^{(k)}}(x_2) \cdot \ldots \cdot \mathbf{1}_{A_n^{(k)}})(x_n),$$

for all $x_i \in \mathcal{X}_i$, $i = 2, \ldots, n$, where we used that $\overline{\mu}_1(B) = \mu_1(B)$ for any $B \in \mathcal{A}_1$. Repeating the same for $i = 2, \ldots, n$, we get

$$\mu_1(A_1) \cdot \ldots \mu_n(A_n) = \sum_{k \in \mathbb{N}} \mu_1(A_1^{(k)}) \cdot \ldots \mu_n(A_n^{(k)}),$$

which is exactly (3.49). □

By the general theory of the Carathéodory extension, Proposition 3.83 immediately yields the following

Corollary 3.84. Let μ_i be a pre-measure on a semi-ring $\mathcal{A}_i$ for all $i = 1, \dots, n$. Then $(\mu_1 \times \ldots \times \mu_n)^*$ is a measure on the σ-algebra $\mathcal{M}\left((\mu_1 \times \ldots \times \mu_n)^*\right)$, which contains $\sigma\left(\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n\right)$.

Assume now that we have measurable spaces $(\mathcal{X}_i, \mathcal{A}_i)$ for all $i = 1, \dots, n$. Then $\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n$ is the collection of boxes in $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ whose sides are measurable. Note that $\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n$ is not a σ-algebra in general, but it is always a semi-ring. The *product* of the σ-algebras $\mathcal{A}_1, \dots, \mathcal{A}_n$ is defined as the σ-algebra generated by all the boxes with measurable sides, i.e.,

$$\otimes_{i\in[n]}\mathcal{A}_i := \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n := \sigma\left(\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n\right).$$

By definition, $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$ is a σ-algebra on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, and hence

$$\otimes_{i\in[n]}(\mathcal{X}_i, \mathcal{A}_i) := (\times_{i\in[n]}\mathcal{X}_i, \otimes_{i\in[n]}\mathcal{A}_i)$$

is a measurable space, that we call the product of the measurable spaces $(\mathcal{X}_i, \mathcal{A}_i)$, $i = 1, \dots, n$.

Assume now that there is also a measure μ_i given on all $\mathcal{A}_i$, i.e., we have measure spaces $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ for all $i = 1, \dots, n$, and consider their product $\mu_1 \times \ldots \times \mu_n$ as defined in (3.46):

$$\begin{aligned}\mu_1 \times \ldots \times \mu_n : \mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n &\to [0, +\infty], \\ A_1 \times \ldots \times A_n &\mapsto \mu_1(A_1) \cdot \ldots \cdot \mu_n(A_n).\end{aligned}$$

By Proposition 3.83 and Corollary 3.84, $(\mu_1 \times \ldots \times \mu_n)^*$ is a measure on the σ-algebra $\mathcal{M}\left((\mu_1 \times \ldots \times \mu_n)^*\right)$, which contains $\sigma\left(\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n\right) = \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$.

Definition 3.85. The measures

$$\begin{aligned}\mu_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \mu_n &:= (\mu_1 \times \ldots \times \mu_n)^*\big|_{\mathcal{M}((\mu_1 \times \ldots \times \mu_n)^*)}, \\ \mu_1 \otimes \ldots \otimes \mu_n &:= (\mu_1 \times \ldots \times \mu_n)^*\big|_{\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n} = \mu_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \mu_n\big|_{\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n}\end{aligned}$$

are called the *complete product* and the *product* of the measures $\mu_1, \dots, \mu_n$, respectively. Correspondingly,

$$\begin{aligned}\overline{\otimes}_{i\in[n]}(\mathcal{X}_i, \mathcal{A}_i, \mu_i) &:= \left(\times_{i\in[n]}\mathcal{X}_i, \mathcal{M}\left((\mu_1 \times \ldots \times \mu_n)^*\right), \overline{\otimes}_{i\in[n]}\,\mu_i\right), \\ \otimes_{i\in[n]}(\mathcal{X}_i, \mathcal{A}_i, \mu_i) &:= \left(\times_{i\in[n]}\mathcal{X}_i, \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n, \otimes_{i\in[n]}\mu_i\right)\end{aligned}$$

are called the *complete product* and the *product* of the measure spaces $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$, $i \in [n]$, respectively.

Remark 3.86. By the general properties of the Carathéodory extension, $\overline{\otimes}_{i\in[n]}(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ is always a complete measure space, even if none of the $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ are complete. On the other hand, $\otimes_{i\in[n]}(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ need not be complete in general, even if all $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ are complete. We will revisit this problem shortly in Proposition 3.91.

Our next goal is to show that the n-dimensional Lebesgue measure is the n-fold complete product of the 1-dimensional Lebesgue measure with itself. This will be an easy consequence of the following:

Lemma 3.87. For every $i \in [n]$, let μ_i be a pre-measure on a semi-ring $\mathcal{S}_i$, let $\mathcal{A}_i$ be a σ-algebra such that $\mathcal{S}_i \subseteq \mathcal{A}_i \subseteq \mathcal{M}(\mu_i^*)$, and let $\overline{\mu}_i := \mu_i^*\big|_{\mathcal{M}(\mu_i^*)}$, $\tilde{\mu}_i := \mu_i^*\big|_{\mathcal{A}_i}$. Then

$$(\overline{\mu}_1 \times \ldots \times \overline{\mu}_n)^* = (\tilde{\mu}_1 \times \ldots \times \tilde{\mu}_n)^* = (\mu_1 \times \ldots \times \mu_n)^* .$$

Proof. The inequalities

$$(\overline{\mu}_1 \times \ldots \times \overline{\mu}_n)^* \leq (\tilde{\mu}_1 \times \ldots \times \tilde{\mu}_n)^* \leq (\mu_1 \times \ldots \times \mu_n)^* .$$

are obvious. The converse inequalities follow by an elementary but slightly tedious argument, which we leave as an exercise. □

Corollary 3.88. In the above setting, we have

$$\begin{aligned}\overline{\mu_1 \times \ldots \times \mu_n} &:= (\mu_1 \times \ldots \times \mu_n)^*\big|_{\mathcal{M}((\mu_1\times\ldots\times\mu_n)^*)} \\ &= (\tilde{\mu}_1 \times \ldots \times \tilde{\mu}_n)^*\big|_{\mathcal{M}((\tilde{\mu}_1\times\ldots\times\tilde{\mu}_n)^*)} = \tilde{\mu}_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \tilde{\mu}_n \\ &= (\overline{\mu}_1 \times \ldots \times \overline{\mu}_n)^*\big|_{\mathcal{M}((\overline{\mu}_1\times\ldots\times\overline{\mu}_n)^*)} = \overline{\mu}_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \overline{\mu}_n,\end{aligned}$$

where the second equalities in each line are by definition, and the first equalities in the second and the third lines are due to Lemma 3.87.

Example 3.89. Taking $\mu_i := \lambda_1$ to be the length function on $\mathrm{Box}(\mathbb{R})$, $\mu_1 \times \ldots \times \mu_n$ is the volume function λ_n on the n-dimensional boxes $\mathrm{Box}(\mathbb{R}^n) = \mathrm{Box}(\mathbb{R})\,(\times)\ldots(\times)\,\mathcal{T}(\mathbb{R})$, and $\overline{\mu_1 \times \ldots \times \mu_n}$ is, by definition, the n-dimensional Lebesgue measure $\overline{\lambda}_n$ on the n-dimensional Lebesgue measurable sets $\Lambda(\mathbb{R}^n)$. By the above, it is equal to the n-fold complete product of the 1-dimensional Lebesgue measure with itself. That is,

$$\overline{\otimes}_{i\in[n]} \left(\mathbb{R}, \Lambda(\mathbb{R}), \overline{\lambda}_1\right) = \left(\mathbb{R}^n, \Lambda(\mathbb{R}^n), \overline{\lambda}_n\right).$$

Taking $\mathcal{A}_i = \mathcal{B}(\mathbb{R})$, Corollary 3.88 tells that the n-fold complete product of the 1-dimensional Lebesgue on the Borel σ-algebra by itself is still the n-dimensional Lebesgue measure on the Lebesgue-measurable sets, i.e.,

$$\overline{\otimes}_{i\in[n]} \left(\mathbb{R}, \mathcal{B}(\mathbb{R}), \overline{\lambda}_1\big|_{\mathcal{B}(\mathbb{R})}\right) = \left(\mathbb{R}^n, \Lambda(\mathbb{R}^n), \overline{\lambda}_n\right).$$

The above example of the Lebesgue measure can be generalized to Lebesgue-Stieltjes measures. For this, it will be convenient to introduce the following notation. For functions $f_i : \mathcal{X}_i \to \mathbb{K}$, let

$$\begin{aligned} f_1 \otimes \ldots \otimes f_n : \mathcal{X}_1 \times \ldots \times \mathcal{X}_n &\to \mathbb{K} \\ (x_1, \ldots, x_n) &\mapsto f_1(x_1) \cdot \ldots \cdot f_n(x_n). \end{aligned}$$

Note that, for instance,

$$\mathbf{1}_{A_1 \times \ldots \times A_n} = \mathbf{1}_{A_1} \otimes \ldots \otimes \mathbf{1}_{A_n}.$$

Example 3.90. Let $F_1, \ldots, F_n : \mathbb{R} \to \mathbb{R}$ be monotone increasing functions that are continuous from the left at each point. Then $F_1 \otimes \ldots F_n$ is also continuous from the left at each point. Moreover, we have

$$\begin{aligned} \Delta^{(1)}_{[a_1,b_1)} \ldots \Delta^{(1)}_{[a_1,b_1)} (F_1 \otimes \ldots F_n) &= (F(b_1) - F(a_1)) \cdot \ldots \cdot (F(b_n) - F(a_n)) \\ &= \lambda_{F_1}([a_1, b_1)) \cdot \ldots \cdot \lambda_{F_n}([a_n, b_n)) \\ &= (\lambda_{F_1} \times \ldots \times \lambda_{F_n})([a_1, b_1) \times \ldots \times [a_n, b_n)) \end{aligned} \tag{3.50}$$

for all $\underline{a}, \underline{b} \in \mathbb{R}^n$; in particular, $\lambda_{F_1 \otimes \ldots F_n} = \lambda_{F_1} \times \ldots \times \lambda_{F_n}$ on $\mathrm{Box}(\mathbb{R}^n) = \mathrm{Box}(\mathbb{R})\,(\times)\ldots(\times)\,\mathcal{T}(\mathbb{R})$. Since the RHS in (3.50) is non-negative for all $\underline{a}, \underline{b} \in \mathbb{R}^n$, we can define

$$\begin{aligned} \overline{\lambda}_{F_1 \otimes \ldots \otimes F_n} &:= \lambda^*_{F_1 \otimes \ldots \otimes F_n}\big|_{\mathcal{M}\left(\lambda^*_{F_1 \otimes \ldots \otimes F_n}\right)} = (\lambda_{F_1} \times \ldots \times \lambda_{F_n})^* \,\big|_{\mathcal{M}\left((\lambda_{F_1} \times \ldots \times \lambda_{F_n})^*\right)} \\ &= \overline{\lambda}_{F_1} \,\overline{\otimes} \ldots \overline{\otimes}\, \overline{\lambda}_{F_n}, \end{aligned}$$

where the first equality is due to (3.50), and the second equality is due to Lemma 3.87. Again, we can rewrite this as

$$\overline{\otimes}_{i \in [n]} \left(\mathbb{R}, \mathcal{M}(\lambda^*_{F_i}), \overline{\lambda}_{F_i}\right) = \left(\mathbb{R}^n, \mathcal{M}\left(\lambda^*_{F_1 \otimes \ldots \otimes F_n}\right), \overline{\lambda}_{F_1 \otimes \ldots \otimes F_n}\right),$$

and we get the same if we restrict all $\overline{\lambda}_{F_i}$ to the Borel σ-algebra:

$$\overline{\otimes}_{i \in [n]} \left(\mathbb{R}, \mathcal{B}(\mathbb{R}), \overline{\lambda}_{F_i}\big|_{\mathcal{B}(\mathbb{R})}\right) = \left(\mathbb{R}^n, \mathcal{M}\left(\lambda^*_{F_1 \otimes \ldots \otimes F_n}\right), \overline{\lambda}_{F_1 \otimes \ldots \otimes F_n}\right).$$

Taking $F_i = \mathrm{id}_{\mathbb{R}}$ for all i yields Example 3.89 as a special case.

Next, we consider the uniqueness and the completeness of the product measure.

Proposition 3.91. Assume that μ_i is a σ-finite measure on a σ-algebra $\mathcal{A}_i \subseteq \mathcal{P}(\mathcal{X}_i)$ for all $i = 1, \ldots, n$. Then

(i) $\mu_1 \otimes \ldots \otimes \mu_n$ is also σ-finite, and it is the unique extension of $\mu_1 \times \ldots \times \mu_n$ onto $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$.

(ii) $\overline{\mu_1 \otimes \ldots \otimes \mu_n} = \overline{\mu}_1 \,\overline{\otimes} \ldots \overline{\otimes}\, \overline{\mu}_n$, where $\overline{\mu_1 \otimes \ldots \otimes \mu_n}$ is the natural extension of $\mu_1 \otimes \ldots \otimes \mu_n$, which also coincides with its completion.

Proof. (i) By assumption, there exist decompositions $\mathcal{X}_i = \cup_{k\in\mathbb{N}} A_k^{(i)}$ with $A_k^{(i)} \in \mathcal{A}_i$ and $\mu_i\left(A_k^{(i)}\right) < +\infty$ for all i, k. Then

$$\mathcal{X}_1 \times \ldots \times \mathcal{X}_n = \cup_{k_1\in\mathbb{N}} \ldots \cup_{k_n\in\mathbb{N}} A_{k_1}^{(1)} \times \ldots \times A_{k_n}^{(n)},$$

and $(\mu_1 \times \ldots \times \mu_n)\left(A_{k_1}^{(1)} \times \ldots \times A_{k_n}^{(n)}\right) = \mu_1(A_{k_1}^{(1)}) \cdot \ldots \cdot \mu_n(A_{k_n}^{(n)}) < +\infty$. Thus, $\mu_1 \times \ldots \times \mu_n$ is (completely) σ-finite on the semi-ring $\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n$. The assertion then follows from the general properties of the Carathéodory extension.

(ii) Since $\mu_1 \times \ldots \times \mu_n$ is (completely) σ-finite on the semi-ring $\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n$, we have, for all σ-algebra $\mathcal{A}$ such that $\sigma(\mathcal{A}_1 (\times) \ldots (\times) \mathcal{A}_n) \subseteq \mathcal{A} \subseteq \mathcal{M}((\mu_1 \times \ldots \times \mu_n)^*)$ that

$$\left((\mu_1 \times \ldots \times \mu_n)^*\big|_{\mathcal{A}}\right)^* = (\mu_1 \times \ldots \times \mu_n)^* = (\overline{\mu}_1 \times \ldots \times \overline{\mu}_n)^*,$$

where the last equality is due to Lemma 3.87. Taking now $\mathcal{A} = \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$, we get

$$(\mu_1 \otimes \ldots \otimes \mu_n)^* = (\overline{\mu}_1 \times \ldots \times \overline{\mu}_n)^*.$$

Taking the restriction of both sides onto $\mathcal{M}((\overline{\mu}_1 \times \ldots \times \overline{\mu}_n)^*)$ yields $\overline{\mu_1 \otimes \ldots \otimes \mu_n} = \overline{\mu}_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \overline{\mu}_n$. Since $\mu_1 \otimes \ldots \otimes \mu_n$ is σ-finite, its natural extension coincides with its completion. □

Example 3.92. Let $F_i : \mathbb{R} \to \mathbb{R}$ be a monotone increasing and left continuous function, and $\mu_i := \overline{\lambda}_{F_1}\big|_{\mathcal{B}(\mathbb{R})}$ for all $i = 1, \ldots, n$. Note that $\overline{\mu}_i = \overline{\lambda}_{F_i}$, and hence

$$\overline{\mu_1 \otimes \ldots \otimes \mu_n} = \overline{\overline{\lambda}_{F_1} \otimes \ldots \otimes \overline{\lambda}_{F_n}} = \overline{\lambda}_{F_1} \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \overline{\lambda}_{F_n} = \overline{\lambda}_{F_1 \otimes \ldots \otimes F_n}.$$

3.8 $\mathcal{L}^p$ spaces

Proposition 3.93. Let $(\mathcal{X}, \tau)$ be a locally compact topological space and μ be a regular Borel measure on a σ-algebra $\mathcal{A} \supseteq \mathcal{B}(\tau)$. Then $C_c(\mathcal{X}, V)$ is dense in $\mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mu; V)$ for any separable Banach space V and any $p \in [1, +\infty)$.

Proof. For every $f \in \mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mu; V)$ there exists a sequence s_n of simple measurable functions such that $\lim_{n\to+\infty} \|f - s_n\|_p = 0$, and all level sets of s_n corresponding to non-zero values have finite measure. Hence, it is enough to show that the characteristic function of any measurable set $A \in \mathcal{A}$ with finite measure can be approximated by arbitrary precision by continuous functions of compact support.

Due to regularity, for any $\varepsilon > 0$ there exist an open set $G \supseteq A$ and a compact set $K \subseteq G$ such that $\mu(G \setminus A) < \varepsilon$, $\mu(G \setminus K) < \varepsilon$. By Proposition 1.19, there exists

a continuous function $f : \mathcal{X} \to [0,1]$ of compact support such that $f|_K \equiv 1$ and $f|_{\mathcal{X}\setminus G} \equiv 0$. Hence,

$$\|\mathbf{1}_A - f\|_p \leq \|\mathbf{1}_A - \mathbf{1}_G\|_p + \|\mathbf{1}_G - f\|_p = \mu(G \setminus A)^{1/p} + \mu(G \setminus K)^{1/p} < 2\varepsilon^{1/p}.$$

□

4 Functional Analysis

4.1 Vector spaces

We assume that the reader is familiar with the basics of linear algebra; however, we summarize some of the most important notions below.

Recall that a *vector space* over the scalar field $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ is a set V with a binary operation $+ : V \times V \to V$ (called addition), that is

- *associative:* $(x + y) + z = x + (y + z),\ \ x, y, z \in V$;
- *commutative:* $x + y = y + x,\ \ x, y, \in V$;
- there exists an element $0 \in V$ such that $0 + x = x + 0 = x,\ \ x \in V$;
- any element $x \in V$ has an *inverse*, denoted by $-x$ such that $x - x := x + (-x) = 0$.

That is, V is an Abelian group with the addition, and the null element 0. Moreover, there exists an operation

$$\mathbb{K} \times V \to V, \qquad (\lambda, x) \mapsto \lambda x,$$

with the properties

- $\lambda(\eta x) = (\lambda\eta)x,\ \ \lambda, \eta \in \mathbb{K},\ x \in V$;
- $(\lambda + \eta)x = \lambda x + \eta x,\ \ \lambda, \eta \in \mathbb{K},\ x \in V$;
- $\lambda(x + y) = \lambda x + \lambda y,\ \ \lambda \in \mathbb{K},\ x, y \in V$.

Example 4.1. (General function space)
Let $\mathcal{X}$ be an arbitrary non-empty set, and define

$$\mathbb{K}^{\mathcal{X}} := \{f : \mathcal{X} \to \mathbb{K} \ \text{ function}\},$$

which is the set of all $\mathbb{K}$-valued functions on $\mathcal{X}$. This is a vector space with the natural *point-wise operations*

$$(f + g)(x) := f(x) + g(x), \qquad (\lambda f)(x) := \lambda \cdot f(x), \qquad x \in \mathcal{X},\ \lambda \in \mathbb{K},\ f, g \in \mathbb{K}^{\mathcal{X}}.$$

We will often consider two special cases; when $\mathcal{X} = [d] := \{1, \dots, d\}$, then

$$\mathbb{K}^d := \mathbb{K}^{[d]} = \{x : [d] \to \mathbb{K}\} = \{(x_1, \dots, x_d) : x_i := x(i) \in \mathbb{K},\ i \in [d]\}$$

is the usual vector space of d-tuples of real or complex numbers.

An "infinite version" of the above is the sequence space

$$\mathbb{K}^{\mathbb{N}} := \{(x_1, x_2, \dots) : x_i := x(i) \in \mathbb{K},\ i \in \mathbb{N}\}.$$

Definition 4.2. Let V be a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. For a finite set of vectors $v_1, \ldots, v_r \in V$, and scalars $\lambda_1, \ldots, \lambda_r$, the expression $\sum_{i=1}^{r} \lambda_i v_i$ is a *linear combination* of the vectors with the given coefficients.

Definition 4.3. A subset $A \subseteq V$ of a vector space V is a *(linear) subspace* if it is closed under linear combinations. We say that a subspace A is a *proper subspace* if $A \neq \{0\}$ and $A \neq V$.

Remark 4.4. By a "subspace" we will always mean a linear subspace, often omitting "linear".

Remark 4.5. It is easy to see that for A being a subspace of V, it is sufficient that for any $x, y \in V$, and any $\lambda \in \mathbb{K}$, $\lambda x \in A$, and $x + y \in A$.

Example 4.6. The set of all continuous functions on an interval $[a, b] \subseteq \mathbb{R}$, denoted by

$$C_{\mathbb{K}}([a, b]) := \{f \in \mathbb{K}^{[a,b]} : f \text{ continuous}\}$$

is easily seen to be a subspace of $\mathbb{K}^{[a,b]}$. To see this, we only have to verify that if f, g are continuous on $[a, b]$ then $f + g$ is also continuous, and for any $\lambda \in \mathbb{K}$, λf is continuous.

It is easy to see that an arbitrary collection of linear subspaces of a vector space V is again a linear subspace of V. In particular, for any set $A \subseteq V$,

$$\text{span}(A) := \cap\{W \text{ linear subspace in } V,\, W \supseteq A\}$$

is a linear subspace of V, which we call the *subspace generated by* A, the subspace *spanned* by A, or simply the *span of* A.

Exercise 4.7. Show that for any subset $A \subseteq V$, $\text{span}(A)$ is the collection of all linear combinations of elements of A, i.e.,

$$\text{span}(A) = \left\{ \sum_{i=1}^{n} \lambda_i v_i : \lambda_i \in \mathbb{K},\, v_i \in A,\, i \in [n],\, n \in \mathbb{N} \right\},$$

and that $\text{span}(A)$ is the *smallest subspace containing* A.

Definition 4.8. Let A be a subset in a vector space V.

- We say that A is a *generating system* for a subspace $W \subseteq V$ if $\text{span}(A) = W$. If $\text{span}(A) = V$ then we say that A is a generating system for V, or simply that it is a generating system.

- We say that A is *linearly independent*, if for any $v_1, \ldots, v_n \in A$, $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$,

$$0 = \sum_{i=1}^{n} \lambda_i v_i \quad \Longrightarrow \quad \lambda_1 = \ldots = \lambda_n = 0.$$

- We say that A is a *basis* in a subspace W in V, if it is linearly independent and a generating system for W. We simply say that A is a basis if it is a basis for V.

Remark 4.9. We will sometimes use "algebraic basis" for the above concept of a basis, to distinguish it from, for instance, orthonormal bases in Hilbert spaces; see Section **??**.

Theorem 4.10. Any vector space has a basis, and the cardinality of any basis in a given vector space is the same; this number is called the (algebraic) dimension of the vector space.

The proof of the above theorem is beyond the scope of these notes. We only mention that existence can be shown easily using the axiom of choice, and in fact, the statement that every vector space has a basis is *equivalent* to the axiom of choice.

Example 4.11. For an arbitrary set $\mathcal{X}$, and $x \in \mathcal{X}$, let

$$\delta_x := \mathbf{1}_{\{x\}} : y \mapsto \begin{cases} 1, & y = x, \\ 0, & y \neq x \end{cases}$$

be the *Dirac delta* concentrated at the point x. It is easy to see that

$$\{\delta_x : x \in \mathcal{X}\} \quad \text{is linearly independent.}$$

When $\mathcal{X}$ is finite, it is also easy to see that it is a generating system, and hence a basis. In fact, for $\mathcal{X} = [d]$, $\delta_i =: e_i$ is just the familiar *canonical basis vector*, whose components are all zero, except for the i-th component, which is 1.

On the contrary, if $\mathcal{X}$ is infinite, then the Dirac deltas do not form a generating system (and hence neither a basis); instead, we have

$$\operatorname{span}(\{\delta_x : x \in \mathcal{X}\}) = \left(\mathbb{K}^{\mathcal{X}}\right)_f := \{f : \mathcal{X} \to \mathbb{K} : |\{x \in \mathcal{X} : f(x) \neq 0\}| < +\infty\},$$

i.e., the subspace generated by the Dirac deltas is the proper subspace of functions that are non-zero only at finitely many points.

Exercise 4.12. Show that the following subsets are linearly independent, but do not form a basis in the indicated vector spaces:

(i) $\mathcal{P}_{\mathbb{K}}([a,b])$ polynomials on $[a,b]$ with $\mathbb{K}$-valued coefficients in $V := C_{\mathbb{K}}([a,b])$.

(ii) $\mathcal{P}_{\mathbb{K}}([a,b])$ polynomials on $[a,b]$ with $\mathbb{K}$-valued coefficients in $V := C_{\mathbb{K}}([a,b])$.

For vector spaces V, W over the same field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, let $\mathrm{Lin}(V;W)$ denote the set of linear operators from a vector space V to a vector space W. When $V = W$, we will also use the short-hand notation $\mathrm{Lin}(V)$. Elements of $\mathrm{Lin}(V;\mathbb{K})$ are called *linear functionals* on V, and $\mathrm{Lin}(V;\mathbb{K})$ is called the *dual space* of V, for which we also use the notation

$$V' = \mathrm{Lin}(V;\mathbb{K}).$$

More generally, we use the notation $\mathrm{Lin}_n(V_1,\dots,V_n;W)$ for the set of n-linear maps from $V_1 \times \dots \times V_n$ to W; i.e., if $A \in \mathrm{Lin}_n(V_1,\dots,V_n;W)$ then

$$\begin{aligned} A(v_1,\dots,v_i+u_i,\dots,v_n) &= A(v_1,\dots,v_i,\dots,v_n) + A(v_1,\dots,u_i,\dots,v_n),\\ A(v_1,\dots,\lambda v_i,\dots,v_n) &= \lambda A(v_1,\dots,v_i,\dots,v_n), \end{aligned}$$

where $v_1,\dots,v_n \in V$, $u_i \in V$, and $\lambda \in \mathbb{C}$. The above lines should be interpreted such that we only add a vector u_i in the i-th position, and have v_j in all the other positions in the first line, and we only multiply the argument by λ in the i-th position in the second line.

It is easy to see that $\mathrm{Lin}_n(V_1,\dots,V_n;W)$ forms a vector space with the usual pointwise operations, i.e., if $A, A_1, A_2 \in \mathrm{Lin}_n(V_1,\dots,V_n;W)$ and $\lambda \in \mathbb{C}$ then

$$\begin{aligned} (A_1+A_2)(v_1,\dots,v_n) &:= A_1(v_1,\dots,v_n) + A_2(v_1,\dots,v_n),\\ (\lambda A)(v_1,\dots,v_n) &:= \lambda \cdot A(v_1,\dots,v_n), \qquad\qquad v_i \in V_i. \end{aligned}$$

Lemma 4.13. Let $\{e_{i,j}\}_{j\in J_i}$ be bases in V_i for every $i = 1,\dots,n$, and $w_{j_1,\dots,j_n} \in W$, $j_i \in J_i$, $i \in [n]$ for some vector space W. Then there exists a unique $\Phi \in \mathrm{Lin}_n(V_1,\dots,V_n;W)$ such that $\Phi(e_{1,j_i},\dots,e_{n,j_n}) = w_{j_1,\dots,j_n}$ for all $j_i, \in J_j$, $i \in [n]$.

Proof. Trivial, exercise. □

Definition 4.14. For a linear operator $A \in \mathrm{Lin}(V,W)$, let

$$\ker(A) := \{v \in V : Av = 0\} \subseteq V$$

be the *kernel* of A, and

$$\mathrm{ran}(A) := \{Av : v \in V\} \subseteq W$$

be the *range* of A.

Remark 4.15. We may also use the notations $\ker A$ and $\operatorname{ran} A$, without brackets around the operator.

Remark 4.16. The range of an operator is also often called its *image*, and denoted by $\operatorname{Im}(A)$.

Exercise 4.17. Show that for $A \in \operatorname{Lin}(V, W)$, $\ker(A)$ and $\operatorname{ran}(A)$ are subspaces in V and W, respectively.

4.2 Normed spaces

Definition 4.18. Let V be a vector space over $\mathbb{K}$. A function $\|\ \| : V \to \mathbb{R}_+$ is a *norm* if it has the following properties:

- $\|x\| \geq 0, \quad x \in V, \quad \text{and} \quad \|x\| = 0 \iff x = 0 \qquad$ (strict positivity);
- $\|\lambda x\| = |\lambda| \, \|x\|, \quad x \in \mathcal{H}, \quad \lambda \in \mathbb{K}; \qquad$ (positive homogeneity);
- $\|x + y\| \leq \|x\| + \|y\|, \quad x, y \in \mathcal{H}, \qquad$ (triangle inequality).

Exercise 4.19. Show that the following define norms on $\mathbb{C}^d$:

$$\|x\|_\infty := \max_{1 \leq i \leq d} |x_i|;$$
$$\|x\|_1 := \sum_{i=1}^{n} |x_i|.$$

Definition 4.20. Let $V_1, \ldots, V_n, W$ be normed spaces. The *operator norm* (or *induced norm*) of an n-linear operator $A \in \operatorname{Lin}_n(V_1, \ldots, V_n; W)$ is defined as

$$\begin{aligned}
\|A\| &:= \inf\{c > 0 : \ \|A(v_1, \ldots, v_n)\| \leq c \, \|v_1\| \cdot \ldots \cdot \|v_n\| : \ v_i \in V_i\} \\
&= \sup\{\|A(v_1, \ldots, v_n)\| : \ \|v_i\| \leq 1, \ v_i \in V_i \setminus \{0\}\} \\
&= \sup\{\|A(v_1, \ldots, v_n)\| : \ \|v_i\| = 1, \ v_i \in V_i \setminus \{0\}\} \\
&= \sup\left\{ \frac{\|A(v_1, \ldots, v_n)\|}{\|v_1\| \cdot \ldots \cdot \|v_n\|} : \ v_i \in V_i \setminus \{0\} \right\}.
\end{aligned}$$

We say that A is *bounded* if $\|A\| < +\infty$, and denote the set of bounded n-linear operators from $V_1 \times \ldots \times V_n$ to W by $\mathcal{B}_n(V_1, \ldots, V_n; W)$.

In particular, for $A \in \operatorname{Lin}(V, W)$ we have

$$\|A\| = \sup\{\|A(v)\| : \ \|v\| \leq 1, \, v \in V\} = \sup\left\{ \frac{\|A(v)\|}{\|v\|} : \ v \in V \setminus \{0\} \right\}.$$

We denote the set of bounded linear operators from V to W by $\mathcal{B}(V, W)$, and if $V = W$, we use the short-hand notation $\mathcal{B}(V)$.

Exercise 4.21. (i) Show that Definition 4.20 indeed defines a norm on $\mathrm{Lin}_n(V_1, \ldots, V_n; W)$.

(ii) Show that $\mathcal{B}_n(V_1, \ldots, V_n; W)$ is a linear subspace of $\mathrm{Lin}_n(V_1, \ldots, V_n; W)$.

Exercise 4.22. Show that the operator norm is *submultiplicative* in the sense that if $(V_k, \|.\|_k)$ are normed spaces for $k = 1, 2, 3$, and $A_1 \in \mathrm{Lin}(V_1, V_2)$, $A_2 \in \mathrm{Lin}(V_2, V_3)$, then

$$\|A_2 A_1\| \leq \|A_2\| \, \|A_1\| .$$

4.3 Dense subspaces

Lemma 4.23. The function h defined by

$$h(x) := \frac{g(x)}{g(x) + g(1-x)}, \qquad g(x) := \begin{cases} e^{\frac{1}{x}}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

is infinitely many times differentiable on $\mathbb{R}$, it is 0 when $x \leq 0$, and it is 1 when $x \geq 1$.

Proof. □

Theorem 4.24. For every $d \in \mathbb{N}$ and every $1 \leq p < +\infty$, $C_c^\infty(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d)$.

Proof. □

Next, we show that functions on a product space can often be well approximated by functions that are the product of functions on the individual spaces. This will be very useful, e.g., in constructing dense sets in $L^p(\mathbb{R}^d)$ with good properties from dense sets in $L^p(\mathbb{R})$.

Recall that the tensor product of functions $f_i \in \mathbb{K}^{\mathcal{X}_i}$, $i \in [n]$, is the n-variable function

$$(f_1 \otimes \ldots \otimes f_n)(x_1, \ldots, x_n) := f_1(x_1) \cdot \ldots \cdot f_n(x_n), \qquad x_i \in \mathcal{X}_i, \, i \in [n].$$

For subspaces $V_i \subseteq \mathbb{K}^{\mathcal{X}_i}$, we define their *tensor product* as

$$V_1 \otimes \ldots \otimes V_n := \mathrm{span}\{f_1 \otimes \ldots \otimes f_n : \, f_i \in V_i, \, i \in [n]\},$$

which is a subspace of $\mathbb{K}^{\mathcal{X}_1 \times \ldots \times \mathcal{X}_n}$. Note that

$$\times_{i=1}^n V_i \ni (f_1, \ldots, f_n) \mapsto f_1 \otimes \ldots \otimes f_n \in V_1 \otimes \ldots \otimes V_n \quad \text{ is } n\text{-linear}.$$

Example 4.25. Note that the indicator function of a disjoint union is the sum of the indicator functions:

$$\mathbf{1}_{\cup_{i=1}^n A_i} = \sum_{i=1}^n \mathbf{1}_{A_i}.$$

Analogously, the indicator function of a Cartesian product of sets is the product of the indicator functions: if $A_i \subseteq \mathcal{X}_i$, $i \in [n]$, then

$$\mathbf{1}_{A_1 \times \ldots \times A_n} = \otimes_{i=1}^n \mathbf{1}_{A_i}.$$

Recall that the d-dimensional Lebesgue measure is the d-fold product of the 1-dimensional Lebesgue measure with itself, and more generally,

$$\lambda_{d_1+\ldots+d_n} = \otimes_{i=1}^n \lambda_{d_i};$$

see Sections 2.3 and 3.7. We are mainly interested in the following statements in this setting, but we state them in higher generality as it requires no extra effort.

Lemma 4.26. Let $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$, $i \in [n]$, be measure spaces and μ be a measure on $\otimes_{i=1}^n \mathcal{A}_i$ that factorizes to the product of the μ_i (see Definition **??**). Let $f_i, \tilde{f}_i \in \mathbb{K}^{\mathcal{X}_i}$, $i \in [n]$, be measurable. Then

$$\forall i: \quad \mu_i\left(\{f_i \neq \tilde{f}_i\}\right) = 0 \quad \Longrightarrow \quad \mu\left(\{f_1 \otimes \ldots \otimes f_n \neq \tilde{f}_1 \otimes \ldots \otimes \tilde{f}_n\}\right) = 0. \quad (4.1)$$

Moreover, for any $p \in [1, +\infty]$,

$$\|f_1 \otimes \ldots \otimes f_n\|_p = \prod_{i=1}^n \|f_i\|_p. \quad (4.2)$$

Proof. (4.1) is immediate from

$$\{f_1 \otimes \ldots \otimes f_n \neq \tilde{f}_1 \otimes \ldots \otimes \tilde{f}_n\} \subseteq \cup_{k=1}^n \left(\{f_k \neq \tilde{f}_k\} \times \left(\times_{j \in [n]\setminus\{k\}} \mathcal{X}_j\right)\right),$$

and (4.2) follows immediately from Exercise **??**. □

Let $[f]$ denote the equivalence class of a function on a measure space $(\mathcal{X}, \mathcal{A}, \mu)$ w.r.t the relation $f \sim \tilde{f}$ if $\mu(\{f \neq \tilde{f}\}) = 0$. Lemma 4.26 yields immediately the following:

Corollary 4.27. In the setting of Lemma 4.26, for any $p \in [1, +\infty)$,

$$\times_{i=1}^n L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i) \ni ([f_1], \ldots, [f_n]) \mapsto [f_1 \otimes \ldots \otimes f_n]$$

is a well-defined map to $L^p(\times_{i=1}^n \mathcal{X}_i, \otimes_{i=1}^n \mathcal{A}_i, \mu)$ that is n-linear, and

$$\|[f_1 \otimes \ldots \otimes f_n]\|_p = \prod_{i=1}^n \|[f_i]\|_p.$$

As usual, we will omit the equivalence class sign, and simply write $f \in L^p(\mathcal{X}, \mathcal{A}, \mu)$ if $[f] \in L^p(\mathcal{X}, \mathcal{A}, \mu)$. With this convention, we define

$$\otimes_{i=1}^n L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i) := \operatorname{span}\{f_1 \otimes \ldots \otimes f_n : f_i \in L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i)\}.$$

Proposition 4.28. In the setting of Lemma 4.26,

$$\overline{\otimes}_{i=1}^n L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i) := \overline{\otimes_{i=1}^n L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i)} \subseteq L^p(\times_{i=1}^n \mathcal{X}_i, \otimes_{i=1}^n \mathcal{A}_i, \mu).$$

If, moreover, all the μ_i are σ-finite, and hence they have a well-defined product measure $\otimes_{i=1}^d \mu_i$ on $\otimes_{i=1}^d \mathcal{A}_i$, then $\otimes_{i=1}^d L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ is dense in $L^p(\times_{i=1}^n \mathcal{X}_i, \otimes_{i=1}^n \mathcal{A}_i, \otimes_{i=1}^n \mu_i)$, i.e.,

$$\overline{\otimes}_{i=1}^n L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i) = L^p(\times_{i=1}^n \mathcal{X}_i, \otimes_{i=1}^n \mathcal{A}_i, \otimes_{i=1}^n \mu_i).$$

In particular,

$$\overline{\otimes}_{i=1}^n L^p(\mathbb{R}^{d_i}, \mathcal{B}(\mathbb{R}^{d_i}), \lambda_{d_i}) = L^p(\mathbb{R}^{d_1+\ldots+d_n}, \mathcal{B}(\mathbb{R}^{d_1+\ldots+d_n}), \lambda_{d_1+\ldots+d_n}).$$

Proof. We only need to show the σ-finite case, as the rest is obvious from the previous considerations. Let $A \in \otimes_{i=1}^n \mathcal{A}_i$ with $\otimes_{i=1}^n \mu_i(A) < +\infty$. By the definition of the product measure and Exercise **??**, for every $\varepsilon > 0$ there exist $B_{1,k,\varepsilon} \times \ldots \times B_{n,k,\varepsilon} \in \mathcal{A}_1 \times \ldots \times \mathcal{A}_n$, $k \in [n_\varepsilon]$, such that

$$\begin{aligned} \varepsilon &> (\otimes_{i=1}^n \mu_i)\left(A \triangle (B_{1,k,\varepsilon} \times \ldots \times B_{n,k,\varepsilon})\right) \\ &= \int \left|\mathbf{1}_A - \mathbf{1}_{B_{1,k,\varepsilon} \times \ldots \times B_{n,k,\varepsilon}}\right|^p \, d(\otimes_{i=1}^n \mu_i) = \int \left|\mathbf{1}_A - \otimes_{i=1}^n \mathbf{1}_{B_{i,k,\varepsilon}}\right|^p \, d(\otimes_{i=1}^n \mu_i), \end{aligned}$$

and hence $\mathbf{1}_A \in \overline{\otimes}_{i=1}^n L^p(\mathbb{R}^{d_i}, \mathcal{B}(\mathbb{R}^{d_i}), \lambda_{d_i})$. Since the subspace spanned by the indicator functions of measurable sets with finite measure is dense in L^p, the statement follows. □

Corollary 4.29. Let $(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$, $i \in [n]$, be σ-finite measure spaces. If for every $i \in [n]$, $\operatorname{span}\{f_{i,j} : j \in \mathcal{J}_i\}$ is dense in $L^p(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ then

$$\operatorname{span}\{f_{\underline{j}} := \otimes_{i=1}^n f_{i,j_i} : \underline{j} \in \times_{i=1}^n \mathcal{J}_i\} \quad \text{is dense in} \quad L^p\left(\times_{i=1}^n \mathcal{X}_i, \otimes_{i=1}^n \mathcal{A}_i, \otimes_{i=1}^n \mu_i\right).$$

4.4 Linear and multilinear operators

Definition 4.30. Let X, Y be vector spaces. A map $A : \operatorname{dom}(A)(\subseteq X) \to Y$ is called a *linear map* or *linear operator* if $\operatorname{dom}(A)$ is a linear subspace in X, and A is preserves linear combinations, i.e., for any $x_1, \ldots, x_r \in \operatorname{dom}(A)$, $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$, $r \in \mathbb{N}$,

$$A\left(\sum_{i=1}^r \lambda_i x_i\right) = \sum_{i=1}^r \lambda_i A(x_i).$$

When $Y = \mathbb{K}$, a linear map $\varphi : \operatorname{dom}(\varphi)(\subseteq X) \to \mathbb{K}$ is called a *linear functional*.

Remark 4.31. For a linear operator we often do not put a bracket around its argument, i.e., we write Ax instead of $A(x)$.

Definition 4.32. For a linear operator $A : \mathrm{dom}(A)(\subseteq X) \to Y$ between vector spaces X, Y, let

$$\ker(A) := \{x \in X :, Ax = 0\} \subseteq X$$

be the *kernel* of A, and

$$\mathrm{ran}(A) := \{Ax : x \in X\} \subseteq Y$$

be the *range* of A.

Remark 4.33. We may also use the notations $\ker A$ and $\mathrm{ran}\, A$, without brackets around the operator.

Remark 4.34. The range of an operator is also often called its *image*, and denoted by $\mathrm{Im}(A)$.

Exercise 4.35. Show that for a linear operator $A : \mathrm{dom}(A)(\subseteq X) \to Y$, $\ker(A)$ and $\mathrm{ran}(A)$ are linear subspaces in X and Y, respectively.

Lemma 4.36. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator between vector spaces X, Y. The following are equivalent:

(i) A is injective, i.e., $x_1, x_2 \in \mathrm{dom}(A)$, $Ax_1 = Ax_2 \Longrightarrow x_1 = x_2$.

(ii) $\ker A = \{0\}$.

(iii) There exists a linear operator $B : \mathrm{ran}(A) \to X$ such that $BA = I_{\mathrm{dom}(A)}$.

(iv) There exists a linear operator $B : \mathrm{ran}(A) \to X$ such that $AB = I_{\mathrm{ran}(A)}$.

(v) There exists a unique linear operator $A^{-1} : \mathrm{ran}(A) \to X$ such that

$$A^{-1}A = I_{\mathrm{dom}(A)}, \qquad AA^{-1} = I_{\mathrm{ran}(A)}.$$

Definition 4.37. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator between vector spaces X, Y. We say that A is *invertible* if any (and hence all) of (i)–(v) in Lemma 4.36 hold, and we call A^{-1} the *inverse* of A.

Definition 4.38. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a map between sets X, Y. The *graph* of A is

$$\mathrm{graph}(A) := \{(x, Ax) : x \in \mathrm{dom}(A)\} \subseteq X \times Y.$$

Remark 4.39. Clearly, if X, Y are vector spaces and A is linear then $\mathrm{graph}(A)$ is a linear subspace of $X \times Y$.

Exercise 4.40. Let X, Y be vector spaces,

$$V : X \times Y \to X \times Y, \qquad V : (x, y) \mapsto (y, x), \quad x \in X,\, y \in Y.$$

Show that a linear operator $A : \mathrm{dom}(A)(\subseteq X) \to Y$ is invertible if and only if $V\,\mathrm{graph}(A)$ is the graph of an operator, and in this case

$$V\,\mathrm{graph}(A) = \mathrm{graph}(A^{-1}). \tag{4.3}$$

Solution: Hidden.

4.5 Operator norm

Definition 4.41. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator between normed spaces X, Y. The *norm* of A is

$$\|A\| := \inf\{M \geq : \|Ax\| \leq M \|x\|, \, x \in \mathrm{dom}(A)\}.$$

A is called *bounded* if $\|A\| < +\infty$, and *unbounded* otherwise. (Note that the infimum of the empty set is $+\infty$.)

Exercise 4.42. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator between normed spaces X, Y. Show that

$$\begin{aligned} \|A\| &= \sup\{\|Ax\| : x \in \mathrm{dom}(A),\, \|x\| \leq 1\} \\ &= \sup\{\|Ax\| : x \in \mathrm{dom}(A),\, \|x\| = 1\} \\ &= \sup\left\{\frac{\|Ax\|}{\|x\|} : x \in \mathrm{dom}(A) \setminus \{0\}\right\}. \end{aligned}$$

Proposition 4.43. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator between normed spaces X, Y. The following are equivalent:

(i) A is continuous;

(ii) A is continuous at 0;

(iii) A is bounded.

Proof. □

Definition 4.44. For normed spaces X, Y, let

$$\mathcal{B}(X,Y) := \{A : X \to Y \text{ everywhere defined bounded linear}\},$$

and

$$\mathcal{B}(X) := \mathcal{B}(X,X).$$

Example 4.45. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, and $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}$ be a measurable function. The *multiplication operator* M_f may be defind on any of the $L^p(\mathcal{X}, \mathcal{A}, \mu)$ spaces as

$$\mathrm{dom}(M_f) := \{g \in L^p(\mathcal{X}, \mathcal{A}, \mu) : fg \in L^p(\mathcal{X}, \mathcal{A}, \mu)\}, \qquad M_f g := fg$$

for any $g \in L^p(\mathcal{X}, \mathcal{A}, \mu)$. We have

$$\|M_f\| = \|f\|_\infty ;$$

in particular, M_f is bounded if and only if f is essentially bounded.

Indeed...

Example 4.46. For $\underline{x} \in \mathbb{R}^d$, let $\pi_k(\underline{x}) := x_k$, $k \in [d]$, be the projection onto the k-th component. It is easy to see that π_k is a measurable function. The corresponding multiplication operator

$$Q_k := M_{\pi_k}, \qquad (Q_k f)(\underline{x}) := x_k f(\underline{x}), \qquad f \in L^2(\mathbb{R}^d),$$

is called in quantum mechanics the k-th *position operator* of a particle moving in $\mathbb{R}^d$. According to the previous example, $\|Q_k\| = +\infty$, i.e., Q_k is unbounded.

Exercise 4.47. Let $f : \mathbb{R}^d \to \mathbb{C}$ be a measurable *locally integrable function*, i.e., for any compact set $K \subseteq \mathbb{R}^d$, $\int_k |f| \, d\lambda < +\infty$. Then f defines a linear functional $\langle f|$ as

$$\mathrm{dom}(\langle f|) := C_0^\infty(\mathbb{R}^d) \subseteq L^p(\mathbb{R}^d), \qquad \langle f| \, g := \langle f, g\rangle := \int_{\mathbb{R}^d} \overline{f} g \, d\lambda, \qquad g \in C_0^\infty(\mathbb{R}^d),$$

where $p \in [1, +\infty)$. Show that $\langle f|$ is bounded if and only if $f \in L^{\frac{p-1}{p}}(\mathbb{R}^d)$.
(Hint: Use the Hölder inequality for the "if" part.)

For a function $f : \mathbb{R}^d \to \mathbb{C}$, let

$$\partial_k f(\underline{x}) := \frac{\partial}{\partial x_k} f(\underline{x}) := \lim_{t \to +\infty} \frac{f(\underline{x} + t\mathbf{1}_{\{k\}}) - f(\underline{x})}{t}$$

if the limit exists; this is the k-th *partial derivative* of f at $\underline{x}$. If f is locally integrable and its d-th partial derivative exists at every point $\underline{x} \in \mathbb{R}^d$, and $\partial_d f$ is continuous, then for every $g \in C_0^\infty(\mathbb{R}^d)$,

$$\begin{aligned}
\langle f | \, \partial_d g &= \int_{\mathbb{R}^d} \overline{f} \, \partial_d g \, d\lambda \\
&= \int_{\mathbb{R}^{d-1}} \left(\int_{\mathbb{R}} \overline{f}(x_1, \ldots, x_d) \partial_d g(x_1, \ldots, x_d) \, d\lambda(x_d) \right) \, d\lambda(x_1, \ldots, x_{d-1}) \\
&= \int_{\mathbb{R}^{d-1}} \left(\underbrace{\left[\overline{f} g \right]_{-\infty}^{+\infty}}_{=0} - \int_{\mathbb{R}} \overline{\partial_d f(x_1, \ldots, x_d)} g(x_1, \ldots, x_d) \, d\lambda(x_d) \right) \, d\lambda(x_1, \ldots, x_{d-1}) \\
&= - \int_{\mathbb{R}^d} \overline{\partial_d f}(x_1, \ldots, x_d) g(x_1, \ldots, x_d) \, d\lambda(x_1, \ldots, x_d) \\
&= \langle -\partial_d f | \, g.
\end{aligned}$$

Similarly, if the k-th partial derivative of f exists at every point $\underline{x} \in \mathbb{R}^d$, and $\partial_k f$ is continuous, then $\langle f | \, \partial_k g = - \langle \partial_k f | \, g$ for every $g \in C_0^\infty(\mathbb{R}^d)$. This motivates to introduce the following:

Definition 4.48. In the setting of Exercise 4.47, the linear functional

$$\mathrm{dom}(\partial_k f) := C_0^\infty(\mathbb{R}^d), \qquad \partial_k f : \, g \mapsto - \int_{\mathbb{R}^d} \overline{f} \, \partial_k g \, d\lambda, \qquad g \in C_0^\infty(\mathbb{R}^d),$$

is called the *distributional derivative* of f.

Example 4.49. The k-th *momentum operator* of a particle moving in $\mathbb{R}^d$ is defined as

$$\mathrm{dom}(P_k) := \{ f \in L^2(\mathbb{R}^d) : \, \partial_k f \in L^2(\mathbb{R}^d) \}, \qquad P_k f := -i \partial_k f.$$

Exercise 4.50. Show that $\mathrm{dom}(P_k)$ is dense, and P_k is unbounded, i.e., $\|P_k\| = +\infty$.

Exercise 4.51. The *evaluation functional* at point x is defined as

$$\mathrm{dom}(\delta_x) := \{ f \in L^p(\mathbb{R}^d) : \, f \text{ continuous at } x \}, \qquad \delta_x f := f(x).$$

Show that for any $x \in \mathbb{R}^d$, $\mathrm{dom}(\delta_x)$ is dense, and δ_x is unbounded, i.e., $\|\delta_x\| = +\infty$, for any $p \in [1, +\infty)$.

4.6 Closed operators

The following relation between boundedness of the operator and closedness of its graph is easy to see:

Lemma 4.52. Consider a linear operator $A : \mathrm{dom}(A)(\subseteq X) \to Y$ between normed spaces X, Y. If $\mathrm{dom}(A)$ is closed and A is bounded then $\mathrm{graph}(A)$ is closed.

Proof. Let $(x_n)_{n\in\mathbb{N}} \subseteq \mathrm{dom}(A)$ be such that

$$x_n \xrightarrow[n\to+\infty]{} x \in X, \qquad Ax_n \xrightarrow[n\to+\infty]{} y \in Y.$$

Closedness of $\mathrm{dom}(A)$ then implies $x \in \mathrm{dom}(A)$, and boundedness of A implies

$$Ax = \lim_{n\to+\infty} Ax_n = y.$$

Thus, $(x, y) \in \mathrm{graph}(A)$. □

A highly non-trivial fact, called the closed graph theorem, is that if X and Y are Banach spaces then the above statement can be extended to the following:

Theorem 4.53. (Closed graph theorem)
Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator between Banach spaces X, Y. Any two of the following properties imply the third:

(i) $\mathrm{dom}(A)$ is closed;

(ii) A is bounded;

(iii) $\mathrm{graph}(A)$ is closed.

Proof. We have seen in Lemma 4.52 that (i) and (ii) implies (iii). The implication (ii)+(iii)$\Longrightarrow$(i) is easy, and we leave its proof to Exercise 4.54.

The implication (i)+(iii)$\Longrightarrow$(ii) is a consequence of the Baire category theorem, and its proof is beyond the scope of these notes. We refer the interested reader to [**?**, Theorem III.12]. □

Exercise 4.54. Prove the implication (ii)+(iii)$\Longrightarrow$(i) in Theorem 4.53.

According to the above theorem, having a closed graph is a weaker property than being everywhere defined and bounded. Most operators that we encounter in quantum physics are defined on a dense but not closed subspace, and they are not bounded. However, they are self-adjoint, and, as we will see below, this implies that their graphs are closed. Hence, it is exactly this weaker property of having a closed graph that will be relevant for our investigations. Therefore, we introduce the following:

Definition 4.55. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator between normed spaces X and Y. We say that A is *closed* if $\mathrm{graph}(A)$ is closed.

We say that A is *closable* if $\overline{\mathrm{graph}(A)}$ is the graph of an operator $\overline{A}$, which we call the *closure* of A.

Exercise 4.56. Show that A is closable if and only if it has a closed extension, and its closure is the smallest such closed extension.

Exercise 4.57. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be an injective linear operator between normed spaces X, Y. Show that A is closable if and only if A^{-1} is closable, and in this case $\overline{A}$ is invertible with

$$\overline{A}^{-1} = \overline{A^{-1}}. \tag{4.4}$$

Solution: Hidden.

As it turns out, bounded operators into a Banach space are always closable:

Proposition 4.58. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator from a normed space X into a Banach space Y. If A is bounded then it has a unique extension to $\overline{\mathrm{dom}(A)}$ with the same norm, and it is exactly the closure of A. Moreover,

$$\mathrm{dom}(\overline{A}) = \overline{\mathrm{dom}(A)}.$$

Proof. Let $x \in \overline{\mathrm{dom}(A)}$. Then there exists a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathrm{dom}(A)$ such that $\lim_n x_n = x$. Boundedness of A implies that $\|Ax_n - Ax_m\| \leq \|A\| \, \|x_n - x_m\| \to 0$ as $n, m \to +\infty$, i.e., $(Ax_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. Since Y is complete, there exist a $y \in Y$ such that $y = \lim_n Ax_n$. It is easy to see that (i) this y does not depend on the particular sequence $(x_n)_{n\in\mathbb{N}}$ converging to y, and hence we can introduce the notation $y =: \hat{A}x$; (ii) the map $x \mapsto \hat{A}x$ is linear; (iii) if $x \in \mathrm{dom}(A)$ then $\hat{A}x = Ax$. This proves that $\hat{A}$ is an extension of A, $\mathrm{dom}(\hat{A}) = \overline{\mathrm{dom}(A)}$, and $\left\|\hat{A}\right\| = \|A\|$ is easy to verify. Since $\mathrm{dom}(\hat{A})$ is closed and $\hat{A}$ is bounded, Lemma 4.52 implies that $\hat{A}$ is closed. Hence, A is closable and $\overline{A} \subseteq \hat{A}$. Finally, if $x \in \mathrm{dom}(\hat{A})$ then, by the above, there exists a exists a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathrm{dom}(A)$ such that $\lim_n x_n = x$ and $(Ax_n)_{n\in\mathbb{N}}$ is convergent, i.e., $x \in \mathrm{dom}(\overline{A})$. Thus, $\hat{A} \subseteq \overline{A}$, and the proof is complete. □

Exercise 4.59. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a closable linear operator between normed spaces X, Y.

(i) Show that

$$\mathrm{dom}(\overline{A}) \subseteq \overline{\mathrm{dom}(A)}, \qquad \mathrm{ran}(\overline{A}) \subseteq \overline{\mathrm{ran}(A)}.$$

(ii) Show that if X, Y are Banach spaces then

$$\mathrm{dom}(\overline{A}) = \overline{\mathrm{dom}(A)} \iff A \text{ is bounded.}$$

Exercise 4.60. Let $A : \mathrm{dom}(A)(\subseteq X) \to Y$ be a linear operator from a normed space X into a Banach space Y. Show that if there exist positive constants $m, M \in (0, +\infty)$ such that

$$m\|x\| \leq \|Ax\| \leq M\|x\|, \qquad x \in \mathrm{dom}(A),$$

then A is closable, and

$$\mathrm{dom}(\overline{A}) = \overline{\mathrm{dom}(A)}, \qquad \mathrm{ran}(\overline{A}) = \overline{\mathrm{ran}(A)}, \tag{4.5}$$

and $m\|x\| \leq \left\|\overline{A}x\right\| \leq M\|x\|$, $x \in \mathrm{dom}(\overline{A})$. Conclude that an isometric operator is always closable and its closure is again an isometric operator.

Solution: Hidden.

4.7 Hilbert spaces

Definition 4.61. Let V, W be vector spaces over the scalar field $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. We say that a map $\gamma : V \times V \to W$ is a *sesquilinear map* if it is linear in its second variable and conjugate linear in the first variable:

$$\begin{aligned}
&\gamma(y, c_1x_1 + c_2x_2) = c_1\gamma(y, x_1) + c_2\gamma(y, x_2), \qquad x_1, x_2, y \in V, \quad c_1, c_2 \in \mathbb{K};\\
&\gamma(c_1y_1 + c_2y_2, x) = \overline{c}_1\gamma(y_1, x) + \overline{c}_2\gamma(y_2, x), \qquad x, y_1, y_2 \in V, \quad c_1, c_2 \in \mathbb{K}.
\end{aligned}$$

When $\mathbb{K} = \mathbb{R}$, $\overline{c} = c$ for every $c \in \mathbb{K}$, and conjugate linearity is the same as linearity, so a sesquilinear map is simply a bilinear map.

Remark 4.62. In the Mathematics literature the convention is usually the opposite to the above, i.e., a sesquilinear map is defined to be linear in its first, and conjugate linear in its second variable. We follow the Physics convention, as it allows for the use of the very convenient Dirac formalism (see later).

Exercise 4.63. Let V, W be complex vector spaces, and $\gamma : V \times V \to W$ be a sesquilinear map. Show that it satisfies the *(complex) polarization identity*

$$\gamma(x, y) = \frac{1}{4}\sum_{k=1}^{4} i^k \gamma(i^k x + y, i^k x + y), \qquad x, y \in V. \tag{4.6}$$

Prove that if $A : V \to V$ is a linear operator then

$$\gamma(x, Ay) = \frac{1}{4}\sum_{k=1}^{4} i^k \gamma(i^k x + y, A(i^k x + y)), \qquad x, y \in V. \tag{4.7}$$

Solution: (4.6) follows by a straightforward computation. Note that $\gamma_A(x,y) := \gamma(x,Ay)$ is again a sesquilinear from, and (4.7) is nothing else but the polarization identity for γ_A.

Exercise 4.64. Let V, W be real vector spaces, and $\gamma : V \times V \to W$ be a bilinear map. Show that it satisfies the *(real) polarization identity*

$$\frac{1}{4}\left[\gamma(x+y,x+y) - \gamma(x-y,x-y)\right] = \frac{1}{2}\gamma(x,y) + \frac{1}{2}\gamma(y,x).$$

Conclude that if γ is symmetric, i.e., $\gamma(x,y) = \gamma(y,x)$, $x, y \in V$, then

$$\begin{aligned}\gamma(x,y) &= \frac{1}{4}\left[\gamma(x+y,x+y) - \gamma(x-y,x-y)\right] \\ &= \sum_{k=0}^{1}(-1)^k\gamma((-1)^k x + y, (-1)^k x + y).\end{aligned}$$

We will mainly be interested in the case where $W = \mathbb{K}$:

Definition 4.65. Let V be a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. A sesquilinear map $\gamma : V \times V \to \mathbb{K}$ is called a *sesquilinear form*.

We say that a sesquilinear form $\gamma : V \times V \to \mathbb{K}$ is

- *Hermitian*, if $\gamma(y,x) = \overline{\gamma(x,y)}$, $x, y \in V$;
- *positive semi-definite*, if $\gamma(x,x) \in \mathbb{R}_{\geq 0}$, $x \in V$;
- *positive definite*, if $\gamma(x,x) \in \mathbb{R}_{>0}$, $x \in V \setminus \{0\}$.

When $\mathbb{K} = \mathbb{R}$, a sesquilinear form is also called a *bilinear form*.

Remark 4.66. Note that when $\mathbb{K} = \mathbb{R}$, a sesquilinear form $\gamma : V \times V \to \mathbb{K}$ is Hermitian if and only if it is *symmetric*, i.e., $\gamma(y,x) = \gamma(x,y)$, $x, y \in V$.

Remark 4.67. Note that for a sesquilinear form $\gamma : V \times V \to \mathbb{K}$, $\gamma(0,0) = 0$, and hence for a positive definite sesquilinear form we have $\gamma(x,x) \geq 0$, with equality if and only if $x = 0$.

Exercise 4.68. Let $\gamma : V \times V \to \mathbb{K}$ be a positive semi-definite sesquilinear form.

(i) Prove that if $\mathbb{K} = \mathbb{C}$ then γ is Hermitian.

(Hint: Use the polarization identity.)

Assume for the rest that γ is Hermitian also when $\mathbb{K} = \mathbb{R}$.

(ii) Show that for any $x, y \in V$,

$$\gamma(x+y, x+y) = \gamma(x,x) + 2\,\mathfrak{Re}\,\gamma(x,y) + \gamma(y,y).$$

(iii) Prove that γ satisfies the *Cauchy-Schwarz inequality*

$$|\gamma(x,y)|^2 \le \gamma(x,x)\gamma(y,y), \qquad x, y \in V. \tag{4.8}$$

(Hint: Use that $\gamma(x+ty, x+ty) \ge 0$ for all $t \in \mathbb{R}$.)

(iv) Prove that $\|x\|_\gamma := \gamma(x,x)^{1/2}$, $x \in V$, defines a semi-norm on V, and it is a norm if and only if γ is positive definite.

(Hint: Use the Cauchy-Schwarz inequality to prove the triangle inequality.)

Assume for the rest that γ is positive definite.

(v) Prove that the Cauchy-Schwarz inequality (4.8) holds with equality if and only if x and y are linearly dependent, if and only if there exists a constant $\lambda \in \mathbb{C}$ such that $x = \lambda y$ or $y = \lambda x$.

(vi) Prove that for $x \in V$,

$$x = 0 \iff \gamma(x,y) = 0 \qquad \forall y \in V.$$

(vii) Prove that for two linear operators $A, B : V \to V$,

$$\begin{aligned} A = B &\iff \gamma(y, Ax) = \gamma(y, Bx) \ \ \forall x, y \in V \\ &\iff \gamma(x, Ax) = \gamma(x, Bx) \ \ \forall x \in V. \end{aligned}$$

Solution: Hidden.

Definition 4.69. A positive semi-definite Hermitian sesquilinear form on a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ is called a *(real or complex) inner product*. We will normally denote an inner product by $\langle ., . \rangle$.

A pair $(\mathcal{H}, \langle ., . \rangle)$, where $\langle ., . \rangle$ is an inner product on the vector space $\mathcal{H}$, is called an *inner product space*. We will normally use the shorter terminology "$\mathcal{H}$ is an inner product space".

When $\mathcal{H}$ is a finite-dimensional inner product space, it is also called a *finite-dimensional Hilbert space* (with respect to that inner product).

Definition 4.70. Let $\mathcal{H}$ be an inner product space. We say that two vectors $x, y \in \mathcal{H}$ are *orthogonal*, in notation, $x \perp y$, if $\langle x, y \rangle = 0$.

We may rewrite the findings of Exercises 4.63 and 4.68 as

- An inner product on $\mathcal{H}$ defines a norm on $\mathcal{H}$ by
$$\|x\| := \sqrt{\langle x, x\rangle}, \qquad x \in \mathcal{H}. \tag{4.9}$$
- $\langle x, y\rangle = \dfrac{1}{4}\displaystyle\sum_{k=1}^{4} i^k \left\|i^k x + y\right\|^2, \quad x, y \in \mathcal{H} \qquad$ (when $\mathbb{K} = \mathbb{C}$),

  $\langle x, y\rangle = \dfrac{1}{4}\displaystyle\sum_{k=0}^{1} (-1)^k \left\|(-1)^k x + y\right\|^2, \quad x, y \in \mathcal{H} \qquad$ (when $\mathbb{K} = \mathbb{R}$),

  (polarization)
- $|\langle x, y\rangle| \leq \|x\|\,\|y\|$ with equality if and only if x, y are linearly dependent (Cauchy-Schwarz)
- $x = 0 \iff x \perp y \ \ \forall y \in \mathcal{H}$.
- For $A, B \in \mathrm{Lin}(\mathcal{H})$: $A = B \iff \langle x, Ax\rangle = \langle x, Bx\rangle \ \ \forall x \in \mathcal{H}$.

The following is the canonical example of a finite-dimensional Hilbert space:

Example 4.71. Let $\mathbb{C}^d = \{\underline{z} = (z_1, \ldots, z_d) : z_i \in \mathbb{C}\}$ denote the vector space of d-tuples of complex numbers, with the usual coordinate-wise addition and multplication of scalars. Then
$$\langle \underline{z}, \underline{w}\rangle := \sum_{i=1}^{d} \bar{z}_i w_i, \qquad \underline{z}, \underline{w} \in \mathbb{C}^d,$$
defines an inner product on $\mathbb{C}^d$, with induced norm
$$\|\underline{z}\| = \sqrt{\sum_{i=1}^{d} |z_i|^2}.$$
These are the standard Euclidean inner product and norm on $\mathbb{C}^d$.

Example 4.72. The following is a slightly more abstract version of the previous example. For a finite set Ω, let
$$\mathcal{H}_\Omega := l^2(\Omega) := \mathbb{C}^\Omega, \qquad \langle f, g\rangle := \sum_{\omega \in \Omega} \overline{f(\omega)} g(\omega), \qquad f, g \in l^2(\Omega).$$
That is, the Hilbert space associated to Ω is the space of complex-valued functions on Ω with its standard inner product. The choice $\Omega := [d]$ gives back Example 4.71.

Exercise 4.73. Let $\mathcal{H}$ be an inner product space and $\|.\|$ be the induced norm. Show that for any $x, y \in \mathcal{H}$,

$$\|x+y\|^2 = \|x\|^2 + 2\,\mathfrak{Re}\,\langle x, y\rangle + \|y\|^2\,,$$

and conclude that the norm $\|.\|$ satisfies the *parallelogram identity*:

$$\|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right), \qquad x, y \in \mathcal{H}. \tag{4.10}$$

Conclude that if $x, y \in \mathcal{H}$ are unit vectors then

$$\left\|\frac{x+y}{2}\right\|^2 = 1 - \frac{1}{4}\|x-y\|^2\,. \tag{4.11}$$

Remark 4.74. John von Neumann showed that the parallelogram identity (4.10) is also sufficient for a norm to be derived from an inner product.

Remark 4.75. A normed space $(V, \|.\|)$ is called *uniformly convex* if there exists a function $\varepsilon : [0,2] \to [0,+\infty)$ such that $\varepsilon(t) > 0$ for $t > 0$, $\lim_{\varepsilon \searrow 0} = 0 = \varepsilon(0)$, and for any unit vectors $x, y \in V$,

$$\left\|\frac{x+y}{2}\right\| \leq 1 - \varepsilon(\|x-y\|).$$

It is easy to see from (4.11) that any Hilbert space is uniformly convex.

Exercise 4.76. (i) Show that for any vector x in an inner product space $\mathcal{H}$,

$$\|x\| = \sup\{|\,\langle y, x\rangle\,| : \|y\| \leq 1,\, y \in \mathcal{H}\}.$$

(ii) Show that for any linear operator $A \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ between inner product spaces $\mathcal{H}, \mathcal{K}$,

$$\|A\| = \sup\{|\,\langle y, Ax\rangle\,| : \|x\|, \|y\| \leq 1,\, x \in \mathcal{H},\, y \in \mathcal{K}\}.$$

Score: 2+2=4 points.

Definition 4.77. We say that a Hermitian positive semi-definite sesquilinear form on a vector space V is a *semi-inner product*.

Note that every inner product is also a semi-inner product, by definition. A semi-inner product $(.,.)$ that is not an inner product is different from an inner product in that $(x,x) = 0$ can happen even if $x \neq 0$, i.e., a non-zero vector may be orthogonal to itself. The polarization identities, and (ii)–(iii) of Exercise 4.68 still hold for semi-inner products.

As the following exercise shows, an inner product space can be obtained from a semi-inner product space by factoring out the vectors that are orthogonal to themselves.

Exercise 4.78. Let $(.,.)$ be a semi-inner product on a vector space V. Define

$$\mathcal{N} := \{x \in V : (x,x) = 0\}.$$

Show that $\mathcal{N}$ is a subspace of V, and let $\mathcal{H} := V/\mathcal{N}$ be the factor space. For every $v \in V$, let $[v] := \{v + x : x \in \mathcal{N}\}$ denote the equivalence class of v. Show that

$$\langle [v], [w] \rangle := (v,w)$$

defines an inner product on $\mathcal{H}$.

The concept of (semi-)inner product can be generalized to more than two vectors in the following way:

Definition 4.79. Let $\langle .,. \rangle$ be a semi-inner product on $\mathcal{H}$, and $(v_i)_{i=1}^r$ be a sequence of vectors in $\mathcal{H}$. We define the corresponding *Gram matrix* $G(\{v_i\}_{i=1}^r)$ as

$$G(\{v_i\}_{i=1}^r) := \{\langle v_i, v_j \rangle\}_{i,j=1}^r .$$

Recall that a matrix $A \in \mathbb{C}^{r\times r}$ is called positive semi-definite (PSD) if $0 \leq \langle \underline{x}, A\underline{x} \rangle = \sum_{i,j=1}^r \overline{x}_i A_{ij} x_j$ for all $\underline{x} \in \mathbb{C}^r$, and positive definite if $0 < \langle \underline{x}, A\underline{x} \rangle$ for all $\underline{x} \in \mathbb{C}^r \setminus \{0\}$.

Exercise 4.80. (i) Show that for any $(v_i)_{i=1}^r$, the corresponding Gram matrix $G(\{v_i\}_{i=1}^r)$ is positive semidefinite.

(ii) Show that if $\langle .,. \rangle$ is an inner product then the Gram matrix $G(\{v_i\}_{i=1}^r)$ is positive definite if and only if $\{v_i\}_{i=1}^r$ is linearly independent. Conclude that $\operatorname{rk} G(\{v_i\}_{i=1}^r) = \dim \operatorname{span}\{v_i\}_{i=1}^r$.

Solution: Let $G_{k,l} := G(\{v_i\}_{i=1}^r)_{k,l} = \langle v_k, v_l \rangle$. For any $c_1, \ldots, c_r \in \mathbb{C}$,

$$\sum_{k,l=1}^r \overline{c}_k G_{k,l} c_l = \left\langle \sum_{k=1}^r c_k v_k, \sum_{l=1}^r c_l v_l \right\rangle = \left\| \sum_{k=1}^r c_k v_k \right\|^2 \geq 0,$$

showing the positive semidefinitess of G. Moreover, $\sum_{k,l=1}^r \overline{c}_k G_{k,l} c_l = 0 \iff \|\sum_{k=1}^r c_k v_k\| = 0$, from which the assertion about the positive definiteness is immediate.

Now, let $v_{i_1}, \ldots, v_{i_m}$, $1 \leq i_1 < \ldots < i_m$, be a basis for $\operatorname{span}\{v_i\}_{i=1}^r$. Then all the columns of G can be expressed as linear combinations of the $i_1, \ldots, i_m$ columns of G, and hence $\operatorname{rk} G \leq m$. On the other hand, the submatrix $\tilde{G} := G(\{v_{i_k}\}_{k=1}^m)$ is positive definite by the above considerations, and hence its columns are linearly independent. This in turn yields that the $i_1, \ldots, i_m$ columns of G are linearly independent, and thus $\operatorname{rk} G \geq m$. □

The converse of the above is also true, in the sense that every PSD matrix is the Gram matrix of some sequence of vectors in a Hilbert space; see Exercise 4.277.

Exercise 4.81. Prove the Cauchy-Schwarz inequality and its equality condition from Exercise 4.80.

Solution: Let $r := 2$, $v_1 := x$, $v_2 := y$. Then

$$\det G(x,y) = \langle x,x\rangle \langle y,y\rangle - \langle x,y\rangle \langle y,x\rangle = \|x\|^2 \|y\|^2 - |\langle x,y\rangle|^2.$$

Positive semi-definiteness of the Gram matrix implies the non-negativity of the determinant, which is equivalent to the Cauchy-Schwarz inequality. Moreover, the determinant is zero if and only if $\operatorname{rk} G(x,y) < 2$, which is equivalent to x and y being linearly dependent, by Exercise 4.80.

Definition 4.82. Let $(\mathcal{H}, \langle .,. \rangle)$ be a semi-inner product space, and $v_1, \ldots, v_r \in \mathcal{H}$. We say that $u_1, \ldots, u_r$ is a *dual system* to $\{v_i\}_{i=1}^r$ if $\langle u_j, v_i\rangle = \delta_{i,j}$ for all $i, j \in [r]$.

Exercise 4.83. Let $(\mathcal{H}, \langle .,. \rangle)$ be an inner product space, and $v_1, \ldots, v_r \in \mathcal{H}$.

(i) Show that a dual system exists for $\{v_i\}_{i=1}^r$ if and only if $\{v_i\}_{i=1}^r$ is linearly independent, and in this case $\{u_i\}_{i=1}^r$ is linearly independent, too. Moreover, any dual system is of the form

$$u_j = \hat{v}_j + v_j^\perp, \quad \text{where} \quad \hat{v}_j := \sum_{i=1}^r \left(G^{-1}\right)_{ij} v_i, \quad \text{and} \quad v_j^\perp \perp \operatorname{span}\{v_1, \ldots, v_r\}.$$

(ii) Show that the dual system is unique if and only if $\{v_i\}_{i=1}^r$ is a basis in $\mathcal{H}$.

(iii) Let $\{v_i\}_{i=1}^r$ be a linearly independent system and $\{u_i\}_{i=1}^r$ a dual system. Show that

$$G\left(\{u_i\}_{i=1}^r\right) \geq G\left(\{\hat{v}_i\}_{i=1}^r\right) = G\left(\{v_i\}_{i=1}^r\right)^{-1},$$

and equality holds in the first inequality if and only if $\operatorname{span}\{u_1, \ldots, u_r\} = \operatorname{span}\{v_1, \ldots, v_r\}$.

Solution: Hidden.

4.8 The Dirac formalism

For a vector $x \in \mathcal{H}$ in an inner product space $\mathcal{H}$, let

$$|x\rangle : \lambda \mapsto \lambda x, \qquad \lambda \in \mathbb{C},$$
$$\langle x| : y \mapsto \langle x, y\rangle, \qquad y \in \mathcal{H}.$$

Then it is easy to see that

$$|x\rangle \in \operatorname{Lin}(\mathbb{K}, \mathcal{H}), \qquad \langle x| \in \operatorname{Lin}(\mathcal{H}, \mathbb{K}) = \mathcal{H}';$$

in particular, $\langle x|$ is a linear functional on $\mathcal{H}$.

Remark 4.84. In physics, the vector x is often identified with the linear map $|x\rangle$. This notation was introduced by P.A.M. Dirac. The vector $\langle x| \in \mathrm{Lin}(\mathcal{H};\mathbb{C})$ is called a "bra" vector, while $|x\rangle \in \mathrm{Lin}(\mathbb{C};\mathcal{H})$ is a "ket" vector, from parts of the word "bracket".

It is also common to write a label instead of a vector in a ket; for instance, the canonical basis of $\mathbb{C}^2$ is often denoted by

$$|0\rangle := \begin{bmatrix}0\\1\end{bmatrix}, \qquad |1\rangle := \begin{bmatrix}1\\0\end{bmatrix},$$

where 0 and 1 are interpreted as the logical 0 and 1 (see Section **??** for details). Clearly, the zero in $|0\rangle$ here is not the zero vector (since then we would have $|0\rangle = 0$), but a label.

Exercise 4.85. Let $\mathcal{H}, \mathcal{K}, \mathcal{L}$ be inner product spaces. Show that

(i) $\langle \lambda x_1 + \eta x_2| = \overline{\lambda}\langle x_1| + \overline{\eta}\langle x_2|, \qquad |\lambda x_1 + \eta x_2\rangle = \lambda|x_1\rangle + \eta|x_2\rangle$
for any $x_1, x_2 \in \mathcal{H}$, $\lambda, \eta \in \mathbb{K}$.

(ii) $\|\langle x|\| = \||x\rangle\| = \|x\|, \qquad x \in \mathcal{H}$.

(iii) $x \mapsto |x\rangle$ is a linear norm-preserving map from $\mathcal{H}$ into $\mathcal{B}(\mathbb{K}, \mathcal{H})$, with inverse $V \mapsto V1$.

(iv) $x \mapsto \langle x|$ is a conjugate linear norm-preserving map from $\mathcal{H}$ into $\mathcal{H}' = \mathcal{B}(\mathcal{H}, \mathbb{K})$.

Norm-preserving linear or conjugate linear maps are called *isometries*; we discuss them in detail in Section 4.21. In particular, the norm-preserving property implies that $x \mapsto \langle x|$ is injective. The Riesz representation theorem, which we prove in Section **??** shows that it is also surjective, and hence $x \mapsto \langle x|$ gives a conjugate linear isometric isomorphism between a Hilbert space $\mathcal{H}$ and its dual space $\mathcal{H}^*$ of continuous linear functionals.

Corollary 4.86. (Riesz representation theorem in finite dimension)
If $\mathcal{H}$ is a finite-dimensional inner product space then for any linear functional $\varphi \in \mathrm{Lin}(\mathcal{H}, \mathbb{K})$, there exists a unique $y_\varphi \in \mathcal{H}$ such that

$$\varphi(x) = \langle y_\varphi, x\rangle, \qquad x \in \mathcal{H}.$$

Proof. This is just a restatement of the surjectivity of the map $x \mapsto \langle x|$. □

Remark 4.87. We will see a different proof of the Riesz representation theorem for finite-dimensional Hilbert spaces in Proposition 4.100.

Remark 4.88. In fact, the statements in Proposition **??** and in Corollary 4.86 are valid also in infinite-dimensional Hilbert spaces.

Next, we turn to the composition of operators defined by bra and ket vectors. For any $x, y \in \mathcal{H}$, the composition $\langle x| \circ |y\rangle$ is a linear map on the one-dimensional Hilbert space $\mathbb{C}$, and hence it is given by the multiplication of a number. One can easily see that this number is $\langle x, y\rangle$, i.e.,

$$\langle x| \circ |y\rangle : \lambda \mapsto \lambda \langle x, y\rangle , \qquad \lambda \in \mathbb{C}.$$

Hence, we use the identification

$$\langle x|y\rangle := \langle x| \, |y\rangle = \langle x| \circ |y\rangle \equiv \langle x, y\rangle .$$

Remark 4.89. In the physics literature, the inner product $\langle x, y\rangle$ is usually denoted by $\langle x|y\rangle$. We interpret this in terms of the above identification.

Next, consider (possibly different) inner product spaces $\mathcal{H}$ and $\mathcal{K}$. For any $x \in \mathcal{H}$ and $y \in \mathcal{K}$,

$$|y\rangle\langle x| := |y\rangle \circ \langle x| \in \mathrm{Lin}(\mathcal{H}, \mathcal{K}), \qquad \text{acts as} \qquad |y\rangle\langle x| : z \mapsto \langle x, z\rangle \, y, \quad z \in \mathcal{H}.$$

Definition 4.90. The operator $|y\rangle\langle x|$ above is called a *diad*, or the *diadic product* of y and x.

Exercise 4.91. (i) $|x\rangle\langle y| \circ |z\rangle\langle w| = \langle y, z\rangle \, |x\rangle\langle w|, \qquad w \in \mathcal{H}, \, y, z \in \mathcal{K}, \, x \in \mathcal{L}.$

(ii) $\| |y\rangle\langle x| \| = \|x\| \cdot \|y\|, \qquad x \in \mathcal{H}, \, y \in \mathcal{K}.$

Remark 4.92. See Exercises **??** for further properties of the bra and ket vectors and diadic products.

4.9 Orthonormal systems and projections

Definition 4.93. Let $\mathcal{H}$ be an inner product space.

(i) A set of vectors $\{x_1, \ldots, x_r\} \subseteq \mathcal{H}$ is an *orthonormal system (ONS)* if they are pairwise orthogonal, and all of them have unit length, i.e.,

$$\langle x_i, x_j\rangle = \delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \qquad i, j \in [r].$$

(ii) An orthonormal system $\{e_i : i \in \mathcal{I}\}$ is called an *orthonormal basis (ONB)* for a subspace $\mathcal{K} \subseteq \mathcal{H}$ if $\mathrm{span}\{e_i : i \in \mathcal{I}\}$ is dense in $\mathcal{K}$.

Exercise 4.94. Let V be a finite-dimensional vector space, and let $\{e_1, \dots, e_d\}$ be an algebraic basis in it. Show that the bilinear map

$$\langle e_i, e_j \rangle := \delta_{i,j} := \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \qquad i, j \in [r],$$

can be uniquely extended to an inner product on V, with respect to which $\{e_1, \dots, e_d\}$ is an orthonormal basis.

Exercise 4.95. Let $\mathcal{H}$ be an inner product space.

(i) Let $\{e_1, \dots, e_r\}$ be an orthonormal system. Show that for any $x \in \mathcal{H}$,

$$x - \sum_{i=1}^{r} \langle e_i, x \rangle \, e_i \;\perp\; e_j, \quad j = 1, \dots, r.$$

(ii) Let $(x_i)_{i \in \mathcal{I}}$ be a sequence of linearly independent vectors, where $\mathcal{I} = [d]$ for some $d \in \mathbb{N}$, or $\mathcal{I} = \mathbb{N}$. Define

$$\begin{aligned} v_1 &:= x_1, & e_1 &:= v_1 / \|v_1\|, \\ v_2 &:= x_2 - \langle e_1, x_2 \rangle \, e_1, & e_2 &:= v_2 / \|v_2\|, \\ &\vdots \\ v_k &:= x_k - \sum_{i=1}^{k-1} \langle e_i, x_k \rangle \, e_i, & e_k &:= v_k / \|v_k\|, \\ &\vdots \end{aligned}$$

Show that $(e_i)_{\in \mathcal{I}}$ is an orthonormal system, and $\operatorname{span}\{x_i : i \in [k]\} = \operatorname{span}\{e_i : i \in [k]\}$ for every $k \in \mathcal{I}$.

(iii) Conclude that in every finite-dimensional inner product space $\mathcal{H}$ there exists an ONB, and the cardinality of any ONB in $\mathcal{H}$ is equal to $\dim \mathcal{H}$.

Definition 4.96. The procedure in Exercise 4.95 is called the *Gram-Schmidt orthogonalization.*

Exercise 4.97. Let $\mathcal{H}$ be an inner product space.

(i) Let $\{x_1, \dots, x_r\} \subseteq \mathcal{H}$ be such that the x_i are pairwise orthogonal and non-zero. Show that $\{x_1, \dots, x_r\}$ is linearly independent.

Score: 3 points.

(ii) Assume that $\mathcal{H}$ is finite-dimensional. Show that any orthonormal basis is also an algebraic basis of $\mathcal{H}$ (i.e., a maximal linearly independent set), and if $\{e_1, \ldots, e_d\}$ is an ONB then every vector $x \in \mathcal{H}$ can be uniquely expanded in the form

$$x = \langle e_1, x\rangle\, e_1 + \ldots + \langle e_d, x\rangle\, e_d, \tag{4.12}$$

and

$$\|x\|^2 = \sum_{i=1}^{d} |\, \langle e_i, x\rangle\, |^2.$$

Remark 4.98. (4.12) is exactly the coordinate expansion of x in the basis $\{e_1, \ldots, e_d\}$.

Example 4.99. For a finite set Ω, consider $l^2(\Omega)$ defined in Example 4.72. It is easy to see that the characteristic functions $\mathbf{1}_{\{\omega\}}$, $\omega \in \Omega$, are easily seen to form an ONB in this space that we call the canonical basis of $l^2(\Omega)$.

Proposition 4.100. (Riesz representation theorem in finite dimension)
Let $\mathcal{H}$ be a finite-dimensional Hilbert space. For any linear functional $\varphi \in \mathrm{Lin}(\mathcal{H}, \mathbb{K})$, there exists a unique $y_\varphi \in \mathcal{H}$ such that

$$\varphi(x) = \langle y_\varphi, x\rangle\,, \qquad x \in \mathcal{H}.$$

Proof. Let $(e_i)_{i=1}^d$ be an orthonormal basis in $\mathcal{H}$. For any $x \in \mathcal{H}$,

$$\varphi(x) = \varphi\left(\sum_{i=1}^{d} \langle e_i, x\rangle\, e_i\right) = \sum_{i=1}^{d} \langle e_i, x\rangle\, \varphi(e_i) = \langle y_\varphi, x\rangle\,,$$

if we define

$$y_\varphi := \sum_{i=1}^{d} \overline{\varphi(e_i)} e_i,$$

showing the existence of y_φ. If $y_1, y_2 \in \mathcal{H}$ are such that $\langle y_1, x\rangle = \langle y_2, x\rangle$ for all $x \in \mathcal{H}$ then $0 = \langle y_1 - y_2, x\rangle$ for all $x \in \mathcal{H}$ implies $y_1 = y_2$, showing the uniqueness of y_φ. □

Exercise 4.101. Let $E := \{e_1, \ldots, e_{d_1}\}$ and $F := \{f_1, \ldots, f_{d_2}\}$ be orthonormal bases in the finite-dimensional Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, respectively. Show that for any $A \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, the (i, j)-entry of the matrix of A in the pair of bases (E, F) is $\langle f_i, Ae_j\rangle$.

Definition 4.102. Let $\mathcal{H}$ be an inner product space, and $S \subseteq \mathcal{H}$ be a subset in $\mathcal{H}$. We say that a vector $y \in \mathcal{H}$ is orthogonal (or perpendicular) to S, in notation $y \perp S$, if it is orthogonal to every element of S, i.e., $y \perp x \ \forall x \in S$.

The set of vectors orthogonal to S is called the *orthocomplement* of S, and is denoted by $S^\perp$, i.e.,

$$S^\perp := \{y \in \mathcal{H} : \langle y, x\rangle = 0 \ \ \forall x \in S\}.$$

Exercise 4.103. Show that for any $S \subseteq \mathcal{H}$,

(i) $S^\perp$ is a subspace of $\mathcal{H}$.

(ii) $S^\perp = (\mathrm{span}(S))^\perp$.

(iii) $S \cap S^\perp = \{0\}$.

Score: 2+2+2=6 points.

Proposition 4.104. Let $\mathcal{H}$ be an inner product space, and $\mathcal{H}_0 \subseteq \mathcal{H}$ be a finite-dimensional subspace.

(i) For any $x \in \mathcal{H}$, there exists a unique $Px \in \mathcal{H}_0$ such that $x - Px \perp \mathcal{H}_0$.

(ii) The map $\mathcal{H} \ni x \mapsto Px$ is linear, and for any ONB $(e_i)_{i=1}^r$ in $\mathcal{H}_0$,

$$P = \sum_{i=1}^r |e_i\rangle\langle e_i| \,.$$

(iii) For any $x \in \mathcal{H}$, Px is the unique closest element of $\mathcal{H}_0$ to x, i.e.,

$$\|x - Px\| = \inf\{\|x - y\| : y \in \mathcal{H}_0\} =: d(x, \mathcal{H}_0),$$

and there are no other $y \in \mathcal{H}_0$ such that $\|x - y\| = d(x, \mathcal{H}_0)$.

Proof. By Exercise 4.95, there exists an ONB in $\mathcal{H}_0$, and let $(e_i)_{i=1}^r$ in $\mathcal{H}_0$ be any such ONB. Let $P := \sum_{i=1}^r |e_i\rangle\langle e_i|$, so that for any $x \in \mathcal{H}$, $Px = \sum_{i=1}^r \langle e_i, x\rangle\, e_i$. By Exercise 4.95, $x - Px \perp e_i$ for all $i \in [r]$, and hence $x - Px \perp \mathrm{span}(\{e_1, \ldots, e_r\}) = \mathcal{H}_0$, by Exercise 4.103. Assume now that $y \in \mathcal{H}_0$ is such that $x - y \perp \mathcal{H}_0$; then

$$\underbrace{(x - y) - (x - P_x)}_{\in \mathcal{H}_0^\perp} = \underbrace{Px - y}_{\in \mathcal{H}_0},$$

by which $Px - y \in \mathcal{H}_0 \cap \mathcal{H}_0^\perp$, and thus $Px - y = 0$, by Exercise 4.103. These prove (i) and (ii).

For any $y \in \mathcal{H}_0$, we have

$$\|x - y\|^2 = \|x - Px + Px - y\|^2 = \|x - P_x\|^2 + \|Px - y\|^2 \geq \|x - Px\|^2 \,,$$

where the second equality follows from $x - Px \perp \mathcal{H}_0 \ni Px - y$. Equality holds in the above inequality if and only if $y = Px$, proving (iii). $\square$

4.10 Orthonormal bases

For every $k \in \mathbb{N}$, let

$$\varphi_k(x) := \frac{1}{\sqrt{2\pi}} e^{ikx}.$$

Note that φ_k is periodic with period 2π. Using the parametrization $z = e^{ix}$ of the one-dimensional torus $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$, we see that

$$\varphi_k(x) = z^k.$$

For this reason, finite linear combinations of the above functions, i.e., functions of the form

$$\sum_{k=-n}^{n} c_k \varphi_k \quad \equiv \quad \sum_{k=-n}^{n} c_k \, \mathrm{id}_{\mathbb{T}}^k$$

are called *trigonometric polynomials*. Obviously, the trigonometric polynomials form a subspace in $L^2([-\pi, \pi])$ (equivalently in $L^2(\mathbb{T})$).

The following can be obtained by a straightforward computation:

Lemma 4.105. $(\varphi_k)_{k\in\mathbb{Z}}$ is an ONS in $L^2([-\pi, \pi])$ (equivalently in $L^2(\mathbb{T})$).

Proof. Exercise. □

As a consequence, for any $f \in L^2([-\pi, \pi])$, the infinite series

$$\mathcal{F}(f) := \sum_{k\in\mathbb{Z}} \langle \varphi_k, f \rangle \, \varphi_k = \lim_{n\to+\infty} \sum_{k=-n}^{n} \langle \varphi_k, f \rangle \, \varphi_k$$

is convergent, and $\mathcal{F}(f)$ is the projection onto the closure of the space of trigonometric polynomials.

Definition 4.106. We call $\mathcal{F}(f)$ the *Fourier expansion* of f, and the sequence $(\langle \varphi_k, f \rangle)_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$ its *Fourier transform*.

Our next aim is to show the following:

Theorem 4.107. The following equivalent statements are true:

(i) The functions $(\varphi_k)_{k\in\mathbb{Z}}$ form on ONB in $L^2([-\pi, \pi])$.

(ii) The subspace of trigonometric polynomials is dense in $L^2([-\pi, \pi])$.

(iii) For any $f \in L^2([-\pi, \pi])$, $\mathcal{F}(f) = f$.

(iv) $\mathcal{F}$ is a unitary from $L^2([-\pi,\pi])$ to $l^2(\mathbb{Z})$.

Proof. In Proposition 4.108 below, we will show that the trigonometric polynomials are dense in max-norm in the space of continuous periodic functions on $[-\pi,\pi]$. Note that for any $f \in L^2([-\pi,\pi])$,

$$\left\| f - f\mathbf{1}_{[-\pi+1/n,\pi+1/n]} \right\|_2^2 = \int_{[-\pi,-\pi+1/n]\cup[\pi-1/n,\pi]} |f|^2 \, d\lambda \to 0$$

as $n \to +\infty$, due to the monotone convergence theorem. Hence, given $\varepsilon > 0$, we can find n such that $\left\| f - f\mathbf{1}_{[-\pi+1/n,\pi+1/n]} \right\|_2 < \varepsilon/3$, and then, by Proposition **??**, a continuous function g such that $g(-\pi) = g(\pi) = 0$ and $\left\| f\mathbf{1}_{[-\pi+1/n,\pi+1/n]} - g \right\|_2 < \varepsilon/3$. Finally, by Proposition 4.108, we can find a trigonometric polynomial h such that

$$\|g - h\|_2 = \left(\int_{[-\pi,\pi]} |g-h|^2 \, d\lambda \right)^{1/2} \leq (2\pi)^{1/2} \|g-h\|_\infty < \varepsilon/3.$$

Putting the above together, $\|f - h\|_2 < \varepsilon$. Thus, the trigonometric polynomials are dense in $L^2([-\pi,\pi])$. □

Let us start preparing the proof of Proposition 4.108 by noting that the n-th partial sum in the Fourier expansion can be expressed as

$$\begin{aligned} \mathcal{F}_n(f)(x) := \left(\sum_{k=-n}^{n} \langle \varphi_k, f \rangle \, \varphi_k \right)(x) &= \sum_{k=-n}^{n} \frac{1}{\sqrt{2\pi}} e^{ikx} \int_{[-\pi,\pi]} \frac{1}{\sqrt{2\pi}} e^{-iky} f(y) \, d\lambda(y) \\ &= \int_{[-\pi,\pi]} f(y) \underbrace{\frac{1}{2\pi} \sum_{k=-n}^{n} e^{ik(x-y)}}_{=:D_n(x-y)} \, d\lambda(y), \end{aligned}$$

where D_n is the *Dirichlet kernel*. A simple summation of geometric terms yields

$$\begin{aligned} (2\pi) D_n(x) = e^{-inx} \sum_{k=0}^{2n} e^{ikx} &= e^{-inx} \frac{e^{i(2n+1)x} - 1}{e^{ix} - 1} \\ &= \frac{e^{i(n+1/2)x} - e^{-i(n+1/2)x}}{e^{ix/2} - e^{-ix/2}} = \frac{\sin(n+1/2)x}{\sin(x/2)}, \end{aligned}$$

whenever $x \neq 0$, and $(2\pi)D_n(0) = 2n+1 = \lim_{x\to 0} D_n(x)$.

As it turns out, $\mathcal{F}_n(f)$ does not converge in general to f in supremum norm, even for a periodic continuous function. One might try to remedy this by looking at the

Césaro sums

$$\tilde{\mathcal{F}}_N(f) := \frac{1}{N}\sum_{n=0}^{N-1} \mathcal{F}_n(f)$$

in the hope that they do converge (which is indeed the case, as we will see). This can be rewritten as

$$\tilde{\mathcal{F}}_N(f)(x) = \int_{[-\pi,\pi]} f(y) \underbrace{\frac{1}{N}\sum_{n=0}^{N-1} D_n(x-y)}_{=:F_n(x-y)}\, d\lambda(y) = \int_{[-\pi,\pi]} f(x-u)F_N(u)\, d\lambda(u),$$

where in the last step we assumed that f is a periodic continuous function, and $x-u$ is defined mod 2π. In the above, F_N is the *Fejér kernel*, given by

$$\begin{aligned}
F_N(x) &:= \frac{1}{N}\sum_{n=0}^{N-1} D_n(x) = \frac{1}{N}\frac{\sum_{n=0}^{N-1} e^{i(n+1/2)x} - \sum_{n=0}^{N-1} e^{-i(n+1/2)x}}{e^{ix/2} - e^{-ix/2}} \\
&= \frac{1}{N}\frac{e^{ix/2}\frac{e^{iNx}-1}{e^{ix}-1} - e^{-ix/2}\frac{e^{-iNx}-1}{e^{-ix}-1}}{e^{ix/2} - e^{-ix/2}} \\
&= \frac{1}{N}\frac{1}{e^{ix/2} - e^{-ix/2}}\left(e^{ix/2}\frac{e^{iNx}-1}{e^{ix}-1} - \underbrace{e^{-ix/2}\frac{e^{-iNx}-1}{e^{-ix}-1}}_{=e^{ix/2}\frac{e^{-iNx}-1}{1-e^{ix}}} \right) \\
&= \frac{1}{N}\underbrace{\frac{e^{ix/2}}{e^{ix/2} - e^{-ix/2}}}_{=1-e^{-ix}}\frac{1}{)(e^{ix}-1)}\left(e^{iNx} - 1 + e^{-iNx} - 1\right) \\
&= \frac{1}{N}\frac{2-2\cos Nx}{2-\cos x} = \frac{1}{N}\frac{\sin^2(Nx/2)}{\sin^2(x/2)}
\end{aligned}$$

when $x \neq 0$, and $F_N(0) = 2n+1 = \lim_{x\to 0} F_n(x)$.

It is clear from their definitions that

$$\int_{[-\pi,\pi]} D_n\, d\lambda = 1, \qquad \text{and hence} \qquad \int_{[-\pi,\pi]} F_n\, d\lambda = 1, \quad n \in \mathbb{N}. \tag{4.13}$$

Moreover, note that for any $0 < \delta < \pi$ there exists a $c_\delta > 0$ such that

$$x \in [-\pi,\pi] \setminus (-\delta,\delta) \implies \sin^2(x/2) > c_\delta \implies F_n(x) \leq \frac{1}{nc_\delta}. \tag{4.14}$$

Now we are in the position to prove the following:

Proposition 4.108. For any $f \in C_{\mathrm{per}}([-\pi,\pi])$,

$$\lim_{n\to+\infty} \left\| f - \tilde{\mathcal{F}}_n(f) \right\|_\infty = 0.$$

In particular, the trigonometric polynomials are dense in $(C_{\mathrm{per}}([-\pi,\pi]), \|.\|_\infty)$.

Proof. Let $M := \max_{x\in[-\pi,\pi]} |f(x)|$. Note that f is uniformly continuous, and hence for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|x-y| < \delta$ (where the difference is again mod 2π) implies $|f(x) - f(y)| < \varepsilon$. With such an ε and δ, we have

$$\begin{aligned}
\left|\tilde{\mathcal{F}}_n(f)(x) - f(x)\right| &= \left| \int_{[-\pi,\pi]} f(x-y)F_n(y)\,d\lambda(y) - f(x)\int_{[-\pi,\pi]} F_n(y)\,d\lambda(y) \right| \\
&= \left| \int_{[-\pi,\pi]} (f(x-y)-f(x))F_n(y)\,d\lambda(y) \right| \\
&\le \int_{[-\pi,\pi]} |f(x-y)-f(x)|F_n(y)\,d\lambda(y) \\
&= \int_{(-\delta,\delta)} \underbrace{|f(x-y)-f(x)|}_{\le\varepsilon} F_n(y)\,d\lambda(y) \\
&+ \int_{[-\pi,\pi]\setminus(-\delta,\delta)} \underbrace{|f(x-y)-f(x)|}_{\le 2M} F_n(y)\,d\lambda(y) \\
&\le \varepsilon + \frac{4M\pi}{nc_\delta} \xrightarrow[n\to+\infty]{} \varepsilon,
\end{aligned}$$

where we used (4.13) and (4.14). Note that the upper bounds are uniform in x, and hence we obtain that for large enough n, $\left\|\tilde{\mathcal{F}}_n(f) - f\right\|_\infty < 2\varepsilon$. □

Finally, we obtain that multi-variable trigonometric polynomials are dense in the L^2 space over closed boxes.

Proposition 4.109. For every $j \in [n]$, let $[a_i, b_i] \subseteq \mathbb{R}$ be a non-degenerate compact interval, and

$$\varphi_{j,k}(x) := \frac{1}{\sqrt{b_i - a_i}}, \qquad x \in [a_j, b_j],\ k \in \mathbb{Z}.$$

Then $\varphi_{\underline{k}} := \otimes_{j=1}^n \varphi_{j,k_j}$, $\underline{k} \in \mathbb{Z}^n$, is an ONB in $L^2\left(\times_{i=1}^n [a_i, b_i]\right)$.

Proof. It follows by a simple change of variables in the integrals that for every j, $(\varphi_{j,k})_{k\in\mathbb{Z}}$ is an ONB in $L^2([a-j, b_j])$, and hence the assertion follows from Corollary 4.29. □

Lemma 4.110. Assume that $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces which contain ONBs of the same cardinality, i.e., there exist ONBs $(e_{k,j})_{j\in J}$ in $\mathcal{H}_k$, $k = 1, 2$. Then there exists a unique unitary operator $U : \mathcal{H}_1 \to \mathcal{H}_2$ such that it maps the first ONB into the second one, i.e., $Ue_{1,j} = e_{2,j}$, $j \in J$.

Proof. Let $\mathcal{H}_{k,0}$ denote the subspace spanned by $\{e_{k,j}\}_{j\in J}$, which is the set of (necessarily finite) linear combinations of $\{e_{k,j}\}_{j\in J}$. Then $U_0\psi := \sum_{j\in J} \langle e_{1,j}, \psi\rangle\, e_{2,j}$, $\psi \in \mathcal{H}_{1,0}$, is a well-defined linear operator with $\operatorname{dom} U_0 = \mathcal{H}_{1,0}$, $\operatorname{ran} U_0 = \mathcal{H}_{2,0}$. Moreover, it is obviously isometric, and hence, by Exercise 4.60, $\overline{U}_0$ is an isometric operator with $\operatorname{dom} \overline{U}_0 = \overline{\mathcal{H}}_{1,0} = \mathcal{H}_1$, and $\operatorname{ran} \overline{U}_0 = \overline{\mathcal{H}}_{2,0} = \mathcal{H}_2$, and thus $\overline{U}_0$ is a unitary with the required property. □

Corollary 4.111. If $(e_j)_{j\in J}$ is an ONB in some Hilbert space $\mathcal{H}$ then there exists a unique unitary $U : \mathcal{H} \to l^2(J)$ such that $Ue_j = \mathbf{1}_{\{j\}}$, $j \in J$. In particular, $\mathcal{H}$ is isomorphic to $l^2(J)$.

Proof. Immediate from Lemma 4.110. □

Note that the unitary in Corollary 4.111 maps every vector in $\mathcal{H}$ into a function on J, as

$$(U\psi)(j) = \langle e_j, \psi\rangle\,.$$

That is, U can be considered as the coordinate expansion in the given basis.

4.11 The adjoint of operators

Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces and $A : \operatorname{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{K}$ be a linear operator. Let

$$\operatorname{dom}(A^*) := \{y \in \mathcal{K} : x \mapsto \langle y, Ax\rangle \text{ is bounded}\}$$

By Proposition 4.58, if $y \in \operatorname{dom}(A^*)$ then $x \mapsto \langle y, Ax\rangle$ has a unique extension to a bounded linear functional on $\overline{\operatorname{dom}(A)}$, and by the Riesz representation theorem, there exists a unique representing vector $A^*y \in \overline{\operatorname{dom}(A)}$ such that

$$\langle A^*y, x\rangle = \langle y, Ax\rangle\,, \qquad x \in \operatorname{dom}(A),\ y \in \operatorname{dom}(A^*).$$

It is clear that $\operatorname{dom}(A^*)$ is a linear subspace, and it is easy to see that the uniqueness of the representing vector implies that A^* is a linear on $\operatorname{dom}(A^*)$, i.e., A^* is a linear operator.

Definition 4.112. Let $A : \operatorname{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{K}$ be a linear operator. The operator A^* defined above is called the *adjoint* of A.

Exercise 4.113. Show that $\operatorname{dom}(A^*)$ is a linear subspace, on which A^* is a linear operator.

Remark 4.114. It is immediate from the definition that

$$\operatorname{ran} A^* \subseteq \overline{\operatorname{dom} A},$$

and

$$(\operatorname{ran} A)^\perp \subseteq \operatorname{dom} A^*; \quad \text{in fact,} \quad (\operatorname{ran} A)^\perp \subseteq \ker A^*.$$

Vice versa, if $y \in \ker A^*$ then $0 = \langle A^* y, x\rangle = \langle y, Ax\rangle$ for every $x \in \operatorname{dom}(A)$, i.e., $y \perp \operatorname{ran} A$. Thus,

$$(\operatorname{ran} A)^\perp = \ker A^*, \qquad \text{i.e.,} \qquad \mathcal{K} = \overline{\operatorname{ran} A} \oplus \ker A^*. \tag{4.15}$$

Remark 4.115. It is easy to see that

$$\|A^* y\| = \sup\{|\langle y, Ax\rangle| : x \in \operatorname{dom}(A),\ \|x\| \le 1\},$$

and hence

$$\begin{aligned}
\|A^*\| &= \sup\{\|A^* y\| : y \in \operatorname{dom}(A^*),\ \|y\| \le 1\} \\
&= \sup\{|\langle y, Ax\rangle| : x \in \operatorname{dom}(A),\ \|x\| \le 1,\ y \in \operatorname{dom}(A^*),\ \|y\| \le 1\} \\
&= \sup\{\sup\{|\langle y, Ax\rangle| : y \in \operatorname{dom}(A^*),\ \|y\| \le 1\} : x \in \operatorname{dom}(A),\ \|x\| \le 1,\} \\
&\le \sup\{\|Ax\| : x \in \operatorname{dom}(A),\ \|x\| \le 1,\} \\
&= \|A\|.
\end{aligned}$$

Moreover, if A is bounded then $\operatorname{dom}(A^*) = \mathcal{K}$, and the inequality above holds as an equality. That is,

$$A \text{ is bounded} \implies \operatorname{dom}(A^*) = \mathcal{K}, \quad \|A^*\| = \|A\| \implies A^* \in \mathcal{B}(\mathcal{K}, \mathcal{H}).$$

In general, A^* need not be bounded, but it is easy to see that if A is densely defined then A^* is closed. Moreover, A^* need not be densely defined even if A was; in fact, it is densely defined if and only if A is closable.

To prove these, we will need the following simple lemma:

Lemma 4.116. Let $U : \mathcal{H} \oplus \mathcal{K} \to \mathcal{H} \oplus \mathcal{K}$ be defined by

$$U(x, y) := (-y, x).$$

Then U is a unitary, and

$$(U \operatorname{graph}(A))^\perp = \operatorname{graph}(A^*) \oplus \big(\operatorname{dom} A^* \oplus (\operatorname{dom} A)^\perp\big) \supseteq \operatorname{graph}(A^*).$$

Proof. For $(y,z) \in \mathcal{K} \oplus \mathcal{H}$, we have

$$\begin{aligned}
&(y,z) \perp U\,\mathrm{graph}(A) \\
&\quad\Longleftrightarrow \forall\, x \in \mathrm{dom}\, A: \quad 0 = \langle (y,z), (-Ax, x)\rangle = -\langle y, Ax\rangle + \langle z, x\rangle \\
&\quad\Longleftrightarrow \forall\, x \in \mathrm{dom}\, A: \quad \langle y, Ax\rangle = \langle z, x\rangle \\
&\quad\Longleftrightarrow y \in \mathrm{dom}\, A^*, \quad A^*y - z \perp \overline{\mathrm{dom}\, A} \\
&\quad\Longleftrightarrow y \in \mathrm{dom}\, A^*, \quad \exists\, w \in (\mathrm{dom}\, A)^\perp: \quad z = A^*y + w
\end{aligned}$$

□

Proposition 4.117. Let $A: \mathrm{dom}(A) (\subseteq \mathcal{H}) \to \mathcal{K}$ be a densely defined linear operator between Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$. The following hold:

(i) A^* is closed.

(ii) A is closable if and only if A^* is densely defined, and in this case

$$\overline{A} = A^{**}, \qquad \overline{A}^* = A^*.$$

Proof. (i) By assumption, $(\mathrm{dom}\, A)^\perp = \{0\}$, and hence, by Lemma 4.116,

$$\mathrm{graph}(A^*) = (U\,\mathrm{graph}(A))^\perp,$$

which, as the orthocomplement of a subspace, is closed.

(ii) See [**?**, Theorem VIII.1].

□

Corollary 4.118. For a linear operator, $A: \mathrm{dom}(A) (\subseteq \mathcal{H}) \to \mathcal{K}$,

$$A \text{ is densely defined and closed} \implies A^{**} = A.$$

In particular, this holds whenever $A \in \mathcal{B}(\mathcal{H})$.

Definition 4.119. A linear operator $A: \mathrm{dom}(A) (\subseteq \mathcal{H}) \to \mathcal{H}$ is

- *normal*, if $A^*A = AA^*$;
- *self-adjoint*, if $A = A^*$.

Obviously, a self-adjoint operator is also normal.

Exercise 4.120. (i) Let $\mathcal{K} \subseteq \mathcal{H}$ be a closed subspace and $P_\mathcal{K}$ the orthogonal projection onto it. Show that $P_\mathcal{K}$ is self-adjoint, and $P_\mathcal{K}^2 = P_\mathcal{K}$.

(ii) Vice versa, show that if $P \in \mathcal{B}(\mathcal{H})$ is such that

$$P^2 = P = P^* \quad \text{then} \quad \operatorname{ran} P \quad \text{is closed, and} \quad P = P_{\operatorname{ran} P}.$$

Definition 4.121. Any operator $\mathcal{B}(\mathcal{H})$ satisfying $P^2 = P = P^*$ is called a *projection*.

Exercise 4.122. Let $V : \operatorname{dom} V (\subseteq \mathcal{H}) \to \mathcal{K}$. Show that the following are equivalent:

(i) V is an *isometry*, i.e., $\|Vx\| = \|x\|$, $x \in \operatorname{dom} V$.

(ii) V preserves the inner product, i.e., $\langle Vy, Vx \rangle = \langle y, x \rangle$, $x, y \in \operatorname{dom} V$.

(iii) $V^*V = I_{\operatorname{dom} V}$.

Proposition 4.123. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ be a measurable function. Then

$$M_f^* = M_{\overline{f}}.$$

In particular, M_f is closed, and it is self-adjoint if and only if f is real-valued μ-a.e.

Proof. □

Exercise 4.124. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space.

(i) Show that for any measurable functions $f, g : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$,

$$M_f M_g = M_{fg} = M_g M_f.$$

(ii) Conclude that

$$M_f^* M_f = M_{|f|^2} = M_f^* M_f,$$

and thus M_f is normal.

4.12 The spectral theorem

The following is easy to see:

Exercise 4.125. Let $A : \operatorname{dom} A (\subseteq \mathcal{H}) \to \mathcal{H}$ be a linear operator and $U : \mathcal{H} \to \mathcal{K}$ be a unitary. Then A is normal/self-adjoint/projection/unitary if and only if $UAU^* : U \operatorname{dom} A (\subseteq \mathcal{K}) \to \mathcal{K}$ has the same property.

Exercise 4.125 and .... yield immediately the following:

Corollary 4.126. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ be a measurable function, and $U : L^2(\mathcal{X}, \mathcal{A}, \mu) \to \mathcal{H}$ be a unitary. Then UM_fU^* is a normal operator on $\mathcal{H}$.

It is a highly non-trivial fact that the implication in Corollary 4.126 holds also in the opposite direction. More precisely, we have the following:

Theorem 4.127. (Spectral theorem, multiplication operator form)
Let $A : \operatorname{dom} A(\subseteq \mathcal{H}) \to \mathcal{H}$ be a linear operator. A is normal if and only if there exists a measure space $(\mathcal{X}, \mathcal{A}, \mu)$, a measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$, and a unitary operator $U : L^2(\mathcal{X}, \mathcal{A}, \mu) \to \mathcal{H}$, such that

$$A = UM_fU^*.$$

The proof of this theorem is beyond the scope of an introductory course; we refer the interested reader to [**?**]. We only mention that one standard way is to first prove it for bounded normal operators, and then use the Cayley transform to obtain the spectral theorem for not necessarily bounded self-adjoint operators, which is sufficient for the purposes of quantum physics (we won't need the spectral theorem for unbounded normal operators that are not self-adjoint). We will discuss the Cayley transform in detail in Section 4.15.

4.13 Unitary groups and generators

As a further important application of the functional calculus, we discuss the Schrödinger equation in quantum mechanics. The dynamics of a closed quantum system is usually specified by the *Schrödinger equation*

$$\frac{d}{dt}\psi(t) = \frac{-i}{\hbar}\hat{H}\psi(t), \tag{4.16}$$

where $\psi(t)$ is the state vector of the system at time t, and $\hat{H}$ is the *Hamilton operator* of the system, a self-adjoint (but often unbounded) operator that is considered to be the energy observable. In the most general case, the Hamilton operator may itself depend on t, but here we only consider the time-independent case. For simplicity of notation, we absorb the constant $\hbar$ in the Hamilton operator, i.e., redefine it as $\hat{H} \to \hat{H}/\hbar$. A formal solution of the above differential equation may be given by

$$\psi(t) = e^{-i\hat{H}t}\psi(0), \tag{4.17}$$

and another formal computation yields that $U(t) = e^{-i\hat{H}t}$ is a unitary operator for every $t \in \mathbb{R}$. It is, however, far from obvious how to make mathematically precise

sense of the expression in (4.17). Indeed, a naive approach may be to define it as the sum of the infinite series

$$\sum_{n=0}^{+\infty} \frac{(-it)^n}{n!} \hat{H}^n \psi(0). \tag{4.18}$$

This works fine as long as $\hat{H}$ is bounded, in which case it is easy to see that the series $U(t) := \sum_{n=0}^{+\infty} \frac{(-it)^n}{n!} \hat{H}^n$ even converges in operator norm, and $\frac{d}{dt} U(t) = -i\hat{H}U(t)$, where the implicit limit in the derivative can again be taken in the operator norm; in particular, $\psi(t) = U(t)\psi$ is a solution of the Schrödinger equation (4.16).

The case of an unbounded $\hat{H}$ is considerably more complicated. Indeed, it is not too difficult to see that in this case, $\mathcal{D}(\hat{H}) \supsetneq \mathcal{D}(\hat{H}^2) \supsetneq \ldots$, and hence $\hat{H}^n \psi(0)$ need not be defined, not to mention the convergence of the series in (4.18), even if $\psi(0) \in \mathcal{D}(\hat{H})$, and hence the Schrödinger equation (4.16) itself makes sense. One way to save the situation can be to find a dense subspace in $\cap_{n=0}^{+\infty} \mathcal{D}(\hat{H}^n)$ for which the sum in (4.18) makes sense and converges, in which case it is also not too difficult to see that it is a solution of the Schödinger equation (4.16); these are called *analytical vectors* for $\hat{H}$. However, this is a strictly smaller subspace than $\mathcal{D}(\hat{H})$, on which the Schrödinger equation is well-defined, not to mention the fact that if the solution is indeed given by unitary operators then those are everywhere defined, simply by the definition of unitarity.

Instead, we may use functional calculus to obtain the following:

Theorem 4.128. Let $H \in \mathrm{PVM}(\mathcal{H}, \mathbb{R})$ be a real-valued PVM, and $\hat{H}$ be the corresponding operator. For every $t \in \mathbb{R}$, let $\varphi_t(x) := e^{-itx}$, $x \in \mathbb{R}$.

(i) $U(t) := e^{-it\hat{H}} := \varphi_t(H)$ is a unitary for every $t \in \mathbb{R}$, and $U(0) = I$.

(ii) $U(t+s) = U(t)U(s)$ for every $t, s \in \mathbb{R}$, i.e., $(U(t))_{t\in\mathbb{R}}$ is a *one-parameter group of unitaries*.

(iii) $\lim_{t\to t_0} U(t)\psi = U(t_0)\psi$ for any $t_0 \in \mathbb{R}$, and $\psi \in \mathcal{H}$.

(iv) The following are equivalent:

a) $\psi \in \mathcal{D}(\hat{H})$.

b) For every $t \in \mathbb{R}$,

$$\exists \frac{d}{dt} U(t)\psi := \lim_{s\to t} \frac{U(s) - U(t)}{s - t} \psi = U(t)(-i\hat{H})\psi = (-i\hat{H})U(t)\psi.$$

c) The limit $\lim_{s\to 0} \frac{U(s)-I}{s}\psi$ exists.

(v) For any $\psi \in \mathcal{D}(\hat{H})$,

$$\hat{H}\psi = i \lim_{s\to 0} \frac{U(s) - I}{s}\psi.$$

Proof. Since $|\varphi_t(x)| = 1$ for all $x \in \mathbb{R}$, $U(t)$ is a unitary according to Exercise 5.66, proving (i). We have $\varphi_{t+s} = \varphi_t\varphi_s$, and thus $U(t+s) = \varphi_{t+s}(\hat{H}) = \varphi_t(\hat{H})\varphi_s(\hat{H})$ according to Proposition 5.58; equality holds because φ_t is bounded. This proves (ii). We have

$$\|U(t)\psi - U(t_0)\psi\|^2 = \left\|(\varphi_t - \varphi_{t_0})(\hat{H})\psi\right\|^2 = \int |\varphi_t - \varphi_{t_0}|^2 \, dP_\psi \xrightarrow[n\to+\infty]{} 0,$$

where the first equality is again due to Proposition 5.58, and the convergence follows from the Lebesgue dominated convergence theorem, as $\lim_{t\to t_0} \varphi_t = \varphi_{t_0}$ pointwise, and $|\varphi_t - \varphi_{t_0}| \leq 2$, which is an integrable dominating function. This proves (iii).

To prove (iv), assume first that $\psi \in \mathcal{D}(\hat{H})$. For every $x \in \mathbb{R}$, $\lim_{s\to t} \frac{\varphi_s(x)-\varphi_t(x)}{s-t} = -ixe^{-itx}$. Moreover,

$$\left|\frac{\varphi_s(x) - \varphi_t(x)}{s-t} + ix\varphi_t(x)\right| = \left|e^{-itx}\right| \left|\frac{e^{-i(s-t)x} - 1}{s-t} + ix\right| \leq 2|x|,$$

where we used that $|e^{iy} - 1| \leq |y|$, $y \in \mathbb{R}$. Thus,

$$\left\|\frac{U(s) - U(t)}{s-t}\psi + (i\hat{H})U(t)\psi\right\|^2 = \int \left|\frac{\varphi_s - \varphi_t}{s-t} + i(\mathrm{id}_{\mathbb{R}}\,\varphi_t)\right|^2 dP_\psi \xrightarrow[n\to+\infty]{} 0,$$

by the Lebesgue dominated convergence theorem, since the integrand goes to zero pointwise, and $4\,\mathrm{id}_{\mathbb{R}}^2$ is an integrable dominating function. This proves $a) \Longrightarrow b)$, and $b) \Longrightarrow c)$ is trivial.

Let us now define

$$\mathcal{D}(A) := \left\{\psi \in \mathcal{H} : \exists \lim_{s\to 0} \frac{U(s) - I}{s}\psi\right\}, \qquad A\psi := i \lim_{s\to 0} \frac{U(s) - I}{s}\psi, \qquad \psi \in \mathcal{D}(A).$$

By $a) \Longrightarrow c)$, $\hat{H} \subseteq A$. Moreover, for any $\psi_1, \psi_2 \in \mathcal{D}(A)$,

$$\begin{aligned}\langle \psi_1, A\psi_2\rangle &= \lim_{s\to 0} \frac{1}{s} \langle \psi_1, i(U(s) - I)\psi_2\rangle = \lim_{s\to 0} \frac{i}{s} \langle (U(s) - I)^*\psi_1, \psi_2\rangle \\ &= \lim_{s\to 0} \frac{i}{s} \langle (U(-s) - I)\psi_1, \psi_2\rangle = i \langle iA\psi_1, \psi_2\rangle = \langle A\psi_1, \psi_2\rangle,\end{aligned}$$

i.e., A is symmetric. Since $\hat{H}$ is self-adjoint by Exercise 5.66, and self-adjoint operators are maximally symmetric, we see that $A = \hat{H}$, from which $c) \Longrightarrow a)$ follows immediately, and we also get (v). □

Remark 4.129. Theorem 4.128 suggests to take the unitary group $U(t) = \varphi_t(\hat{H})$ as the fundamental object describing the time evolution of a quantum system instead of the Schrödinger equation, leading to the following postulate:

> If the initial state of the system at time 0 is $\psi(0)$ then the state of the system at time t is $\psi(t) = U(t)\psi(0)$, where $U(t) := \varphi_t(\hat{H})$, and the Schrödinger equation (4.16) holds for every $t \in \mathbb{R}$ whenever $\psi(0) \in \mathcal{D}(\hat{H})$.

Definition 4.130. A collection of unitary operators $(U(t))_{t\in\mathbb{R}} \subseteq \mathcal{B}(\mathcal{H})$ is called a *one-parameter group of unitaries*, or a *unitary group* for short, if $U(t+s) = U(t)U(s)$ for every $t, s \in \mathbb{R}$. This means that $t \mapsto U(t)$ is a unitary representation of the commutative group $(\mathbb{R}, +)$ on $\mathcal{H}$.

A unitary group is called *strongly continuous*, if $\lim_{t\to 0} U(t)\psi = \psi$ for every $\psi \in \mathcal{H}$, and *norm continuous* if $\lim_{t\to 0} \|U(t) - I\| = 0$.

Theorem 4.128 can be reversed in the following precise sense:

Theorem 4.131. (Stone - von Neumann)
Let $(U(t))_{t\in\mathbb{R}}$ be a strongly continuous unitary group, and define

$$\mathcal{D}(\hat{H}) := \left\{ \psi \in \mathcal{H} : \exists \lim_{t\to 0} \frac{1}{t}(U(t)\psi - \psi) \right\},$$

$$\hat{H}\psi := \lim_{t\to 0} \frac{-i}{t}(U(t)\psi - \psi), \qquad \psi \in \mathcal{D}(\hat{H}).$$

Then $\mathcal{D}(\hat{H})$ is a dense subspace, and $\hat{H}$ is a self-adjoint operator on it. Moreover, $U(t) = e^{it\hat{H}}$, $t \in \mathbb{R}$, where the latter is defined by functional calculus using the spectral PVM H of $\hat{H}$, as in Theorem 4.128.

We omit the rather non-trivial proof, and refer to Theorem VIII.8 in [**?**] instead.

Theorems 4.128 and 4.131 give a one-to-one correspondence between self-adjoint operators (equivalently, real-valued PVMs) and strongly continuous unitary groups on a given Hilbert space. As the following two exercises show, bounded self-adjoint operators correspond exactly to norm continuous unitary groups under this correspondence.

Exercise 4.132. Let $\hat{H}$ be an unbounded self-adjoint operator, with corresponding spectral PVM H, and $(U(t))_{t\in\mathbb{R}}$ be the generated unitary group, as defined in Theorems 4.128. Show that there exists a sequence $(t_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ such that $\lim_{n\to+\infty} t_n = 0$, and a sequence $(\psi_n)_{n\in\mathbb{N}}$ of unit vectors in $\mathcal{H}$, such that

$$\lim_{n\to+\infty} \|U(t_n)\psi_n - \psi_n\| = 2. \tag{4.19}$$

Conclude that the unitary group is not norm continuous.

Solution: Note that for any $t \in \mathbb{R}$ and $\psi \in \mathcal{H}$,

$$\begin{aligned}\|U(t)\psi - \psi\|^2 &= \|U(t)\psi\|^2 + \|\psi\|^2 - \langle U(t)\psi, \psi\rangle - \langle \psi, U(t)\psi\rangle \\ &= 2\|\psi\|^2 - \int (e^{itx} + e^{-itx})\, dP_\psi(x) = 2\|\psi\|^2 - 2\int \cos(tx)\, dP_\psi(x).\end{aligned}$$

Clearly,

$$\left|\int \cos(tx)\, dP_\psi(x)\right| \le \int \underbrace{|\cos(tx)|}_{\le 1}\, dP_\psi(x) \le 1 \cdot P_\psi(\mathbb{R}) = \|\psi\|^2,$$

by which $\|U(t)\psi - \psi\| \le 2$ for any unit vector ψ.

By assumption, there exists a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \operatorname{supp} H$ such that $\lim_{n\to+\infty} |x_n| = +\infty$. For every $n \in \mathbb{N}$, let $t_n := \pi/x_n$. By assumption, $P_n := P((x_n - 1, x_n + 1)) \neq 0$, and thus we can choose a unit vector $\psi_n \in \operatorname{ran} P_n$ for every $n \in \mathbb{N}$. Then $\operatorname{supp} P_{\psi_n} \subseteq (x_n - 1, x_n + 1)$, and thus

$$\begin{aligned}\int \cos(t_n x)\, dP_{\psi_n}(x) &= \int_{(x_n-1, x_n+1)} \cos(t_n x)\, dP_{\psi_n}(x) \\ &\le \cos(\pi - 1/x_n) \underbrace{P_{\psi_n}((x_n - 1, x_n + 1))}_{=\|P_n\psi_n\|^2=1} \xrightarrow[n\to+\infty]{} -1,\end{aligned}$$

where the inequality holds for all large enough n. Thus, $\liminf_{n\to+\infty} \|U(t_n)\psi_n - \psi_n\|^2 \ge 4\|\psi_n\|^2 = 4$, completing the proof.

By (4.19), $\liminf_{n\to+\infty} \|U(t_n) - I\| \ge 2$, and hence U is not norm continuous.

Exercise 4.133. Let $\hat{H} \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ be a bounded self-adjoint operator on $\mathcal{H}$.

(i) Show that for every $t \in \mathbb{R}$, the series

$$U(t) := \sum_{n=0}^{+\infty} \frac{1}{n!}(-it\hat{H})^n$$

is absolute convergent, and defines a unitary operator.

(ii) Let H be the spectral PVM of $\hat{H}$. Show that $U(t) = \varphi_t(H)$, $t \in \mathbb{R}$, with the notations of Theorem 4.128.

(Hint: Use Corollary 5.60.)

(iii) Show that for any $t \in \mathbb{R}$,

$$\lim_{s\to t} \frac{U(s) - U(t)}{s - t} = -i\hat{H}U(t) = -iU(t)\hat{H},$$

where the limit is in the operator norm. That is, $t \mapsto U(t)$ is *norm differentiable* at any $t \in \mathbb{R}$, and its derivative is $(-i\hat{H})U(t)$.

(iv) Conclude that

$$\hat{H} = \lim_{t\to 0} \frac{-i}{t}(U(t) - I),$$

where the limit is again in the operator norm.

4.14 Symmetric operators

Lemma 4.134. For a linear operator $A : \mathrm{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{H}$,

$$\langle y, Ax\rangle = \langle Ay, x\rangle, \quad x, y \in \mathrm{dom}(A) \quad \Longleftrightarrow \quad A \subseteq A^*. \tag{4.20}$$

Proof. Trivial. □

Definition 4.135. An operator satisfying either (and hence both) properties in (4.20) is called *symmetric*.

Corollary 4.136. (i) The adjoint of a densely defined symmeric operator is densely defined.

(ii) A densely defined symmetric operator is closable, and its closure is again symmetric.

Proof. (i) and the first assertion in (ii) are immediate from the defintion and Proposition 4.117. To see the second assertion in (ii), let $x, y \in \mathrm{dom}(\overline{A})$, so that there exist sequences $(x_n)_{n\in\mathbb{N}} \subseteq \mathrm{dom}(A)$, $(y_n)_{n\in\mathbb{N}} \subseteq \mathrm{dom}(A)$, such that

$$\lim_n x_n = x, \qquad \lim_n y_n = y, \qquad \lim_n Ax_n = \overline{A}x, \qquad \lim_n Ay_n = \overline{A}y.$$

Then

$$\left\langle y, \overline{A}x\right\rangle = \lim_n \langle y_n, Ax_n\rangle = \lim_n \langle Ay_n, x_n\rangle = \left\langle \overline{A}, x\right\rangle,$$

showing that $\overline{A}$ is indeed symmetric. □

It is clear from the definition that self-adjoint operators are symmetric, but not the other way around.

Lemma 4.137. Let $A : \mathrm{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{H}$ be a symmetric operator. For any $\lambda \in \mathbb{C}$ with $\mathfrak{Im}\,\lambda \neq 0$, $A + \lambda I$ is injective, and its inverse is bounded. Moreover,

$$\big(\ker(A^* + \overline{\lambda}I)\big)^\perp = \overline{\mathrm{ran}(A + \lambda I)} = \mathrm{ran}(\overline{A} + \lambda I). \tag{4.21}$$

Proof. We have

$$\|(A+\lambda I)x\|^2 = \|Ax + (\mathfrak{Re}\,\lambda)x\|^2 + |\,\mathfrak{Im}\,\lambda|^2\,\|x\|^2 \geq |\,\mathfrak{Im}\,\lambda|^2\,\|x\|^2\,,$$

from which the injectivity of $A+\lambda I$ is immediate, and $\|(A+\lambda)^{-1}\| \leq |\,\mathfrak{Im}\,\lambda|^{-1}$, showing the boundedness of $(A+\lambda)^{-1}$. According to Proposition 4.58, boundedness of $(A+\lambda I)^{-1}$ implies that it has a unique extension onto the closure of its domain, and it is given by

$$\overline{(A+\lambda I)^{-1}} = \big(\overline{A+\lambda I}\big)^{-1} = (\overline{A}+\lambda I)^{-1}, \tag{4.22}$$

where the first equality is due to Exercise 4.57, and the second equality follows from Exercise **??**. Finally,

$$\begin{aligned}\overline{\mathrm{ran}(A+\lambda I)} = \overline{\mathrm{dom}(A+\lambda I)^{-1}} &= \mathrm{dom}(\overline{(A+\lambda I)^{-1}}) \\ &= \mathrm{dom}\,\big((\overline{A}+\lambda I)^{-1}\big) = \mathrm{ran}(\overline{A}+\lambda I),\end{aligned}$$

where the first equality is by definition, the second equality is due to Proposition 4.58, the third equality is due to (4.22), and the last equality is again by definition. □

Corollary 4.138. Every eigenvalue of a symmetric operator is real.

Proof. According to Lemma 4.137, $A-\lambda I$ is injective when $\mathfrak{Im}\,\lambda \neq 0$, from which the assertion follows immediately. □

Proposition 4.139. For a symmetric operator $A:\ \mathrm{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{H}$, the following are equivalent:

(i) A is essentially self-adjoint.

(ii) For any $\lambda \in \mathbb{C}\setminus\mathbb{R}$, $\ker(A^*+\lambda I) = \{0\}$.

(iii) For some $\lambda \in \mathbb{C}\setminus\mathbb{R}$, $\ker(A^*+\lambda I) = \{0\} = \ker(A^*+\overline{\lambda} I)$.

(iv) For any $\lambda \in \mathbb{C}\setminus\mathbb{R}$, $\overline{\mathrm{ran}(A+\lambda I)} = \mathcal{H}$.

(v) For some $\lambda \in \mathbb{C}\setminus\mathbb{R}$, $\overline{\mathrm{ran}(A+\lambda I)} = \mathcal{H} = \overline{\mathrm{ran}(A+\overline{\lambda} I)}$.

Proof. (i)$\Longrightarrow$(ii): By assumption, $(A^*)^* = \overline{A} = (\overline{A})^* = A^*$, where the first and the last equality hold for any closable operator. In particular, A^* is symmetric, and (ii) follows immediately from Lemma 4.137.

(ii)$\Longleftrightarrow$(iv) and (iii)$\Longleftrightarrow$(v) are immediate from Lemma **??**, and (ii)$\Longrightarrow$(iii) and (iv)$\Longrightarrow$(v) are obvious.

(v)$\Longrightarrow$(i): Symmetry means that $A \subseteq A^*$ and thus $\overline{A} \subseteq A^* = (\overline{A})^*$. Hence, we only have to prove that $A^* \subseteq \overline{A}$. By assumption, $\mathcal{H} = \overline{\mathrm{ran}(A+\lambda I)} = \mathrm{ran}(\overline{A}+\lambda I)$, where the second equality is due to (4.21). Hence, for any $y \in \mathrm{dom}(A^*)$ there exists an $x \in \mathrm{dom}(A)$ such that $(A^*+\lambda I)y = (\overline{A}+\lambda I)x = (A^*+\lambda I)x$, where the last equality is due to $\overline{A} \subseteq A^*$. Hence, $y-x \in \ker(A^*+\lambda I) = \mathrm{ran}(A+\overline{\lambda} I)^{\perp} = \{0\}$, by which $y = x \in \mathrm{dom}(A)$. □

4.15 The Cayley transform

An essentially self-adjoint operator has exactly one self-adoint extension. In general, however, a symmetric operators may have multiple self-adjoint extensions, or no self-adjoint extension at all. A useful tool in studying self-adoint extensions is the Cayley transform that establishes an order-preserving bijection between symmetric operators and isometries with no fixed point, under which self-adjoint operators correspond to unitaries. Using the Cayley transform is also the easiest way to obtain the spectral theorem of unbounded self-adjoint operators from that of bounded normal operators.

Lemma 4.140. Let $A : \mathrm{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{H}$ be a symmetric operator.

(i) $V(A)_+$ and $V(A)_-$, defined as

$$V(A)_\pm : \mathrm{graph}(A) \to \mathrm{ran}(A \pm iI), \qquad V(A)_\pm(x, Ax) := Ax \pm ix, \quad x \in \mathrm{dom}\, A,$$

are isometric bijections.

(ii) A symmetric operator B is an extension of A if and only if $V(B)_+$ (resp. $V(B)_-$) is an extension of $V(A)_+$ (resp. $V(A)_-$).

(iii) We have

$$V(\overline{A})_\pm = \overline{V(A)_\pm}, \quad \text{and} \quad \mathrm{ran}(\overline{A} \pm iI) = \overline{\mathrm{ran}(A \pm iI)}.$$

Proof. (i) Immediate from

$$\|Ax \pm ix\|^2 = \|Ax\|^2 + \|x\|^2 = \|(x, Ax)\|^2 .$$

(ii) Trivial.

(iii) Clearly, $\overline{V(A)_+}$ is an isometric bijection from $\overline{\mathrm{graph}(A)} = \mathrm{graph}(\overline{A})$ onto $\overline{\mathrm{ran}(A + iI)}$ (see Exercise 4.60). If $x \in \mathrm{dom}(\overline{A})$ then there exists a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathrm{dom}(A)$ such that $\lim_n (x_n, Ax_n) = (x, \overline{A}x)$. Hence,

$$\begin{aligned}\overline{V(A)_+}(x, \overline{A}x) &= \lim_n V(A)_+(x_n, Ax_n) = \lim_n (Ax_n + ix_n) = \overline{A}x + ix \\ &= V(\overline{A})_+(x, \overline{A}x).\end{aligned}$$

Thus, $\overline{V(A)_+}$ coincides with $V(\overline{A})_+$ on $\mathrm{graph}(A)$, which is dense in $\overline{\mathrm{graph}(A)} = \mathrm{graph}(\overline{A})$, and hence $\overline{V(A)_+} = V(\overline{A})_+$. In particular, $\overline{\mathrm{ran}(A + iI)} = \mathrm{ran}\, \overline{V(A)_+} = \mathrm{ran}(V(\overline{A})_+) = \mathrm{ran}(\overline{A} + iI)$. The proof for $V(A)_-$ goes the same way. □

Definition 4.141. Let $A : \mathrm{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{H}$ be a symmetric operator. The operator

$$\kappa(A) := V(A)_-^{-1} V(A)_+ = (A - iI)(A + iI)^{-1}$$

is called the *Cayley transform* of A.

Proposition 4.142. (i) The Cayley transform is a bijection between

$$\mathcal{S}_{\mathcal{H}} := \{A : \mathrm{dom}(A)(\subseteq \mathcal{H}) \to \mathcal{H} \ \text{ symmetric}\} \ \text{ and}$$
$$\mathcal{I}_{\mathcal{H}} := \{W : \mathrm{dom}\, W(\subseteq \mathcal{H}) \to \mathcal{H} \ \text{ isometry}, \ \ker(I - W) = \{0\}\},$$

with inverse

$$\kappa^{-1}(W) = i(I + W)(I - W)^{-1}, \qquad W \in \mathcal{I}_{\mathcal{H}}.$$

(ii) For $A, B \in \mathcal{S}_{\mathcal{H}}$,

$$A \subseteq B \iff \kappa(A) \subseteq \kappa(B).$$

(iii) For any $A \in \mathcal{S}_{\mathcal{H}}$,

$$\kappa(\overline{A}) = \overline{\kappa(A)}.$$

Proof. (i)

□

4.16 Analytic vectors

Definition 4.143. For a linear operator $A : \mathrm{dom}\, A(\subseteq \mathcal{H}) \to \mathcal{H}$, let

$$\mathrm{dom}^\infty A := \cap_{n \in \mathbb{N}} \mathrm{dom}\, A^n,$$

and for every $\psi \in \mathrm{dom}^\infty A$, let

$$R_A(\psi) := \left(\limsup_{n \to +\infty} \sqrt[n]{\frac{\|A^n \psi\|}{n!}} \right)^{-1},$$

with the convention $1/0 := +\infty$, $1/+\infty := 0$. We say that ψ is an *analytic vector* for A, if $R_A(\psi) > 0$, and denote the set of analytic vectors for A by $C^\infty(A)$.

Remark 4.144. For every $\psi \in C^\infty(A)$, $R_A(\psi)$ is the supremum of all numbers $R > 0$ such that

$$\sum_{n=0}^{+\infty} \frac{(zA)^n}{n!} \psi \quad \text{is absolutely and unformly convergent on} \quad |z| \leq R.$$

Lemma 4.145. For every $\psi, \psi_1, \psi_2 \in \operatorname{dom}^\infty(A)$ and $\lambda \in \mathbb{C} \setminus \{0\}$,

(i) $R_A(\lambda\psi) = R_A(\psi)$,

(ii) $R_A(\psi_1 + \psi_2) \geq \left(R_A(\psi_1)^{-1} + R_A(\psi_2)^{-1}\right)^{-1} \geq \min\{R_A(\psi_1), R_A(\psi_2)\}$,

(iii) $R_A(A\psi) = R_A(\psi)$.

Proof. The first property is obvious. The second one follows immediately from

$$\sqrt[n]{\frac{\|A^n(\psi_1+\psi_2)\|}{n!}} \leq \sqrt[n]{\frac{\|A^n\psi_1\|}{n!} + \frac{\|A^n\psi_2\|}{n!}} \leq \sqrt[n]{\frac{\|A^n\psi_1\|}{n!}} + \sqrt[n]{\frac{\|A^n\psi_2\|}{n!}}.$$

The third one follows as

$$\begin{aligned} R_A(A\psi)^{-1} &= \limsup_{n\to+\infty} \sqrt[n]{\frac{\|A^{n+1}\psi\|}{n!}} = \limsup_{n\to+\infty} \sqrt[n]{\frac{\|A^{n+1}\psi\|}{(n+1)!}(n+1)} \\ &= \lim_{n\to+\infty} \sqrt[n]{n+1} + \exp\left(\limsup_{n\to+\infty} \frac{n+1}{n} \log \sqrt[n+1]{\frac{\|A^{n+1}\psi\|}{n!}}\right) \\ &= R_A(\psi)^{-1}. \end{aligned}$$

□

Definition 4.146. For a linear operator $A : \operatorname{dom} A(\subseteq \mathcal{H}) \to \mathcal{H}$ and any $\psi \in \cap_{n\in\mathbb{N}} \operatorname{dom} A^n$, let

$$\Omega_A(\psi) := \{\psi, A\psi, A^2\psi, \ldots\}$$

be the *orbit* of ψ under the action of A.

Lemma 4.145 yields immediately the following:

Lemma 4.147. Let $A : \operatorname{dom} A(\subseteq \mathcal{H}) \to \mathcal{H}$ be a linear operator.

(i) $C^\infty(A)$ is a subspace.

(ii) For any $\psi \in C^\infty(A)$,

$$R_A(\phi) \geq R_A(\psi), \quad \phi \in \operatorname{span} \Omega_A(\psi).$$

In particular, $\operatorname{span} \Omega_A(\psi) \subseteq C^\infty(A)$.

Lemma 4.148. Let $A : \operatorname{dom} A(\subseteq \mathcal{H}) \to \mathcal{H}$ be a symmetric operator. If $\psi \in C^\infty(A)$ then $A\big|_{\operatorname{span}\Omega_A(\psi)}$ is essentially self-adjoint as an operator on $\mathcal{H}_\psi := \overline{\operatorname{span}\Omega_A(\psi)}$.

Proof. Let $A_\psi := A\big|_{\operatorname{span}\Omega_A(\psi)}$. Note that for any $(a_n)_{n\in\mathbb{N}} \subseteq \mathbb{C}$ such that only finitely many $a_n \neq 0$,

$$\begin{aligned}\left\|\sum_n \overline{a}_n A^n\psi\right\|^2 &= \sum_{n\in\mathbb{N}} |a_n|^2 \left\|A^n\psi\right\|^2 + \sum_{n\neq m} a_n\overline{a}_m \underbrace{\langle A^n\psi, A^m\psi\rangle}_{=\langle A^m\psi, A^n\psi\rangle} \\ &= \overline{\sum_{n\in\mathbb{N}} |a_n|^2 \left\|A^n\psi\right\|^2 + \sum_{n\neq m} \overline{a}_n a_m \langle A^n\psi, A^m\psi\rangle} \\ &= \overline{\left\|\sum_n a_n A^n\psi\right\|^2} = \left\|\sum_n a_n A^n\psi\right\|^2,\end{aligned}$$

where $\langle A^n\psi, A^m\psi\rangle = \langle A^m\psi, A^n\psi\rangle$ due to the symmetry of A. This shows that

$$C : \operatorname{span}\Omega_A(\psi) \to \operatorname{span}\Omega_A(\psi), \qquad C\left(\sum_n a_n A^n\psi\right) := \sum_n \overline{a}_n A^n\psi$$

is a well-defined conjugate linear isometry, and it is easy to see that $A_\psi C = CA_\psi$. Hence, by Lemma **??**, A_ψ has a self-adjoint extension $\hat{A}_\psi$. The proof will be complete if we show that this is the only self-adjoint extension of A_ψ, for which it is sufficient to show that the unitary group $(e^{it\hat{A}_\psi})_{t\in\mathbb{R}}$ is uniquely determined by A_ψ.

Let $\phi \in \operatorname{span}\Omega_A(\psi)$. If $|t| < R_A(\psi)$ then

$$\begin{aligned}\int_{\mathbb{R}} e^{|t||x|}\, dP_\phi^{\hat{A}_\psi}(x) &= \sum_{n=0}^{+\infty} \frac{|t|^n}{n!} \int_{\mathbb{R}} |x|^n\, dP_\phi^{\hat{A}_\psi}(x) \\ &\leq \sum_{n=0}^{+\infty} \frac{|t|^n}{n!} \underbrace{\left(\int_{\mathbb{R}} |x|^{2n}\, dP_\phi^{\hat{A}_\psi}(x)\right)^{1/2}}_{=\left\|\hat{A}_\psi^n\phi\right\| = \|A^n\phi\|} \underbrace{\left(\int_{\mathbb{R}} 1\, dP_\phi^{\hat{A}_\psi}(x)\right)^{1/2}}_{=\|\phi\|} \\ &\leq \|\phi\| \sum_{n=0}^{+\infty} \frac{\|A^n\phi\|}{n!} |t|^n < +\infty, \end{aligned} \tag{4.23}$$

where the first equality is due to the monotone convergence theorem, the first inequality is due to the Cauchy-Schwarz inequality, and the last inequality is due to Lemma 4.147. Thus, for $|t| < R_A(\psi)$,

$$\begin{aligned}\left\langle \phi, e^{it\hat{A}_\psi}\phi\right\rangle &= \int_{\mathbb{R}} e^{itx}\, dP_\phi^{\hat{A}_\psi}(x) = \int_{\mathbb{R}} \sum_{n\in\mathbb{N}} \frac{(itx)^n}{n!}\, dP_\phi^{\hat{A}_\psi}(x) \\ &= \sum_{n\in\mathbb{N}} \frac{(it)^n}{n!} \underbrace{\int_{\mathbb{R}} x^n\, dP_\phi^{\hat{A}_\psi}(x)}_{=\left\langle \phi, \hat{A}_\psi^n\phi\right\rangle = \langle \phi, A^n\phi\rangle} = \sum_{n\in\mathbb{N}} \frac{(it)^n}{n!} \langle \phi, A^n\phi\rangle,\end{aligned}$$

where the third equality follows by the Lebesgue dominated convergence theorem due to (4.23). Now, if $\tilde{A}_\psi$ is any other self-adjoint extension of A_ψ then by applying the exact same argument to it, we get that

$$\left\langle \phi, e^{it\hat{A}_\psi}\phi \right\rangle = \sum_{n\in\mathbb{N}} \frac{(it)^n}{n!} \langle \phi, A^n\phi\rangle = \left\langle \phi, e^{it\tilde{A}_\psi}\phi \right\rangle, \qquad |t| < R_A(\psi), \quad \phi \in \operatorname{span}\Omega_A(\psi).$$

Since $e^{it\hat{A}_\psi}$ and $e^{it\tilde{A}_\psi}$ are bounded, the same holds for every $\phi \in \mathcal{H}_\psi$, and thus $e^{it\hat{A}_\psi} = e^{it\tilde{A}_\psi}$ for $|t| < R_A(\psi)$. Therefore, by Theorem 4.128, $\hat{A}_\psi = \tilde{A}_\psi$. □

Theorem 4.149. (Nelson)
If $A : \operatorname{dom} A(\subseteq \mathcal{H}) \to \mathcal{H}$ is a symmetric operator such that $C^\infty(A)$ is dense then A is essentially self-adjoint.

Proof. By Proposition 4.139, it is sufficient to prove that $\operatorname{ran}(A \pm iI)$ are dense. Let $\eta \in \mathcal{H}$ be arbitrary. By assumption, for every $\varepsilon > 0$ there exists an $\psi_\varepsilon \in C^\infty(A)$ such that $\|\eta - \psi_\varepsilon\| < \varepsilon$. By Lemma 4.148, A_{ψ_ε} is essentially self-adjoint, and hence, again by Proposition 4.139, $\operatorname{ran}(A_{\psi_\varepsilon} \pm iI)$ is dense in $\mathcal{H}_{\psi_\varepsilon}$. Thus, there exists a $\phi_\varepsilon \in \operatorname{dom}(A_{\psi_\varepsilon}) = \operatorname{span}\Omega_A(\psi_\varepsilon) \subseteq \operatorname{dom}(A)$ such that $\varepsilon > \|\psi_\varepsilon - (A_{\psi_\varepsilon} + iI)\phi_\varepsilon\| = \|\psi_\varepsilon - (A + iI)\phi_\varepsilon\|$, and hence $\|\eta - (A + iI)\phi_\varepsilon\| < 2\varepsilon$. This shows the density of $\operatorname{ran}(A + iI)$, and the density of $\operatorname{ran}(A - iI)$ follows the same way. □

4.17 The adjoint of bounded operators

Proposition 4.150. Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces, and let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be a bounded linear operator. There exists a unique operator $A^* \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ such that

$$\langle A^*y, x\rangle = \langle y, Ax\rangle, \qquad x \in \mathcal{H},\ y \in \mathcal{K}.$$

Proof. For every $y \in \mathcal{K}$, the map

$$x \mapsto \langle y, Ax\rangle$$

is a linear functional on $\mathcal{H}$. By the Riesz representation theorem (see Proposition 4.100), there exists a unique vector, which we denote by A^*y, such that

$$\langle y, Ax\rangle = \langle A^*y, x\rangle, \qquad x \in \mathcal{H}.$$

It is easy to see that the uniqueness implies that the map $y \mapsto A^*y$ is a linear map from $\mathcal{K}$ to $\mathcal{H}$. Boundedness is automatic when either of the Hilbert spaces is finite-dimensional, and it follows from Exercise 4.152 below in the general case. □

Definition 4.151. For $A \in \operatorname{Lin}(\mathcal{H}, \mathcal{K})$, the map $A^* \in \operatorname{Lin}(\mathcal{K}, \mathcal{H})$ is called the *adjoint* of A.

Exercise 4.152. Show that for any $A \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$,

$$\|A\| = \|A^*\|, \qquad \text{and} \qquad \|A^*A\| = \|A\|^2.$$

(Hint: Use Exercise 4.76 and Exercise 4.22.)

Score: 3+3=6 points.
Solution: We only prove the last assertion. It is obvious that $\|A^*A\| \leq \|A^*\| \, \|A\| = \|A\|^2$. For the converse, note that for any $x \in \mathcal{H}$,

$$\|Ax\|^2 = \langle Ax, Ax \rangle = \langle x, A^*Ax \rangle \leq \|x\| \, \|A^*Ax\| \leq \|x\|^2 \, \|A^*A\|.$$

Taking the supremum over all x in the unit sphere of $\mathcal{H}$, we get $\|A\|^2 \leq \|A^*A\|$.

Exercise 4.153. (i) Show that $A \mapsto A^*$ is a conjugate linear map from $\mathcal{B}(\mathcal{H}, \mathcal{K})$ to $\mathcal{B}(\mathcal{K}, \mathcal{H})$, i.e.,

$$(A+B)^* = A^* + B^*, \qquad (\lambda A)^* = \bar{\lambda} A^*, \qquad A, B \in \mathrm{Lin}(\mathcal{H}, \mathcal{K}), \quad \lambda \in \mathbb{C}.$$

(ii) Show that the adjoint reverses the product, i.e., for any $B \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ and $A \in \mathcal{B}(\mathcal{K}, \mathcal{L})$,

$$(AB)^* = B^*A^*.$$

(iii) Show that taking the adjoint is an involution, i.e., for any $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$,

$$(A^*)^* = A.$$

Score: 2+2+2=6 points.

Exercise 4.154. Show that if $A_{ij} := \langle e_i, Ae_j \rangle$ are the matrix elements of an $A \in \mathrm{Lin}(\mathcal{H})$ in an ONB $\{e_i\}_{i=1}^d$ then the matrix elements of A^* in the same basis are given by $(A^*)_{ij} = \bar{A}_{ij}$. That is, the matrix of the adjoint is the entry-wise conjugate of the transpose of the matrix of A.

Definition 4.155. We define the adjoint A^* of a square matrix $A \in \mathbb{C}^{d \times d}$ as $(A^*)_{ij} := \bar{A}_{ji}$, i.e., it is the entry-wise conjugate of the transpose of A.

The following exercise shows that it is justified to denote the adjoint of a matrix by the same notation as the adjoint of an operator.

Exercise 4.156. Let $e_i = \mathbf{1}_{\{i\}}$, $i = 1, \ldots, d$ be the canonical basis of $\mathbb{C}^d$. Show that every matrix $A \in \mathbb{C}^{d\times d}$ defines a linear operator on $\mathcal{H} := \mathbb{C}^d$ (also denoted by A), given by

$$Ax := \sum_{i=1}^{d} \left(\sum_{j=1}^{d} A_{ij} x_i \right) e_i.$$

Show that the adjoint of this operator is the operator corresponding (in the above way) to the adjoint of the matrix A.

Exercise 4.157. Let $\mathcal{H}, \mathcal{K}, \mathcal{L}_1, \mathcal{L}_2$ be Hilbert spaces.

(i) Show that for any $x \in \mathcal{H}$,

$$\langle x|^* = |x\rangle\,, \qquad |x\rangle^* = \langle x|\,,$$

where the inner product on the scalar field is the canonical $\langle \lambda, \eta\rangle := \bar{\lambda}\eta$, $\lambda, \eta \in \mathbb{K}$.

(ii) Show that for any $x \in \mathcal{H}$, $y \in \mathcal{K}$,

$$|y\rangle\langle x|^* = |x\rangle\langle y|\,.$$

(iii) Show that for any $x \in \mathcal{H}$, $y \in \mathcal{K}$, and any $A \in \mathcal{B}(\mathcal{L}_1, \mathcal{H})$, $B \in \mathcal{B}(\mathcal{K}, \mathcal{L}_2)$,

$$B\,|y\rangle\langle x|\,A = |By\rangle\langle A^* x|\,.$$

Score: 3+2+2=7 points.

Exercise 4.158. Let $A \in \mathcal{B}(\mathcal{H})$ and $\mathcal{K}$ be a subspace of $\mathcal{H}$. Show that if $\mathcal{K}$ is invariant under A, i.e., $A\mathcal{K} \subseteq \mathcal{K}$, then $\mathcal{K}^\perp$ is invariant under A^*.

Exercise 4.159. Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces and $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$.

(i) Show that

$$\ker A = \ker(A^* A). \tag{4.24}$$

(Hint: Consider $\|Ax\|^2$.)

(ii) Show that for any $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$,

$$(\operatorname{ran} A)^\perp := \{y \in \mathcal{K} : \langle y, z\rangle = 0,\ \forall z \in \operatorname{ran} A\} = \ker A^*. \tag{4.25}$$

Conclude that $(\operatorname{ran} A^*)^\perp = \ker A$.

Score: 3+3=6 points.

Solution:

(i) We have $\|Ax\|^* = \langle Ax, Ax\rangle = \langle x, A^*Ax\rangle$, and hence if $x \in \ker(A^*A)$ then $x \in \ker A$. In the opposite direction, if $x \in \ker A$ then $A^*Ax = A^*0 = 0$, and hence $x \in \ker(A^*A)$.

(ii) We have $y \in (\operatorname{ran} A)^\perp \iff 0 = \langle y, Ax\rangle = \langle A^*y, x\rangle$ for all $x \in \mathcal{H}$, which is equivalent to $A^*y = 0$, i.e., $y \in \ker A^*$.

Remark 4.160. The relation in (4.25) may also be expressed as

$$\mathcal{H} = \overline{\operatorname{ran} A^*} \oplus \ker A, \qquad \mathcal{K} = \overline{\operatorname{ran} A} \oplus \ker A^*. \tag{4.26}$$

Definition 4.161. We say that a bounded linear map $A \in \mathcal{B}(\mathcal{H})$ is

- *normal*, if $A^*A = AA^*$;
- *self-adjoint*, if $A^* = A$.

Obviously, every self-adjoint operator is normal.

Recall that the adjoint of a matrix $A \in \mathbb{K}^{[m]\times[n]}$ is defined as $(A^*)_{ij} := \overline{A_{ji}}$. A matrix $A \in \mathbb{K}^{[n]\times[n]}$ is called self-adjoint if $A_{ji} = \overline{A_{ij}}$, and it is called normal if $AA^* = A^*A$. When $\mathbb{K} = \mathbb{R}$, the adjoint of a matrix coincides with its transpose, and a self-adjoint matrix is also called *symmetric*.

It is easy to see that the self-adjointness/normality of an operator and of a matrix are in the expected relation:

Exercise 4.162. Let $A \in \mathbb{K}^{[d]\times[d]}$ be a matrix. Show that it is normal/self-adjoint if and only if it is normal/self-adjoint as a linear operator on $\mathbb{K}^d$.

Exercise 4.163. Show that for an $A \in \mathcal{B}(\mathcal{H})$, the following are equivalent:

(i) A is normal/self-adjoint.

(ii) The matrix of A in any ONB is normal/self-adjoint.

(iii) The matrix of A in some ONB is normal/self-adjoint.

Exercise 4.164. Show an example of an operator $A \in \mathcal{B}(\mathcal{H})$ that is a) not normal; b) normal but not self-adjoint; c) self-adjoint.

(Hint: Consider operators of the form $|y\rangle\langle x|$.)

Score: 6 points.

Exercise 4.165. (i) Show that for an operator $A \in \mathcal{B}(\mathcal{H})$,

$$A \text{ is normal} \quad \Longleftrightarrow \quad \|Ax\| = \|A^*x\|, \quad x \in \mathcal{H}.$$

(ii) Conclude that if A is normal then

$$\ker A = \ker A^*A = \ker AA^* = \ker A^*.$$

(iii) Show that if $A \in \mathcal{B}(\mathcal{H})$ is normal then

$$(\operatorname{ran} A)^\perp = \ker A, \quad \text{and} \quad \operatorname{ran} A = \operatorname{ran} A^*. \tag{4.27}$$

Score: 2+2+4=8 points.

Remark 4.166. Combining (4.27) with (4.26) yields that for a normal operator $A \in \mathcal{B}(\mathcal{H})$,

$$\mathcal{H} = \overline{\operatorname{ran} A} \oplus \ker A = \overline{\operatorname{ran} A^*} \oplus \ker A^*.$$

By the (4.26), for any linear operator $A \in \mathcal{B}(\mathcal{H})$ on a finite-dimensional Hilbert space, there exists an ONB $e_1, \dots, e_r, e_{r+1}, \dots, e_d$, such that $e_i \in \operatorname{ran} A^*$, $i = 1, \dots, r$, and $e_i \in \ker A$, $i = r+1, \dots, d$. Hence, the matrix of A in this ONB is of the form

$$A = \begin{array}{|c|c|} \hline A_{11} & 0_{12} \\ \hline A_{21} & 0_{22} \\ \hline \end{array},$$

where $A_{11} \in \mathcal{B}(\operatorname{ran} A^*, \operatorname{ran} A^*)$, $A_{21} \in \mathcal{B}(\operatorname{ran} A^*, \ker A)$ may be arbitrary operators, and $0_{12} \in \mathcal{B}(\ker A, \operatorname{ran} A^*)$, $0_{22} \in \mathcal{B}(\ker A, \operatorname{ran} A^*)$ are the 0-operators between the respective spaces. By (4.27), if A is normal, then $\operatorname{ran} A = \operatorname{ran} A^*$. Since $\operatorname{ran} A$ is clearly invariant under A, we get that the matrix of A in an ONB as above is of the form

$$A = \begin{array}{|c|c|} \hline A_{11} & 0_{12} \\ \hline 0_{21} & 0_{22} \\ \hline \end{array}.$$

That is, A acts non-trivially only on $\operatorname{ran} A$, and exactly there. This motivates to introduce the following:

Definition 4.167. Let $A \in \mathcal{B}(\mathcal{H})$ be a normal operator. The *support* of A is defined as

$$\operatorname{supp} A := \overline{\operatorname{ran} A} = (\ker A)^\perp.$$

It is clear that $\mathcal{B}(\mathcal{H})_{\text{sa}}$ is closed under the adjoint, and under real linear combinations, i.e., it is a real vector space w.r.t. the usual pointwise operations on linear operators. However, it is not closed under the operator product:

Exercise 4.168. Show that the product of two self-adjoint operators is self-adjoint if and only if the two operators commute.

Score: 2 points.

Exercise 4.169. Let $A \in \mathcal{B}(\mathcal{H})$. Show that A is self-adjoint if and only if $\langle x, Ax\rangle \in \mathbb{R}$ for every $x \in \mathcal{H}$.

Score: 5 points.
Solution: We have

$$\langle x, A^*x\rangle = \langle Ax, x\rangle = \overline{\langle x, Ax\rangle}, \qquad x \in \mathcal{H}. \tag{4.28}$$

Thus,

$$\begin{aligned} A = A^* &\iff \langle x, Ax\rangle = \langle x, A^*x\rangle \\ &\iff \langle x, Ax\rangle = \overline{\langle x, Ax\rangle} \iff \langle x, Ax\rangle \in \mathbb{R}, \end{aligned}$$

where the first equivalence is due to Exercise 4.63, and the second is due to (4.28).

It is easy to see that any operator $A \in \mathrm{Lin}(\mathcal{H})$ can be decomposed as $A = A_1 + iA_2$, where A_1 and A_2 are self-adjoint; indeed

$$A = \frac{A + A^*}{2} + i\frac{A - A^*}{2i}. \tag{4.29}$$

Exercise 4.170. Show that the above decomposition is unique, i.e., if $A = A_1 + iA_2$ is a decomposition of A such that A_1, A_2 are self-adjoint then $A_1 = \frac{A+A^*}{2}$ and $A_2 = \frac{A-A^*}{2i}$.

Exercise 4.171. Show that $A \in \mathcal{B}(\mathcal{H})$ is normal if and only if there exist two commuting self-adjoint operators $A_1, A_2 \in \mathcal{B}(\mathcal{H})$ such that $A = A_1 + iA_2$.

4.18 The Fourier transform

Definition 4.172. For every $n \in \mathbb{N}$, the *n-th Hermite polynomial* is defined by

$$H_n(x) := (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}. \tag{4.30}$$

A straightforward computation shows that

$$H_0(x) \equiv 1, \qquad H_1(x) = 2x, \qquad H_2(x) = 4x^2 - 2.$$

Differentiating (4.30) yields that for every $n \in \mathbb{N}$,

$$H_n'(x) = 2xH_n(x) - H_{n+1}(x). \tag{4.31}$$

This in turn immediately implies the following:

Lemma 4.173. (i) For every $n \in \mathbb{N}$, H_n is a degree n polynomial in which the coefficient of x^n is 2^n.

(ii) The Hermite polynomials are linearly independent, and for every $n \in \mathbb{N}$,

$$\operatorname{span}\{H_k : k = 0, \dots, n\} = \operatorname{span}\{x^k : k = 1, \dots, n\},$$

where $x^0 := 1$.

Proof. (i) Clearly, the statement is true for $n = 0$. Since $H_{n+1}(x) = 2xH_n(x) - H_n'(x)$ by (4.31), the assertion follows by induction on n.

(ii) Immediate from the previous point.

□

Let us now define the *normalized Hermite functions*

$$\tilde{H}_n(x) := \left(2^n n! \sqrt{\pi}\right)^{-1/2} e^{-x^2/2} H_n(x), x \in \mathbb{R}, \qquad n \in \mathbb{N}.$$

Proposition 4.174. The normalized Hermite functions form an ONB in $L^2(\mathbb{R})$.

Proof. Let $n \geq m$. Then

$$\begin{aligned}
\int_{\mathbb{R}} e^{-x^2/2} H_n(x) e^{-x^2/2} H_m(x)\, d\lambda(x) &= \int_{\mathbb{R}} e^{-x^2} H_n(x) H_m(x)\, d\lambda(x) \\
&= \int_{\mathbb{R}} (-1)^n \left(\frac{d^n}{dx^n} e^{-x^2}\right) H_m(x)\, d\lambda(x) \\
&= \int_{\mathbb{R}} e^{-x^2} \frac{d^n}{dx^n} H_m(x)\, d\lambda(x) \\
&= \begin{cases} 0, & n \neq m, \\ 2^n n! \int_{\mathbb{R}} e^{-x^2}\, d\lambda(x), & n = m, \end{cases}
\end{aligned}$$

where the third equality follows by partial integration, and the last equality follows from (i) of Lemma 4.173. This yields that the normalized Hermite functions form an ONS.

Assume now that some $f \in L^2(\mathbb{R})$ is orthogonal to all the Hermite functions. By (ii) of Lemma 4.173, this is equivalent to

$$\int_{\mathbb{R}} e^{-x^2/2} x^n f(x)\, d\lambda(x) = 0, \qquad n \in \mathbb{N}. \tag{4.32}$$

This implies that for every $z \in \mathbb{C}$,

$$F(z) := \int_{\mathbb{R}} e^{zx} e^{-x^2/2} f(x)\, d\lambda(x) = \sum_{n=0}^{+\infty} \frac{z^n}{n!} \int_{\mathbb{R}} e^{-x^2/2} x^n f(x)\, d\lambda(x) = 0,$$

where the first equality follows by the Lebesgue dominated convergence theorem, and the second equality by (4.32). Substituting $z = -it$, $t \in \mathbb{R}$, yields that the Foruier transform of $e^{-\mathrm{id}^2/2} f$ is zero, and hence $f = 0$. □

Exercise 4.175. (i) Let $f(x) := e^{-x^2}$. Show that for every $n = 1, 2, \ldots$,

$$f^{(n+1)}(x) = -2x f^{(n)}(x) - 2n f^{(n-1)}(x),$$

and conclude that

$$H_{n+1}(x) = 2x H_n(x) - 2n H_{n-1}(x).$$

(Hint: Use induction.)

(ii) Show that for any $n = 1, 2, \ldots$,

$$H_n'(x) = 2n H_{n-1}(x).$$

Solution: Hidden.

(i)

4.19 Positive semi-definite operators and the PSD order

Definition 4.176. We say that an operator $A \in \mathcal{B}(\mathcal{H})$ is *positive semidefinite* (or simply *positive*) if

$$\langle x, Ax \rangle \geq 0, \qquad x \in \mathcal{H}.$$

We say that A is *positive definite*, or *strictly positive*, if

$$\langle x, Ax \rangle > 0, \qquad x \in \mathcal{H} \setminus \{0\}.$$

We denote positivity of an operator A as $A \geq 0$, and strict positivity as $A > 0$. We denote the set of positive operators on $\mathcal{H}$ by $\mathcal{B}(\mathcal{H})_{\geq 0}$, and the set of strictly positive operators by $\mathcal{B}(\mathcal{H})_{>0}$.

Exercise 4.177. Show that if $A \in \mathcal{B}(\mathcal{H})_{\geq 0}$ and $X \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ then $X^*AX \in \mathcal{B}(\mathcal{K})_{\geq 0}$.

Exercise 4.178. Show that every positive operator is self-adjoint. (Hint: Use Exercise 4.169.)

Solution: If A is positive then, by definition, $\langle x, Ax\rangle \in \mathbb{R}_+$ for every $x \in \mathcal{H}$, and hence, by Exercise 4.169, A is self-adjoint.

Exercise 4.179. Let $B \in \mathcal{B}(\mathcal{H})_{\geq 0}$ be a positive semi-definite operator.

(i) Show that

$$\langle x, y\rangle_B := \langle x, By\rangle\,, \qquad x, y \in \mathcal{H},$$

defines a semi-inner product on $\mathcal{H}$, and it is an inner product if and only if B is positive definite.

(ii) Prove that for any $x, y \in \mathcal{H}$,

$$|\langle x, By\rangle|^2 \leq \langle x, Bx\rangle \langle y, By\rangle\,. \tag{4.33}$$

(iii) Prove that if $B > 0$ then for any $x, y \in \mathcal{H}$,

$$|\langle x, y\rangle|^2 \leq \langle x, Bx\rangle \left\langle y, B^{-1}y\right\rangle. \tag{4.34}$$

Score: 4+2+2=8 points.
Solution:

(i) The sesquilinearity of $\langle ., .\rangle_B$ is trivial from the sesquilinearity of the original inner product and the linearity of B. We have

$$\overline{\langle x, y\rangle_B} = \overline{\langle x, By\rangle} = \langle By, x\rangle = \langle y, Bx\rangle = \langle y, x\rangle_B\,, \qquad x, y \in \mathcal{H},$$

due to the self-adjointness of B. Finally,

$$\langle x, x\rangle_B = \langle x, Bx\rangle \geq 0 \;\text{ with equality if and only if }\; x = 0,$$

due to the positive semidefiniteness and the strict positive definiteness of B, respectively.

(ii) The inequality in (4.33) is simply the Cauchy-Schwarz inequality for the inner product $\langle ., .\rangle_B$.

(iii) Simply replace y with $B^{-1}y$ in (4.33).

□

Exercise 4.180. Show that the set of positive semi-definite operators on a Hilbert space $\mathcal{H}$ forms a convex cone, i.e.,

$$\begin{aligned} A, B \in \mathcal{B}(\mathcal{H})_{\geq 0} &\implies A + B \in \mathcal{B}(\mathcal{H})_{\geq 0}, \\ A \in \mathcal{B}(\mathcal{H})_{\geq 0},\ \lambda \in \mathbb{R}_{\geq 0} &\implies \lambda A \in \mathcal{B}(\mathcal{H})_{\geq 0}. \end{aligned}$$

Exercise 4.181. Show that $\mathcal{B}(\mathcal{H})_{\geq 0}$ is a *pointed cone*, i.e.,

$$\text{if } \ A \in \mathcal{B}(\mathcal{H})_{\geq 0} \ \text{ and } \ -A \in \mathcal{B}(\mathcal{H})_{\geq 0} \quad \text{ then } \quad A = 0,$$

or equivalently,

$$\mathcal{B}(\mathcal{H})_{\geq 0} \cap -\mathcal{B}(\mathcal{H})_{\geq 0} = \{0\}.$$

Exercise 4.182. Show that the relation

$$A \geq B \quad \text{ if } \quad A - B \in \mathcal{B}(\mathcal{H})_{\geq 0}$$

defines a partial order on $\mathcal{B}(\mathcal{H})$, which we call the *positive semidefinite order*. When we want to emphasize which order we mean, we use the notation $A \geq_{\text{PSD}} B$.

Remark 4.183. Note that the restriction of a partial order on a set X to a subset $X_0 \subseteq X$ gives a partial order on X_0. In particular, we may restrict the PSD order to any observable algebra $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$. Also, it is common to consider the PSD order only on $\mathcal{B}(\mathcal{H})_{\text{sa}}$, the self-adjoint operators on $\mathcal{H}$.

Exercise 4.184. Show that if $A \in \mathcal{B}(\mathcal{H})_{\text{sa}}$ and $B \geq_{\text{PSD}} A$ then B is also self-adjoint.

Exercise 4.185. Show that the PSD order is compatible with the vector space structure of $\mathcal{B}(\mathcal{H})$ in the sense that for any $A, B \in \mathcal{B}(\mathcal{H})$,

$$\begin{aligned} A \leq B &\implies \lambda A \leq \lambda B, \qquad \lambda \in \mathbb{R}_{\geq 0}, \\ A \leq B &\implies A + C \leq B + C, \qquad C \in \mathcal{B}(\mathcal{H}). \end{aligned}$$

Remark 4.186. See Section **??** for a more general treatment of convex cones, partially ordered vector spaces, and related notions.

Exercise 4.187. Show that if $\dim \mathcal{H} > 1$ then the PSD order is not a complete order on $\mathcal{B}(\mathcal{H})_{\text{sa}}$, i.e., there exist self-adjoint operators $A, B \in \mathcal{B}(\mathcal{H})_{\text{sa}}$ such that neither $A \geq B$ nor $A \leq B$ holds.

Remark 4.188. The above is more generally true for the self-adjoint part of any von Neumann algebra $\mathcal{A}$ on a finite-dimensional Hilbert space $\mathcal{H}$ with $\dim \mathcal{H} \geq 2$, such that $\mathcal{A} \neq \mathbb{C}I$; see Section 4.27 for the definition of a von Neumann algebra.

Exercise 4.189. (i) Show that the cone $\mathcal{B}(\mathcal{H})_{\geq 0}$ is *self-dual*, i.e., for any $A \in \mathcal{B}(\mathcal{H})$,

$$A \in \mathcal{B}(\mathcal{H})_{\geq 0} \quad \iff \quad \operatorname{Tr} AB \geq 0 \ \ \forall B \in \mathcal{B}(\mathcal{H})_{\geq 0}.$$

Solution: Hidden.

Remark 4.190. See Section **??** for the general notion of duality in ordered vector spaces.

Exercise 4.191. Let $A, B \in \mathcal{B}(\mathcal{H})_{>0}$ be strictly positive definite operators on a Hilbert space $\mathcal{H}$. Prove that

$$A \geq B \iff \langle x, Ax\rangle \left\langle y, B^{-1}y\right\rangle \geq |\langle x, y\rangle|^2, \quad x, y \in \mathcal{H}.$$

Solution: Hidden.

Remark 4.192. Note that (4.34) can be written as

$$\left\langle x \otimes y, (B \otimes B^{-1})x \otimes y\right\rangle \geq \langle x \otimes y, Fx \otimes y\rangle, \qquad x, y \in \mathcal{H},$$

i.e., $B \otimes B^{-1} - F$ is block-positive, where F is the flip operator defined by $Fx \otimes y := y \otimes x$, $x, y \in \mathcal{H}$. Does this offer any useful insight into (4.34)?

4.20 Projections

Definition 4.193. An operator P on a Hilbert space $\mathcal{H}$ is a *projection* if

$$P^* = P = P^2.$$

Remark 4.194. Sometimes an operator P such that $P^2 = P$ is also called a projection, and if it is, moreover, self-adjoint, then it is called an *orthogonal projection*. In our terminology introduced in Definition 4.193, which is commonly used in the theory of Hilbert spaces, a projection is always self-adjoint.

Exercise 4.195. (i) Prove that every projection is PSD.

(ii) Show that if P is a projection then $I - P$ is a projection as well, and hence, in particular, $I - P \geq 0$, i.e., $0 \leq P \leq I$.

(iii) Let P be a projection. Show that $\|Px\| \leq \|x\|$ for all $x \in \mathcal{H}$, and

$$x \in \operatorname{ran} P \iff Px = x \iff \|Px\| = \|x\|\,. \tag{4.35}$$

Conclude that

a) The support of P is its fixed point set.

b) $\operatorname{ran} P = \ker(I-P)$, $\ker P = \operatorname{ran}(I-P)$, and

$$\mathcal{H} = \operatorname{ran} P \oplus \operatorname{ran}(I-P), \tag{4.36}$$

i.e., $I-P$ is the projection onto the orthocomplement of $\operatorname{ran} P$.

c) For all $x \in \mathcal{H}$, $\|x\|^2 = \|Px\|^2 + \|(I-P)x\|^2$.

d) If $P \neq 0$ then $\|P\| = 1$.

(iv) Let $P, Q \in \mathcal{B}(\mathcal{H})$ be projections. Show that

$$P \leq Q \iff \operatorname{ran} P \subseteq \operatorname{ran} Q \iff PQ = QP = P.$$

Solution: Hidden.

Exercise 4.196. Show that for any vector $v \in \mathcal{H}$, $|v\rangle\langle v|$ is positive, and it is a projection if and only if $\|v\| = 1$.

Solution: Hidden.

Exercise 4.197. (i) Show that if P and Q are projections then

$$P+Q \text{ is a projection} \iff QP = PQ = 0 \iff \operatorname{ran} P \perp \operatorname{ran} Q.$$

Show that if any (and hence all) of the above holds then $P+Q$ is the projection onto $\operatorname{span}\{\operatorname{ran} P, \operatorname{ran} Q\}$.

(ii) Let $P_1, \ldots, P_r$ be non-zero projections on a Hilbert space $\mathcal{H}$. Show that

$$P_1 + \ldots + P_r \leq I \iff P_i P_j = 0,\ \forall i \neq j,$$

i.e., the projections are pairwise orthogonal.

Score: 6+6=12 points.

Solution: Hidden.

(i) Obviously, $(P+Q)^* = P+Q$, and $(P+Q)^2 = P^2 + PQ + QP + Q^2 = P + Q + PQ + QP$. Hence, $P+Q$ is a projection if and only if $PQ+QP = 0$. Multiplying by P from both sides, we get $0 = PQP = (PQ)(QP) = (QP)^*(QP)$, and hence $QP = 0$. Taking the adjoint, we get $PQ = 0$. Conversely, if $PQ = QP = 0$ then $(P+Q)^2 = (P+Q)$, and hence $P+Q$ is a projection.

(ii) Assume that there exist $i \neq j$ such that $P_i P_j \neq 0$. By the previous part, this means that there exists an $x \in \operatorname{ran} P_i$ such that $x \notin \ker P_j$, i.e., $0 < \|P_j x\|^2 = \langle P_j x, P_j x \rangle = \langle x, P_j x \rangle$. We can assume without loss of generality that $\|x\| = 1$, and hence we get

$$1 = \|x\|^2 = \langle x, Ix \rangle = \left\langle x, \sum_k P_k x \right\rangle = \sum_k \langle x, P_k x \rangle \geq \langle x, P_i x \rangle + \langle x, P_j x \rangle$$
$$= 1 + \langle x, P_j x \rangle > 1,$$

a contradiction.

Exercise 4.198. Let P, Q be projections. Show that

$$PQ \text{ is a projection} \iff PQ = QP,$$

and that in this case PQ is the projection onto $\operatorname{ran} P \cap \operatorname{ran} Q$.

Score: 6 points.

In general, a finite set of self-adjoint operators does not have a smallest upper bound or largest lower bound among the self-adjoint operators w.r.t. the PSD ordering; see Section 4.30. The situation is different if we restrict to the set of projections, as Exercise 4.200 shows.

Definition 4.199. For two projections P, Q, let $P \vee Q$ be the projection onto $\operatorname{span}\{\operatorname{ran} P \cup \operatorname{ran} Q\}$, and $P \wedge Q$ be the projection onto $\operatorname{ran} P \cap \operatorname{ran} Q$.

Exercise 4.200. Show that $P \vee Q$ is the smallest upper bound, and $P \wedge Q$ is the largest lower bound, to $\{P, Q\}$ in the PSD ordering.

4.21 Isometries and unitaries

In this section we discuss how Hilbert spaces, or, more generally, subspaces of Hilbert spaces can be identified. Recall that vector spaces are isomorphic if there exists a bijective linear map (an isomorphism) between them. It is well known that two vector spaces are isomorphic if and only if they have the same dimension. Moreover, isomoprhisms are exactly those linear maps that map a basis of one space into a basis of the other space.

Definition 4.201. We say that two Hilbert spaces $\mathcal{H}, \mathcal{K}$ is *isomorphic* if there exists a bijective linear map $U : \mathcal{H} \to \mathcal{K}$ that preserves the inner product, i.e.,

$$\langle Ux, Uy \rangle = \langle x, y \rangle .$$

Any such map U is called a *unitary*.

When we want to stress that $\mathcal{H}$ and $\mathcal{K}$ can be identified as Hilbert spaces, and not only as vector spaces, i.e., there exists a unitary from $\mathcal{H}$ to $\mathcal{K}$, then we say that $\mathcal{H}$ and $\mathcal{K}$ are *isometrically isomorphic*. Part of the reason for this terminology is that, as we can see in the next exercise, a linear map preserves the inner product if and only if it preserves the norm. Norm-preserving linear maps are called isometries:

Definition 4.202. An operator $V \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ is an *isometry* if

$$\|Vx\| = \|x\|, \qquad x \in \mathcal{H}.$$

Exercise 4.203. Let $V \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$. Show that the following are equivalent:

(i) V is an isometry.

(ii) V preserves the inner product, i.e.,

$$\langle Vx, Vy\rangle = \langle x, y\rangle, \qquad x, y \in \mathcal{H}.$$

(iii) V maps every orthonormal system in $\mathcal{H}$ into an orthonormal system in $\mathcal{K}$.

(iv) There exists an orthonormal basis in $\mathcal{H}$ that V maps into an orthonormal system in $\mathcal{K}$.

(v) $V^*V = I$.

Note that an isometry is always injective. In particular, if there exists an isometry from $\mathcal{H}$ to $\mathcal{K}$, then necessarily $\dim \mathcal{H} \leq \dim \mathcal{K}$. It is easy to see that the converse is also true.

Exercise 4.204. Show that there exists an isometry from $\mathcal{H}$ to $\mathcal{K}$ if and only if $\dim \mathcal{H} \leq \dim \mathcal{K}$.

Exercise 4.205. Give an explicit isometry from $\mathbb{C}^2$ to $\mathbb{C}^3$.

Exercise 4.206. Show that for any Hilbert space $\mathcal{H}$,

$$\mathcal{H} \ni v \mapsto |v\rangle \in \mathcal{B}(\mathbb{C}, \mathcal{H}), \qquad \mathcal{B}(\mathbb{C}, \mathcal{H}) \ni V \mapsto V1 \in \mathcal{H}$$

are linear maps that are inverses of each other, and thus they give explicit isomorphisms from $\mathcal{H}$ to $\mathcal{B}(\mathbb{C}, \mathcal{H})$ and back, under which unit vectors correspond to isometries.

Exercise 4.207. Show that the map $\mathcal{H} \ni x \mapsto |x\rangle \in \mathcal{B}(\mathbb{K}, \mathcal{H})$ gives a one-to-one correspondence between unit vectors in $\mathcal{H}$ and isometries from $\mathbb{K}$ to $\mathcal{H}$. (See also Exercise 4.85).

By definition, isometries are different from unitaries in that they need not be surjective. In fact, isometries identify a Hilbert space with a subspace of a Hilbert space, and any isometry V from $\mathcal{H}$ to $\mathcal{K}$ is a unitary from $\mathcal{H}$ to $\operatorname{ran} V$. An isometry is a unitary if and only if it is surjective (and hence bijective), that can be summiraized concisely the following way:

Exercise 4.208. Show that for a linear map $U : \mathcal{H} \to \mathcal{K}$ we have

$$U \text{ is a unitary} \quad \Longleftrightarrow \quad U^*U = I \text{ and } UU^* = I.$$

Conclude that every unitary is a normal operator, and for a unitary U,

$$U^* = U^{-1}.$$

Exercise 4.209. Show that if $V : \mathcal{H} \to \mathcal{K}$ is an isometry, and $\dim \mathcal{H} = \dim \mathcal{K}$, then V is also a unitary.

As for vector spaces, we see that two finite-dimensional Hilbert spaces can be identified if and only their dimensions are the same.

Exercise 4.210. Let $\mathcal{H}, \mathcal{K}$ be finite-dimensional Hilbert spaces. Show that there exists a unitary $U : \mathcal{H} \to \mathcal{K}$ if and only if $\dim \mathcal{H} = \dim \mathcal{K}$. Conclude that every d-dimensional Hilbert space is isometrically isomorphic to $\mathbb{C}^d$, where the latter is equipped with its standard inner product $\langle x, y\rangle := \sum_{k=1}^{d} \bar{x}_k y_k$, $x, y \in \mathbb{C}^d$.

We have seen above that an isometry $V : \mathcal{H} \to \mathcal{K}$ identifies $\mathcal{H}$ with a subspace of $\mathcal{K}$. It is useful to further generalize this concept, and introduce a name for operators that identify a subspace in their domain with a subspace in their target space.

Definition 4.211. A linear map $V : \mathcal{H} \to \mathcal{K}$ is a *partial isometry* if it is an isometry on $(\ker V)^\perp$, i.e.,

$$\|Vx\| = \|x\|, \qquad x \in (\ker V)^\perp.$$

Lemma 4.212. For a $V \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, the following are equivalent:

(i) V is a partial isometry.

(ii) V^*V is the projection onto $(\ker V)^\perp$.

(iii) V^*V is a projection.

(iv) V^* is a partial isometry.

(v) VV^* is the projection onto $\operatorname{ran} V$.

(vi) VV^* is a projection.

(vii) There exist orthonormal systems $(e_i)_{i=1}^r$ in $\mathcal{H}$ and $(f_i)_{i=1}^r$ in $\mathcal{K}$ such that

$$V = \sum_{i=1}^{r} |f_i\rangle\langle e_i| \,. \tag{4.37}$$

Proof. (i)$\Longrightarrow$ (ii) Assume that V is an isometry on $(\ker V)^\perp$, and let P be the projection onto $(\ker V)^\perp$. Then for every $x \in \mathcal{H}$ we have

$$\langle x, V^*Vx\rangle = \langle Vx, Vx\rangle = \langle VPx, VPx\rangle = \|VPx\|^2 = \|Px\|^2 = \langle Px, Px\rangle = \langle x, Px\rangle \,,$$

where the fourth identity follows by assumption (i). Hence, by Exercise **??**, $V^*V = P$.

(ii)$\Longrightarrow$ (iii) Trivial.

(iii)$\Longrightarrow$ (i) Assume that V^*V is a projection, and let $x \in \ker(V)^\perp$. Note that $\ker V = \ker V^*V = \operatorname{ran}(V^*V)^\perp$ (in fact, for any operator V), and hence $x \in \operatorname{ran} V^*V$. Since V^*V is a projection, this means that $x = V^*Vx$. Thus,

$$\|Vx\|^2 = \langle Vx, Vx\rangle = \langle x, V^*Vx\rangle = \langle x, x\rangle = \|x\|^2 \,,$$

i.e., V is indeed an isometry on $\ker(V)^\perp$, proving (i).

Now the equivalences (iv)$\Longleftrightarrow$(v)$\Longleftrightarrow$(vi) follow by replacing V with V^* in the above.

(ii)$\Longrightarrow$ (vi) Obviously, VV^* is self-adjoint, and we have $(VV^*)(VV^*) = V(V^*V)V^* = VP_{(\ker V)^\perp}V^* = VV^*$. (v)$\Longrightarrow$ (iii) Follows the same way by replacing V with V^*.

(vii)$\Longrightarrow$ (iii) Obviously, $V^*V = \sum_{i,j=1}^r |e_j\rangle\langle f_j|\, |f_i\rangle\langle e_i| = \sum_{i,j=1}^r \delta_{i,j}\, |e_j\rangle\langle e_i| = \sum_{i=1}^r |e_i\rangle\langle e_i|$ is a projection.

(i)$\Longrightarrow$ (vii) Let $(e_i)_{i=1}^r$ be an ONB in $(\ker V)^\perp$. Since V is an isometry on $(\ker V)^\perp$, $f_i := Ve_i$, $i = 1, \ldots, r$ is an ONS in $\mathcal{K}$ according to Exercise 4.203, and we have (4.37). □

Remark 4.213. We may write "$V : \mathcal{H}_0 \to \mathcal{K}_0$ partial isometry" to mean that V is a partial isometry with $\ker V = \mathcal{H}_0^\perp$ and $\operatorname{ran} V = \mathcal{K}_0$.

Remark 4.214. By (4.37) above, V is a partial isometry if and only if all of its non-zero singular values are equal to 1; cf. Corollary 4.362.

Exercise 4.215. Let $\{x_i\}_{i\in\mathcal{I}} \subseteq \mathcal{H}$ and $\{y_i\}_{i\in\mathcal{I}} \subseteq \mathcal{K}$. Show that the following are equivalent:

(i) There exists a partial isometry $V : \overline{\operatorname{span}\{x_i\}_{i\in\mathcal{I}}} \to \overline{\operatorname{span}\{y_i\}_{i\in\mathcal{I}}}$ such that $Vx_i = y_i$, $i \in \mathcal{I}$.

(ii) $\langle x_i, x_j\rangle = \langle y_i, y_j\rangle$ for all $i, j \in \mathcal{I}$.

(iii) $G(\{x_i\}_{i\in\tilde{\mathcal{I}}}) = G(\{y_i\}_{i\in\tilde{\mathcal{I}}})$, for any finite subset $\tilde{\mathcal{I}} \subseteq \mathcal{I}$, where G stands for the Gram matrix.

Show that if $\mathcal{I}$ is finite then the above is further equivalent to $G(\{x_i\}_{i\in\mathcal{I}}) = G(\{y_i\}_{i\in\mathcal{I}})$.

(Hint: Check that the map $V(\sum_i c_i x_i) := \sum_i c_i y_i$ is a well-defined partial isometry under condition (ii).)

Solution: Hidden.

4.22 The trace and the Hilbert-Schmidt inner product

Exercise 4.216. Let $A \in \mathrm{Lin}(\mathcal{H})$ be a linear operator one a Hilbert space $\mathcal{H}$, and let $\{e_1, \ldots, e_d\}$ and $\{f_1, \ldots, f_d\}$ be two orthonormal bases in $\mathcal{H}$. Show that

$$\sum_{i=1}^{d} \langle e_i, Ae_i\rangle = \sum_{i=1}^{d} \langle f_i, Af_i\rangle\,.$$

Solution:

$$\begin{aligned}
\sum_{i=1}^{d} \langle e_i, Ae_i\rangle &= \sum_{i=1}^{d} \left\langle e_i, A\left(\sum_{k=1}^{d} |f_k\rangle\langle f_k|\right) e_i\right\rangle \\
&= \sum_{k=1}^{d}\sum_{i=1}^{d} \langle e_i, Af_k\rangle \langle f_k, e_i\rangle \\
&= \sum_{k=1}^{d}\sum_{i=1}^{d} \langle f_k, e_i\rangle \langle e_i, Af_k\rangle \\
&= \sum_{k=1}^{d} \left\langle f_k, \left(\sum_{i=1}^{d} |e_i\rangle\langle e_i|\right) Af_k\right\rangle \\
&= \sum_{k=1}^{d} \langle f_k, Af_k\rangle\,.
\end{aligned}$$

Definition 4.217. Let $A \in \mathrm{Lin}(\mathcal{H})$. The *trace* of A (in notation, $\mathrm{Tr}\, A$) is defined as

$$\mathrm{Tr}\, A := \sum_{i=1}^{d} \langle e_i, Ae_i\rangle\,, \tag{4.38}$$

where $\{e_1, \ldots, e_d\}$ is an arbitrary orthonormal basis in $\mathcal{H}$. According to Exercise 4.216, the value of the sum in (4.38) is independent of the choice of the orthonormal basis.

Exercise 4.218. Show the following properties of the trace:

(i) Linearity:

$$\operatorname{Tr}(A+B) = \operatorname{Tr} A + \operatorname{Tr} B, \qquad \operatorname{Tr} \lambda A = \lambda \operatorname{Tr} A, \qquad A, B \in \operatorname{Lin}(\mathcal{H}),\ \lambda \in \mathbb{C}.$$

(ii) Cyclic property:

$$\operatorname{Tr} AB = \operatorname{Tr} BA, \qquad A \in \operatorname{Lin}(\mathcal{H}, \mathcal{K}), \mathcal{B} \in \operatorname{Lin}(\mathcal{K}, \mathcal{H}).$$

(iii) Self-adjointness:

$$\operatorname{Tr} A^* = \overline{\operatorname{Tr} A}, \qquad A \in \operatorname{Lin}(\mathcal{H}).$$

Score: 2+2+2=6 points.
Solution: Straightforward computation.

Exercise 4.219. Let $A \in \mathcal{B}(\mathcal{H})_+$ be a PSD operator.

(i) Prove that $\operatorname{Tr} A \geq 0$.

(ii) Let $e_1, \ldots, e_d$ be an ONB in $\mathcal{H}$. Prove that if $\langle e_i, Ae_i \rangle = 0$ for some i then $\langle e_j, Ae_i \rangle = 0$ and $\langle e_i, Ae_j \rangle = 0$ for all j, i.e., if a diagonal element of the matrix of A in the given ONB is zero then the corresponding row and column are zero as well.

(Hint: Use Exercise 4.179.)

(iii) Prove that if all diagonal elements of A in an ONB are zero then $A = 0$.

(iv) Prove that $\operatorname{Tr} A = 0 \iff A = 0$.

Score: 1+3+1+2=7 points.
Solution:

(i) We have $\operatorname{Tr} A = \sum_{i=1}^{d} \langle e_i, Ae_i \rangle$, where $e_1, \ldots, e_d$ is any ONB, and $A \geq 0$ implies $\langle e_i, Ae_i \rangle \geq 0$ for all i by definition.

(ii) By Exercise 4.179, $\gamma_A(x, y) := \langle x, Ay \rangle$ is a PSD hermitian sesquilinear form. By the Cauchy-Schwarz inequality for this form,

$$\begin{aligned} 0 \leq |\langle e_j, Ae_i \rangle| &= |\gamma_A(e_j, e_i)| \leq \gamma_A(e_j, e_j)^{1/2} \gamma_A(e_i, e_i)^{1/2} \\ &= \langle e_j, Ae_j \rangle \langle e_j, Ae_i \rangle . \end{aligned}$$

Thus, if $\langle e_i, Ae_i \rangle = 0$ then $\langle e_j, Ae_i \rangle = 0$ for every j. The other assertion follows the same way.

(iii) Immediate from the previous point.

(iv) The direction $A = 0 \implies \operatorname{Tr} A = 0$ is obvious from the trace being a linear functional. In the converse direction, note that $\langle e_i, Ae_i\rangle \geq 0$ for all i, and hence $0 = \operatorname{Tr} A = \sum_{i=1}^{d} \langle e_i, Ae_i\rangle$ implies $\langle e_i, Ae_i\rangle = 0$ for all i, and the assertion follows from the previous point.

Definition 4.220. Let $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ be a *-subalgebra. We say that a linear functional $\varphi \in \operatorname{Lin}(\mathcal{A}, \mathbb{K})$ is *positive*, in notation $\varphi \geq 0$, if

$$A \geq 0 \implies \varphi(A) \geq 0.$$

We say that a positive linear functional φ is *faithful*, if $\varphi(A) = 0$ for an $A \in \mathcal{B}(\mathcal{H})_+$ implies that $A = 0$.

Remark 4.221. By Exercise 4.219, the trace is a faithful positive functional on $\mathcal{B}(\mathcal{H})$.

Recall that for $A, B \in \mathcal{B}(\mathcal{H})$, we write $A \leq B$ if $B - A$ is PSD.

Exercise 4.222. Show that if φ is a positive linear functional on $\mathcal{B}(\mathcal{H})$ then it is *monotone*, i.e., for any $A, B \in \mathcal{B}(\mathcal{H})$,

$$A \leq B \implies \varphi(A) \leq \varphi(B).$$

Conclude that the trace is monotone, i.e.,

$$A \leq B \implies \operatorname{Tr} A \leq \operatorname{Tr} B.$$

Score: 3 points.
Solution: We have

$$A \leq B \iff 0 \leq B - A \implies 0 \leq \varphi(B - A) = \varphi(B) - \varphi(A),$$

where both the equivalence and the implication are by definition, and the last equality is by the linearity of φ. The assertion about the trace follows from this, as it is a positive linear functional by Exercise 4.219.

Lemma 4.223. Let $\mathcal{H}, \mathcal{K}$ be finite-dimensional Hilbert spaces. Then

$$\langle A, B\rangle_{HS} := \operatorname{Tr} A^* B, \qquad A, B \in \operatorname{Lin}(\mathcal{H}, \mathcal{K}),$$

defines an inner product on $\operatorname{Lin}(\mathcal{H}, \mathcal{K})$, that we call the *Hilbert-Schmidt inner product*. Its induced norm is given by

$$\|A\|_{HS}^2 = \operatorname{Tr} A^* A, \qquad A \in \operatorname{Lin}(\mathcal{H}, \mathcal{K}). \tag{4.39}$$

Proof. Sesquilinearity of $\langle .,. \rangle_{HS}$ is obvious from the definition, and we have

$$\langle B, A \rangle_{HS} = \operatorname{Tr} B^* A = \operatorname{Tr}(A^* B)^* = \overline{\operatorname{Tr} A^* B} = \langle A, B \rangle_{HS} ,$$

showing the hermiticity of $\langle .,. \rangle_{HS}$. Given a basis $\{e_i\}_{i=1}^{d_1}$, we have

$$\|A\|_{HS}^2 = \langle A, A \rangle_{HS} = \sum_{i=1}^{d_1} \langle e_i, A^* A e_i \rangle = \sum_{i=1}^{d_1} \|A e_i\|^2 ,$$

which is obviously non-negative, and equal to zero if and only if $Ae_i = 0$ for all $i = 1, \ldots, d_1$, which is equivalent to A being zero. This proves that $\langle .,. \rangle_{HS}$ is indeed an inner product, and hence (4.39) defines a norm. □

Corollary 4.224. Since $\langle .,. \rangle_{HS}$ is an inner product, it satisfies the Cauchy-Schwarz inequality, i.e.,

$$|\operatorname{Tr} A^* B| \leq (\operatorname{Tr} A^* A)^{\frac{1}{2}} (\operatorname{Tr} B^* B)^{\frac{1}{2}}, \qquad A, B \in \operatorname{Lin}(\mathcal{H}, \mathcal{K}).$$

Exercise 4.225. Show that for any $A, B \in \mathcal{B}(\mathcal{H}, \mathcal{K})$,

$$\overline{\langle A, B \rangle}_{HS} = \langle A^*, B^* \rangle_{HS} .$$

Solution: Hidden.

Exercise 4.226. Let $\{e_i\}_{i=1}^{d_1}$, $\{f_j\}_{j=1}^{d_2}$ be orthonormal bases in the finite-dimensional Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, respectively.

(i) Show that

$$\{E_{ij} := |f_i\rangle \langle e_j|\}_{i=1,j=1}^{d_2,\ d_1} \qquad \text{is an orthonormal basis in } (\operatorname{Lin}(\mathcal{H}, \mathcal{K}),\ \langle .,. \rangle_{HS}).$$

(Hint: Use Exercise 4.157.)

(ii) Show that the expansion coefficients of an $A \in \operatorname{Lin}(\mathcal{H}, \mathcal{K})$ in the above ONB coincide with the usual matrix elements of A in the given pair of bases, i.e.,

$$\langle E_{ij}, A \rangle_{HS} = \langle f_i, A e_j \rangle , \quad i \in [d_2],\ j \in [d_1].$$

(iii) Conclude that every $A \in \operatorname{Lin}(\mathcal{H}, \mathcal{K})$ can be uniquely expanded as

$$A = \sum_{i=1}^{d_2} \sum_{j=1}^{d_1} \langle f_i, A e_j \rangle \, |f_i\rangle \langle e_j| .$$

(iv) Show that for any $A, B \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$,

$$\langle A, B\rangle_{HS} = \sum_{i=1}^{d_1}\sum_{j=1}^{d_2} \overline{\langle f_j, Ae_i\rangle}\, \langle f_j, Be_i\rangle\,,$$

i.e., the Hilbert-Schmidt inner product coincides with the standard inner product on $\mathbb{C}^{[d_1]\times[d_2]}$ when operators in $\mathrm{Lin}(\mathcal{H}, \mathcal{K})$ are represented as matrices in the given bases.

Score: 2+2+2+2=8 points.
Solution:

(i) We have

$$\begin{aligned}\langle E_{ij}, E_{kl}\rangle_{HS} &= \mathrm{Tr}(|f_i\rangle\langle e_j|)^* |f_k\rangle\langle e_l| = \mathrm{Tr}\,|e_j\rangle\langle f_i|\,|f_k\rangle\langle e_l| = \delta_{i,k}\,\mathrm{Tr}\,|e_j\rangle\langle e_l| \\ &= \delta_{i,k}\langle e_l, e_j\rangle = \delta_{i,k}\delta_{j,l} = \delta_{(i,j),(k,l)},\end{aligned}$$

and hence the matrix units form an ONS. Since their number is $d_1 d_2 = \dim \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, they form an ONB. Now, for any $A \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, we have

$$\langle E_{ij}, A\rangle_{HS} = \mathrm{Tr}(|f_i\rangle\langle e_j|)^* A = \mathrm{Tr}\,|e_j\rangle\langle f_i|\,A = \langle f_i, Ae_j\rangle\,,$$

which is exactly the (i, j) entry of the matrix of A in the given pair of ONBs.

(ii) Immediate from the previous point.

(iii) Immediate from the previous point.

An alternative proof can be given as follows. Expanding the trace in the basis $\{e_i\}_{i=1}^{d_1}$,, and inserting an identity of the form $I_{\mathcal{K}} = \sum_{j=1}^{d_2} |f_j\rangle\langle f_j|$, we get

$$\begin{aligned}\langle A, B\rangle_{HS} &= \sum_{i=1}^{d_1}\langle e_i, A^*Be_i\rangle = \sum_{i=1}^{d_1}\langle e_i|\,A^*\left(\sum_{j=1}^{d_2}|f_j\rangle\langle f_j|\right)B\,|e_i\rangle \\ &= \sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\langle e_i, A^*f_j\rangle\langle f_j, Be_i\rangle = \sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\overline{\langle f_j, Ae_i\rangle}\,\langle f_j, Be_i\rangle\end{aligned}$$

Definition 4.227. The set of operators $\{|f_j\rangle\langle e_i|\}_{i=1,j=1}^{d_1,\,d_2}$ in Exercise 4.226 is called the set of *matrix units* corresponding to the ONB pair $\{e_i\}_{i=1}^{d_1}$, $\{f_j\}_{j=1}^{d_2}$.

Exercise 4.228. (i) Let $e_1, \ldots, e_d$ be an ONB in a finite-dimensional Hilbert space $\mathcal{H}$, and $E_{ij} := |e_i\rangle\langle e_j|$ be the corresponding set of matrix units. Show that for all $i, j, k, l \in [d]$,

$$E_{ij}^* = E_{ji}, \qquad E_{ij}E_{kl} = \delta_{j,k}E_{il}. \tag{4.40}$$

(ii) Let $\mathcal{H}$ be a finite-dimensional Hilbert space, and $E_{ij} \in \mathrm{Lin}(\mathcal{H})$, $i, j \in [d]$, be a set of operators satisfying (4.40). Show that either all $E_{ij} = 0$, or there exists an ONS $\{e_{i,a} : i \in [d],\, a \in [m]\}$ for some $m \in \mathbb{N}$, such that

$$E_{ij} = \sum_{a=1}^{m} |e_{i,a}\rangle \langle e_{j,a}|, \qquad i, j \in [d]. \tag{4.41}$$

While in the view of Exercise 4.226, a set of matrix units may seem like a very natural ONB in the space of operators, in quantum information theory, it is often beneficial to work with different families of ONBs. When $\mathcal{H} = \mathcal{K}$, a very important ONB for $\mathrm{Lin}(\mathcal{H})$ can be constructed from any ONB $e_0, \ldots, e_{d-1}$ of $\mathcal{H}$, in the following way. Define

$$Xe_k := e_{k+1 (\mathrm{mod}\, d)}, \qquad Ze_k := e^{i\frac{2\pi}{d}k} e_k, \qquad k = 0, \ldots, d-1.$$

(Recall that for $m, n \in \mathbb{N}$, $a = b \,(\mathrm{mod}\, d)$ if and only if $a - b$ is an integer multiple of d.)

Definition 4.229. For any ONB $e_0, \ldots, e_{d-1}$ in a finite-dimensional Hilbert space $\mathcal{H}$, the corresponding *discrete Weyl unitaries* are defined as

$$W_{a,b} := X^a Z^b, \qquad a, b = 0, \ldots, d-1.$$

Exercise 4.230. Let $W_{a,b} := X^a Z^b$, $a, b \in \{0, \ldots, d-1\}$, be the discrete Weyl operators corresponding to some ONB $e_0, \ldots, e_{d-1}$ in a finite-dimensional Hilbert space $\mathcal{H}$.

(i) Show that $W_{a,b}$ is a unitary for every a, b.

(ii) Show that $ZX = e^{i\frac{2\pi}{d}} XZ$, and the following relations hold:

$$W_{0,0} = I, \qquad W_{a,b}^* = e^{i\frac{2\pi}{d}ab} W_{-a,-b}, \qquad \mathrm{Tr}\, W_{a,b} = d\delta_{a,0}\delta_{b,0}.$$

(iii) Show that X, Z and $W_{a,b}$, $a, b \in \{0, \ldots, d-1\}$, are unitaries, and

$$\left\{ \frac{1}{\sqrt{d}} W_{a,b} : a, b \in \{0, \ldots, d-1\} \right\}$$

forms an orthonormal basis in $\mathrm{Lin}(\mathcal{H})$ with respect to the Hilbert-Schmidt inner product.

(iv) Prove that

$$\frac{1}{d^2} \sum_{a,b=0}^{d-1} W_{a,b} A W_{a,b}^* = I \frac{\mathrm{Tr}\, A}{d}, \qquad A \in \mathrm{Lin}(\mathcal{H}). \tag{4.42}$$

Score: 4+7+3+4=18 points.

Solution:

(i) Obviously, $W_{0,0} = I$ is a unitary. We have

$$\langle f_l, X f_k\rangle = \langle f_l, f_{k+1}\rangle = \delta_{l,k+1} = \delta_{l-1,k} = \langle f_{l-1}, f_k\rangle\,, \qquad k,l \in \{0,\dots,d-1\},$$

where everything is again modulo d. Hence, $X^* f_l = f_{l-1}\ \forall l$, i.e., $X^* = X^{-1}$; in particular, X is a unitary. For Z we get

$$\langle f_l, Z f_k\rangle = \left\langle f_l, e^{i\frac{2\pi}{d}} f_k\right\rangle = e^{i\frac{2\pi}{d}}\,\langle f_l, f_k\rangle = \left\langle e^{-i\frac{2\pi}{d}} f_l, f_k\right\rangle, \qquad k,l \in \{0,\dots,d-1\},$$

and hence $Z^* = Z^{-1}$, i.e., Z is a unitary. Since the product of unitaries is again a unitary, according to Exercise **??**, $W_{a,b} = X^a Z^b$ is a unitary for every $a,b \in \{0,\dots,d-1\}$.

(ii) By definition,

$$ZX f_k = Z f_{k+1} = e^{i\frac{2\pi}{d}(k+1)} f_{k+1} = X e^{i\frac{2\pi}{d}(k+1)} f_k = e^{i\frac{2\pi}{d}} XZ f_k,$$

where we used that $e^{i\frac{2\pi}{d}k} = e^{i\frac{2\pi}{d}l}$ exactly when $k = l\ (mod\ d)$.

Using now the unitarity of X and Z, and the commutation relation $ZX = e^{i\frac{2\pi}{d}} XZ$,

$$\begin{aligned} W_{a,b}^* &= (X^a Z^b)^* = (Z^*)^b (X^*)^a = Z^{d-b} X^{d-a} = e^{i\frac{2\pi}{d}(d-a)(d-b)} X^{d-a} Z^{d-b} \\ &= e^{i\frac{2\pi}{d}ab} X^{-a} Z^{-b} = e^{i\frac{2\pi}{d}ab} W_{-a,-b}. \end{aligned}$$

It is obvious that $\operatorname{Tr} W_{0,0} = \operatorname{Tr} I = d$. Assume now that $(a,b) \neq (0,0)$. Then

$$\operatorname{Tr} W_{a,b} = \sum_{k=0}^{d-1} \left\langle f_k, X^a Z^b f_k\right\rangle = \sum_{k=0}^{d-1} e^{i\frac{2\pi}{d}bk}\,\langle f_k, f_{k+a}\rangle\,.$$

If $a \neq 0$ then the above expression is 0, as all terms $\langle f_k, f_{k+a}\rangle = 0$. If $a = 0$ then, since $b \neq 0$ by assumption,

$$\operatorname{Tr} W_{a,b} = \sum_{k=0}^{d-1} e^{i\frac{2\pi}{d}bk} = \frac{e^{i\frac{2\pi}{d}bd} - 1}{e^{i\frac{2\pi}{d}b} - 1} = 0.$$

(iii) We have

$$W_{a,b}W_{m,n} = X^a Z^b X^m Z^n = e^{i\frac{2\pi}{d}mb} X^{a+m} Z^{b+n},$$

and hence

$$\begin{aligned}\langle W_{a,b}, W_{m,n}\rangle_{HS} &= \operatorname{Tr} W_{a,b}^* W_{m,n} = \operatorname{Tr} e^{i\frac{2\pi}{d}ab} W_{-a,-b} W_{m,n} \\ &= e^{i\frac{2\pi}{d}ab} e^{-i\frac{2\pi}{d}mb} \operatorname{Tr} W_{m-a,n-b} = d\delta_{a,m}\delta_{b,n},\end{aligned}$$

where we used $\operatorname{Tr} W_{a,b} = \delta_{a,0}\delta_{b,0}d$ from the previous point. Thus,

$$\left\langle \frac{1}{\sqrt{d}} W_{a,b}, \frac{1}{\sqrt{d}} W_{m,n} \right\rangle_{HS} = \delta_{(a,b),(m,n)},$$

showing that $\left\{ \frac{1}{\sqrt{d}} W_{a,b} : a, b \in \{0, \dots, d-1\} \right\}$ is an orthonormal system in $\operatorname{Lin}(\mathcal{H})$. Since its cardinality is $d^2 = \dim \operatorname{Lin}(\mathcal{H})$, it is an ONB.

(iv) Since the relation in (4.42) is linear, it is enough to check it on an orthonormal basis for $\operatorname{Lin}(\mathcal{H})$; we choose $\{|f_i\rangle\langle f_j|\}_{i,j=1}^d$. Then

$$\begin{aligned}&\sum_{a,b=0}^{d-1} W_{a,b} |f_k\rangle\langle f_l| W_{a,b}^* \\ &= \sum_{a,b=0}^{d-1} |W_{a,b} f_k\rangle\langle W_{a,b} f_l| = \sum_{a,b=0}^{d-1} e^{i\frac{2\pi}{d}bk} e^{-i\frac{2\pi}{d}bl} |f_{k+a}\rangle\langle f_{l+a}| \\ &= \sum_{a=0}^{d-1} |f_{k+a}\rangle\langle f_{l+a}| \sum_{b=0}^{d-1} e^{i\frac{2\pi}{d}b(k-l)} = \sum_{a=0}^{d-1} |f_{k+a}\rangle\langle f_{l+a}| \, d\delta_{k,l} \\ &= d\delta_{k,l} \sum_{a=0}^{d-1} |f_{k+a}\rangle\langle f_{k+a}| = d\delta_{k,l} I = dI \operatorname{Tr} |f_k\rangle\langle f_l|,\end{aligned}$$

which is exactly what we wanted to prove.

Exercise 4.231. Define the Pauli matrices as

$$\sigma_0 := I, \qquad \sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_1 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Show that $\left(\frac{1}{\sqrt{2}}\sigma_k\right)_{k=0}^3$ forms on ONB in $\mathcal{B}(\mathbb{C}^2)$.

Score: 6 points.
Solution: Straightforward computation.

Exercise 4.232. Let $\mathrm{Lin}(\mathcal{H},\mathcal{K})$ be equipped with the Hilbert-Schmidt inner product, and for every $A \in \mathrm{Lin}(\mathcal{K})$ and $B \in \mathrm{Lin}(\mathcal{H})$, define the left multiplication L_A and the right multiplication R_B as

$$L_A : X \mapsto AX, \qquad R_B : X \mapsto XB, \qquad X \in \mathrm{Lin}(\mathcal{H},\mathcal{K}).$$

Show the following:

(i) $L_{A_1} = L_{A_2} \iff A_1 = A_2$ and $R_{B_1} = R_{B_2} \iff B_1 = B_2$.

(ii) $(L_A)^* = L_{A^*}, \qquad (R_B)^* = R_{B^*}$.

(iii) L_A is normal/unitary/self-adjoint/positive/projection if and only if so is A, and the same holds for R_B and B.

(iv) If A is normal with spectral PVM P^A then L_A is normal with spectral PVM $P^{L_A} = L_{P^A}$. Formulate and prove the same statement for R_B.

(v) Let $\mathcal{K} = \mathcal{H}$, and for positive definite $A, B \in \mathrm{Lin}(\mathcal{H})$, define the *relative modular operator*

$$\Delta_{A/B} := L_A R_{B^{-1}}.$$

Show that $\Delta_{A/B}$ is a positive operator on $\mathrm{Lin}(\mathcal{H})$, and find its spectral decomposition.

Exercise 4.233. Let $\varphi \in \mathrm{Lin}(\mathrm{Lin}(\mathcal{H}),\mathbb{C})$ be a linear functional on $\mathrm{Lin}(\mathcal{H})$. Show that there exists a unique $\hat{\varphi} \in \mathrm{Lin}(\mathcal{H})$ such that

$$\varphi(X) = \mathrm{Tr}\,\hat{\varphi}X, \qquad X \in \mathrm{Lin}(\mathcal{H}).$$

Moreover,

$$\|\varphi\| = \|\hat{\varphi}\|_1,$$

and the following hold:

(i) $\hat{\varphi}$ is self-adjoint $\iff \varphi(X^*) = \overline{\varphi(X)}$, $X \in \mathrm{Lin}(\mathcal{H})$.

(ii) $\hat{\varphi} \geq 0 \iff \varphi \geq 0$, i.e., $\varphi(X) \geq 0 \ \ \forall X \in \mathrm{Lin}(\mathcal{H})_+$.

(iii) $\mathrm{Tr}\,\hat{\varphi} = 1 \iff \varphi(I) = 1$.

Prove that $\varphi \leftrightarrow \hat{\varphi}$ gives an identification of the set of positive linear functionals on $\mathrm{Lin}(\mathcal{H})$ that take the value 1 on the identity I, and the set of density operators on $\mathcal{H}$.

Show that if $\varphi(X^*) = \overline{\varphi(X)}$ for all $X \in \mathrm{Lin}(\mathcal{H})$ then there exist positive linear functionals φ_1 and φ_2 such that $\varphi = \varphi_1 - \varphi_2$.

(Hint: Use the results of Sections 4.11, 4.32 and 4.19.)

Solution: Hidden.

Remark 4.234. Let $(\mathcal{X}, \mathcal{F})$ be a measurable space, and T be a positive measure on $\mathcal{F}$. Every bounded measurable function $f \in L^\infty(\mathcal{X}, \mathcal{F}, T)$ defines a multiplication operator

$$M_f : g \mapsto fg, \qquad g \in L^2(\mathcal{X}, \mathcal{F}, T),$$

which is a bounded linear operator on $L^2(\mathcal{X}, \mathcal{F}, T)$. Note that the set of multiplication operators forms a vector space, and given any probability measure μ on $\mathcal{F}$, it defines a linear functional

$$\varphi_\mu(M_f) := \int f \, d\mu$$

on $\{M_f : f \in L^\infty(\mathcal{X}, \mathcal{F}, T)\}$, which is nothing else than the expectation value of f w.r.t. μ. This linear functional is positive, i.e., if $f \geq 0$ T-almost everywhere then $\varphi_\mu(f) \geq 0$, and it is normalized, i.e., $\varphi_\mu(I) = \varphi_\mu(M_{\mathbf{1}}) = 1$, where $\mathbf{1}$ denotes the constant 1 function.

Now, if μ is absolutely continuous w.r.t. T then, by the Radon-Nikodym theorem, there exists a T-integrable function $\hat{\mu}$ such that

$$\varphi_\mu(M_f) = \int f \, d\mu = \int \hat{\mu} f \, dT.$$

for all bounded measurable f. This $\hat{\mu}$ is called the *density function* of μ (w.r.t. T). When $\mathcal{X}$ is finite, and T is the counting measure, we have $L^2(\mathcal{X}, \mathcal{F}, \mu) = \mathbb{C}^{\mathcal{X}}$ with its canonical inner product, the multiplication operators can be represented by matrices that are diagonal in the canonical basis $\{e_x\}_{x \in \mathcal{X}}$ of $\mathbb{C}^{\mathcal{X}}$, and $\hat{\mu}$ is simply the function $\hat{\mu}(x) := \mu(x)$, $x \in \mathcal{X}$. The linear functional φ_μ can be extended to a positive linear functional on $\mathrm{Lin}(\mathbb{C}^{\mathcal{X}})$ by $\varphi_\mu(A) := \sum_{x \in \mathcal{X}} \mu(x) \langle e_x, Ae_x \rangle$, and we have

$$\varphi_\mu(A) = \sum_{x \in \mathcal{X}} \mu(x) \langle e_x, Ae_x \rangle = \sum_{x \in \mathcal{X}} \langle e_x, M_{\hat{\mu}} A e_x \rangle = \langle M_{\hat{\mu}}, A \rangle_{HS}.$$

Hence, $M_{\hat{\mu}} = \hat{\varphi}_\mu$, i.e., the density operator of φ_μ is the multiplication by the density function of μ. This is the origin of the terminology "density operator".

4.23 The spectral decomposition

Recall that a number $a \in \mathbb{K}$ is an *eigen-value* of $A \in \mathrm{Lin}(V)$, where V is a vector space over $\mathbb{K}$, if there exists a non-zero vector $v \in V$ such that

$$Av = av.$$

Any such vector v is called an *eigen-vector* of A, corresponding to the eigen-value a. The subspace spanned by the eigen-vectors corresponding to the same eigen-value $a \in A$ is called the *eigen-subspace* corresponding to a. The dimension of the eigen-subspace corresponding to an eigen-value a is called the *multiplicity* of the eigen-value.

Definition 4.235. The *spectrum* of A, denoted as $\mathrm{spec}(A)$, is the set of eigen-values of A.

If there exists a basis $(v_i)_{i=1}^d$ of V consisting of eigen-vectors of A, called an *eigen-basis* of A, then A can be *diagonalized* in this basis, i.e., its matrix in its eigen-basis is of the form

$$A = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{bmatrix},$$

where in the diagonal we have the eigen-values of A. Note that not every operator has an eigen-basis; a canonical example is $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

In the Hilbert space setting we are interested in the operators have an *orthonormal* eigen-basis. As the following theorem shows, these are exactly the normal operators.

Theorem 4.236. For every normal operator A on a finite-dimensional complex Hilbert space, there exists an orthonormal basis of $\mathcal{H}$ that consists of eigenvectors of A.

Proof. We prove by induction on the dimension of the Hilbert space. The assertion is trivial when $\dim \mathcal{H} = 1$, and assume that it is true in every Hilbert space of dimension at most d. Let $\mathcal{H}$ be a Hilbert space of dimension $d+1$ and let A be a normal operator on $\mathcal{H}$. Let $p(z) := \det(zI - A)$ be the *characteristic polynomial* of A. Since the complex field is algebraically closed, there exists at least one complex root of p. Let z_0 be such a root; then $0 = p(z_0) = \det(z_0 I - A)$ and hence there exists a non-zero vector e_0 such that $(z_0 I - A)e_0 = 0$, i.e., e_0 is an eigenvector of A with eigenvalue z_0. We can assume without loss of generality that $\|e_0\| = 1$.

Obviously, the subspace $\mathbb{C}e_0$ is invariant under A. Using now that $z_0 I - A$ is normal, we have $(\bar{z}_0 I - A^*)x = 0$, due to Exercise 4.165; in particular, $\mathbb{C}e_0$ is also invariant under A^*. Thus, by Exercise 4.158, the orthocomplement $\{e_0\}^\perp$ is also invariant under both A and A^*. The decomposition $\mathcal{H} = \mathbb{C}e_0 \oplus \{e_0\}^\perp$ yields the decomposition of A as

$$A = \begin{bmatrix} z_0 & 0 \\ 0 & \tilde{A} \end{bmatrix},$$

where $\tilde{A}$ is the restriction of A onto $\{e_0\}^\perp$. One can easily see that $\tilde{A}$ is again normal and hence, by the induction hypothesis, there exists an orthonormal basis in $\{e_0\}^\perp$ that consists of the eigenvectors of $\tilde{A}$; let them be $e_1, \dots, e_d$. Thus, $\{e_0, e_1, \dots, e_d\}$ is the required basis of $\mathcal{H}$. □

Corollary 4.237. Let A be a normal operator on a finite-dimensional complex Hilbert space $\mathcal{H}$. Then A can be written in the form

$$A = \sum_{k=1}^{\dim \mathcal{H}} a_k \, |e_k\rangle\langle e_k| \,, \tag{4.43}$$

where $e_1, \dots, e_{\dim \mathcal{H}}$ are eigenvectors of A that form an orthonormal basis for $\mathcal{H}$, and $a_1, \dots, a_{\dim \mathcal{H}}$ are the corresponding eigenvalues.

Proof. By Theorem 4.236, there exists an orthonormal basis of $\mathcal{H}$ that consists of eigenvectors of A; let us denote it by $e_1, \dots, e_d$, and the corresponding eigenvalues by $a_1, \dots, a_d$. We can expand any vector $x \in \mathcal{H}$ as $x = \sum_{k=1}^d \langle e_k, x\rangle \, e_k$, and thus

$$Ax = \sum_{k=1}^d \langle e_k, x\rangle \underbrace{Ae_k}_{=a_k e_k} = \sum_{k=1}^d a_k e_k \, \langle e, x\rangle = \left(\sum_{k=1}^d a_k \, |e_k\rangle\langle e_k| \right) x.$$

Since this holds for every $x \in \mathcal{H}$, we have the equality $A = \sum_{k=1}^d a_k \, |e_k\rangle\langle e_k|$. □

Definition 4.238. A decomposition as in (4.43) is called a *eigen-decomposition* of A.

Remark 4.239. It is important to note that Theorem 4.236 is only valid in complex Hilbert spaces. Indeed, it is easy to see that

$$A_\vartheta := \begin{bmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{bmatrix},$$

where ϑ is a real number, does not have any real eigen-values unless $\vartheta = k\pi$ for some $k \in \mathbb{Z}$.

Hence, for the rest of this section we assume that **the Hilbert space is complex**, unless otherwise stated.

Remark 4.240. An eigen-decomposition may not be unique; indeed, we have $I = \sum_{i=1}^{d} |e_i\rangle\langle e_i|$ for any ONB $(e_i)_{i=1}^{d}$. More generally, if an eigenvalue $a \in \mathrm{spec}(A)$ has a multiplicity at least two, then we have an infinite freedom of choosing an ONB in the corresponding eigen-subspace, and hence infinitely many different decompositions of the form (4.43). Note also that a further, trivial ambiguity arises from the fact that if e_i is an element of an orthonormal eigen-basis of A then it can be replaced with $\tilde{e}_i := \lambda e_i$, where $|\lambda| = 1$, and $|e_i\rangle\langle e_i| = |\tilde{e}_i\rangle\langle\tilde{e}_i|$.

The ambiguities pointed out in Remark 4.240 can be removed by using the spectral decomposition instead of an eigen-decomposition. Consider any eigen-decomposition of A as in (4.43), and for any $a \in \mathbb{C}$, let

$$P^A(a) := \sum_{i:\, a_i = a} |e_i\rangle\langle e_i| \,.$$

If a is an eigen-value, then this is nothing else than the projection onto the eigen-subspace corresponding to a (see Proposition 4.104). If a is not an eigen-value then, by the definition of an empty sum, this is the zero operator, which is also a projection, onto the zero-dimensional subspace $\{0\}$. With this notation, the eigen-decomposition (4.43) can be rewritten as

$$\begin{aligned} A = \sum_{i=1}^{d} a_i\, |a_i\rangle\langle a_i| &= \sum_{a \in \mathrm{spec}(A)} a \underbrace{\sum_{i:\, a_i = a} |e_i\rangle\langle e_i|}_{= P^A(a)} \\ &= \sum_{a \in \mathrm{spec}(A)} a P^A(a). \end{aligned} \tag{4.44}$$

Definition 4.241. The expression in (4.44) is called the *spectral decomposition* of A.

Remark 4.242. Note that (4.44) can also be written as

$$A = \sum_{a \in \mathbb{C}} a P^A(a),$$

since $P^A(a) = 0$ when $a \notin \mathrm{spec}(A)$.

Definition 4.243. For a normal operator $A \in \mathcal{B}(\mathcal{H})$, and any $H \subseteq \mathbb{C}$, let

$$P^A(H) := \sum_{a \in H} P^A(a). \tag{4.45}$$

This is called the *spectral projection* of A corresponding to the set H.

Remark 4.244. Note that the sum in (4.45) is well-defined, as at most finitely many $P^A(a) \neq 0$. That $P^A(H)$ is indeed a projection follows by Exercise 4.197. Indeed, $P^A(H)$ is the projection onto the subspace spanned by the eigenvectors corresponding to all the eigenvalues of A that fall in H. Finally, note that we have

$$P^A(\{a\}) = P^A(a), \qquad a \in \mathbb{C}.$$

Example 4.245. Recall that the support of a normal operator $A \in \mathcal{B}(\mathcal{H})$ is defined as $\operatorname{supp} A := \operatorname{ran} A = (\ker A)^\perp$. Using the above notations, the projection $P_{\operatorname{supp} A}$ onto the support of A is given by

$$P_{\operatorname{supp} A} = \sum_{a \in \operatorname{spec}(A) \setminus \{0\}} P^A(a) = P^A(\operatorname{spec}(A) \setminus \{0\}) = P^A(\mathbb{C} \setminus \{0\}).$$

Restated slightly differently, the spectral decomposition shows that any normal operator is the linear combination of pairwise orthogonal projections. It is easy to see that the converse is also true:

Exercise 4.246. (i) Show that $A \in \mathcal{B}(\mathcal{H})$ is normal if and only if there exist pairwise orthogonal projections $(P_i)_{i=1}^r$ such that

$$\sum_{i=1}^r P_i = I, \qquad \text{and} \qquad A = \sum_{i=1}^r a_i P_i \tag{4.46}$$

with some numbers $a_1, \ldots, a_r \in \mathbb{C}$.

(ii) Show that for any decomposition of A as in (4.46), $\operatorname{spec}(A) = \{a_i\}_{i \in [r]}$, and

$$P^A(a) = \sum_{i:\, a_i = a} P_i, \qquad a \in \mathbb{C}.$$

Exercise 4.247. Show that $P^A(.)$ is σ-additive, i.e., for any countable collection $\{B_i\}_i$ of pairwise disjoint subsets of $\mathbb{C}$, we have

$$P^A\left(\cup_i B_i\right) = \sum_i P^A(B_i).$$

The above example shows that $P^A(.)$ is a so-called projective valued measure (PVM), that we call the *spectral PVM of* A. See Section **??** for more details.

Definition 4.248. We define the *support* of P^A, denoted as $\operatorname{supp} P^A$, as the smallest closed subset $H \subseteq \mathbb{C}$ such that $P^A(H) = I$.

Exercise 4.249. Show that

$$\operatorname{supp} P^A = \{\lambda \in \mathbb{C} : \ P^A(\{\lambda\}) \neq 0\} = \operatorname{spec}(A).$$

Exercise 4.250. Let $\mathcal{H}$ be a finite-dimensional Hilbert space, and $A \in \mathcal{B}(\mathcal{H})$. Show that

(i) A is self-adjoint if and only if A is normal and $\operatorname{spec}(A) = \operatorname{supp} P^A \subseteq \mathbb{R}$.

(ii) A is positive if and only if A is normal and $\operatorname{spec}(A) = \operatorname{supp} P^A \subseteq \mathbb{R}_+$.

(iii) A is a projection if and only if A is normal and $\operatorname{spec}(A) = \operatorname{supp} P^A \subseteq \{0, 1\}$.

(iv) A is unitary if and only if A is normal and $\operatorname{spec}(A) = \operatorname{supp} P^A$ is a subset of the complex unit circle.

Exercise 4.251. Show that for a normal operator $A \in \mathcal{B}(\mathcal{H})$,

$$\|A\| = \max\{|a| : \ a \in \operatorname{spec}(A)\} = \max\{|\langle x, Ax\rangle| : \ x \in \mathcal{H}, \ \|x\| = 1\}.$$

Show that if A is self-adjoint then

$$\lambda_{\min}(A) = \min\{\langle x, Ax\rangle : \ x \in \mathcal{H}, \ \|x\| = 1\},$$
$$\lambda_{\max}(A) = \max\{\langle x, Ax\rangle : \ x \in \mathcal{H}, \ \|x\| = 1\},$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimal and the maximal eigen-values of A, respectively.

Score: 4+4=8 points.

Definition 4.252. For an operator $A \in \mathcal{B}(\mathcal{H})$,

$$W(A) := \{\langle \psi, A\psi\rangle : \ \psi \in \mathcal{H}, \ \|\psi\| = 1\}$$

is called the *numerical range* of A.

Exercise 4.253. (i) Show that for an arbitrary operator $A \in \mathcal{B}(\mathcal{H})$,

$$\operatorname{spec}(A) \subseteq W(A).$$

(ii) Show that for a normal operator $A \in \mathcal{B}(\mathcal{H})$,

$$W(A) \subseteq \operatorname{conv}(\operatorname{spec}(A)).$$

(iii) Use the Hausdorff-Toeplitz theorem below to conclude that for a normal operator $A \in \mathcal{B}(\mathcal{H})$,

$$W(A) = \text{conv}(\text{spec}(A)).$$

Solution: Hidden.

Theorem 4.254. (Hausdorff-Toeplitz) The numerical range of any operator $A \in \mathcal{B}(\mathcal{H})$ is convex.

Exercise 4.255. Show that the Gram matrix $G(\{u_i\}_{i=1}^r) \in \mathbb{K}^{r\times r}$ and the operator $\sum_{i=1}^r |u_i\rangle\langle u_i| \in \mathcal{B}(\mathcal{H})$ have the same strictly positive eigenvalues, counted with multiplicities. (Hint: consider the vector $\psi := \sum_{i=1}^r e_i \otimes u_i \in \mathbb{C}^r \otimes \mathcal{H}$, and take the partial traces of $|\psi\rangle\langle\psi|$.)

Solution: Hidden.

4.24 Functional calculus

Any operator $A \in \mathcal{B}(\mathcal{H})$ can be substituted into a polynomial $p = c_0 + \sum_{k=1}^n c_k \,\text{id}^k$ as

$$p(A) := c_0 I + \sum_{k=1}^n c_k A^k.$$

This is called the *polynomial functional calculus*. It has some nice algebraic properties, e.g., linearity and multiplicativity:

$$(\lambda p + \eta q)(A) = \lambda p(A) + \eta q(A), \qquad (pq)(A) = p(A)q(A).$$

It can be extended to the larger class of functions that are analytic on the spectrum of A; for instance, the exponential of any operator A can be defined as $e^A := I + \sum_{n=1}^{+\infty} A^n/n!$.

For normal operators, a more general concept of functional calculus can be defined with the help of the spectral decompositoin, in which we can take $f(A)$ for any function f that is defined at least on the spectrum of A.

Definition 4.256. Let A be a normal operator with spectral PVM P^A. For any function $f : \mathcal{D} \to \mathbb{C}$ such that $\text{spec}(A) \subseteq \mathcal{D} \subseteq \mathbb{C}$, we define $f(A)$ as

$$f(A) := \sum_{a\in\text{spec}(A)} f(a)P^A(a).$$

Exercise 4.257. (Algebraic properties of the functional calculus)
Let $A \in \mathcal{B}(\mathcal{H})$ be a normal operator. Show the following properties of the functional calculus as a map $\mathbb{C}^{\mathrm{spec}(A)} \to \mathcal{B}(\mathcal{H})$. Below $f, g \in \mathbb{C}^{\mathrm{spec}(A)}$, and $\lambda, \eta \in \mathbb{C}$.

(i) Linearity:

$$(\lambda f + \eta g)(A) = \lambda f(A) + \eta g(A).$$

(ii) Multiplicativity:

$$(fg)(A) = f(A)g(A).$$

(iii) Adjoint-preserving:

$$\overline{f}(A) = f(A)^*,$$

where $\overline{f}$ is the pointwise conjugate of f.

(iv) Unitality:

$$\mathbf{1}(A) = I,$$

where $\mathbf{1}$ is the constant 1 function.

(v) Show that $\mathrm{id}(A) = A$, and for any polynomial $p(z) = c_0 + \sum_{k=1}^n c_k z^k$, we have

$$p(A) = c_0 I + \sum_{k=1}^n c_k A^k,$$

where A^k is defined as the k-fold product $A \cdot \ldots \cdot A$. That is, we get an extension of the polynomial functional calculus.

(vi) Show that

$$\|f(A)\| = \max_{a \in \mathrm{spec}(A)} |f(a)| =: \|f\|_\infty .$$

Conclude that if a sequence f_k, $k \in \mathbb{N}$, converges to f in supremum norm on the spectrum of A then $f_k(A)$ converges to $f(A)$ in the operator norm (and hence in any other norm on $\mathrm{Lin}(\mathcal{H})$.

Exercise 4.258. Let $A \in \mathcal{B}(\mathcal{H})$ be a normal operator. Show that for any function f defined on some subset containing $\mathrm{spec}(A)$, there exists a polynomial $p(z) = \sum_{k=0}^n c_k z_k$ of degree at most $|\mathrm{spec}(A)| - 1$ such that $f(A) = p(A)$.
(Hint: Consider the Lagrange interpolation polynomials.)

Exercise 4.259. (Isometric invariance of the functional calculus)
Let $A \in \mathcal{B}(\mathcal{H})$ be a normal operator, and $f : \mathcal{D} \to \mathbb{C}$ be a function such that $\operatorname{spec}(A) \cup \{0\} \subseteq \mathcal{D} \subseteq \mathbb{C}$. Show that for any partial isometry $V : \mathcal{H} \to \mathcal{K}$ such that $\operatorname{ran} A \subseteq (\ker V)^\perp$, we have

$$f(VAV^*) = Vf(A)V^* + f(0)(I - VV^*).$$

Conclude that if $f(0) = 0$ or V is surjective then $f(VAV^*) = Vf(A)V^*$.

Solution: Hidden.

Exercise 4.260. (Spectral mapping theorem)
Let A be a normal operator, and f be a complex-valued function defined on $\operatorname{spec}(A)$. Show that $f(A)$ is normal, with spectrum and spectral PVM

$$\operatorname{spec}(f(A)) = f(\operatorname{spec} A), \qquad P^{f(A)}(B) = P^A\left(\{\lambda \in \mathbb{C} : \ f(\lambda) \in B\}\right), \quad B \subseteq \mathbb{C}.$$

Conclude that

(i) $f(A)$ is self-adjoint $\iff f(\operatorname{spec} A) \subseteq \mathbb{R}$;

(ii) $f(A)$ is positive $\iff f(\operatorname{spec} A) \subseteq \mathbb{R}_+$;

(iii) $f(A)$ is a projection $\iff f(\operatorname{spec} A) \subseteq \{0, 1\}$;

(iv) $f(A)$ is a unitary $\iff f(\operatorname{spec} A)$ is a subset of the complex unit circle.

We have defined the functions of a normal operator using its spectral decomposition. It is easy that the spectral projections can also be obtained by functional calculus.

Exercise 4.261. Let $A \in \mathcal{B}(\mathcal{H})$ be a normal operator, with spectral PVM P^A. For every $M \subseteq \mathbb{C}$, let $\mathbf{1}_M$ denote the indicator function of M, i.e.,

$$\mathbf{1}_M(z) := \begin{cases} 1, & z \in M, \\ 0, & z \notin M. \end{cases}$$

Show that for every $M \subseteq \mathbb{C}$,

$$\mathbf{1}_M(A) = \sum_{a \in M} P(a_i) = P^A(M)$$

is the spectral projection of A corresponding to M.

Exercise 4.262. Let $A \in \mathrm{Lin}(\mathcal{H})$ be a normal operator. Show that the projection P onto the support of A is its spectral projection corresponding to $\mathbb{C} \setminus \{0\}$, i.e.,

$$P = \mathbf{1}_{\mathbb{C}\setminus\{0\}}(A) = \sum_{a \in \mathrm{spec}(a):\, a \neq 0} P^A(a).$$

Exercise 4.263. Let $A, B \in \mathrm{Lin}(\mathcal{H})$ be normal operators. Show that the following are equivalent:

(i) $AB = BA$.

(ii) $p(A)q(B) = q(B)p(A)$ for any polynomials p, q.

(iii) $f(A)g(B) = g(B)f(A)$ for any functions f, g defined on $\mathrm{spec}(A)$ and $\mathrm{spec}(B)$, respectively.

(iv) $P^A(X)P^B(Y) = P^B(Y)P^A(X)$ for any $X, Y \subseteq \mathbb{C}$.

(v) $P^A(a)P^B(b) = P^B(b)P^A(a)$ for any $a \in \mathrm{spec}(A)$, $b \in \mathrm{spec}(B)$.

(vi) There exists an orthonormal basis in which the matrices of both A and B are diagonal.

(Hint: Use Exercise 4.198, and prove that $\mathcal{H} = \oplus_{a \in \mathrm{spec}(A),\, b \in \mathrm{spec}(B)} \mathrm{ran}(P^A(a)P^B(b))$.)

Score: 12 points.

When $A \in \mathcal{B}(\mathcal{H})$ is self-adjoint, we introduce the notation

$$\{A \geq c\} := \mathbf{1}_{[c,+\infty)}(A) = P^A\left([c, +\infty)\right)$$

for the spectral projection of A corresponding to the set $[c, +\infty)$. We introduce the notations

$$\{A > c\}, \qquad \{A \leq c\}, \qquad \{a \leq A \leq b\}, \qquad \text{etc.}$$

by an obvious modification of the above definition.

Exercise 4.264. Show that for any self-adjoint $A \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ and any $a, b \in \mathbb{R}$,

$$a\{a \leq A \leq b\} \leq \{a \leq A \leq b\}A = A\{a \leq A \leq b\} \leq b\{a \leq A \leq b\}.$$

Score: 3 points.

For a self-adjoint A, the projections

$$\{A > 0\} := \sum_{a \in \mathrm{spec}(A):\, a > 0} P^A(a), \qquad \{A < 0\} := \sum_{a \in \mathrm{spec}(A):\, a < 0} P^A(a)$$

are of particular interest. These the projections onto the subspace spanned by the eigenvectors corresponding to the stricly positive/strictly negative eigenvalues of A, respectively.

Definition 4.265. For a self-adjoint operator $A \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$, let

$$A_+ := \sum_{a \in \mathrm{spec}(A):\, a>0} a P^A(a), \qquad A_- := \sum_{a \in \mathrm{spec}(A):\, a<0} (-a) P^A(a),$$

be the positive and the negative part of A, respectively.

Exercise 4.266. Let $A \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ be self-adjoint. Show that

$$A_+ = A\{A \geq 0\} = \mathrm{id}_+(A) \geq 0$$
$$A_- = -A\{A \leq 0\} = \mathrm{id}_-(A) \geq 0,$$

where $\mathrm{id}_+(t) := \max\{t, 0\}$, $\mathrm{id}_-(t) := -\min\{t, 0\}$, $t \in \mathbb{R}$. Show also that

$$A_+ A_- = 0, \qquad A = A_+ - A_-, \qquad |A| = A_+ + A_-,$$

and conclude that

$$A_+ = \frac{|A| + A}{2}, \qquad A_- = \frac{|A| - A}{2}.$$

Score: 8 points.

As we have seen in Section 4.11, any operator $A \in \mathrm{Lin}(\mathcal{H})$ can be uniquely decomposed as $A = A_1 + iA_2$, where A_1 and A_2 are self-adjoint; indeed, we have

$$A = \frac{A + A^*}{2} + i\frac{A - A^*}{2i}.$$

As we have seen in Exercise 4.266, the spectral decomposition (alternatively, the functional calculus) allows us to further decompose self-adjoint operators as the difference of two positive operators. Hence, we get the following:

Corollary 4.267. Any operator $A \in \mathrm{Lin}(\mathcal{H})$ can be decomposed as a linear combination of at most four PSD operators.

An important application of the functional calculus is that every PSD operator has a unique PSD square root. Indeed, if $A \in \mathcal{B}(\mathcal{H})_+$ then it admits a spectral decomposition $A = \sum_{a \in \mathrm{spec}(A)} a P^A(a)$, where $a \geq 0$ for all $a \in \mathrm{spec}(A)$, and hence we may define

$$\sqrt{A} := A^{1/2} := \sum_{a \in \mathrm{spec}(A)} \sqrt{a} P^A(a).$$

It is straightforward to verify that $\sqrt{A}^2 = A$.

Exercise 4.268. Let $A \in \mathcal{B}(\mathcal{H})$. Show that the following are equivalent:

(i) A is PSD.

(ii) There exists a PSD operator $B \in \mathcal{B}(\mathcal{H})$ such that $A = B^2$.

(iii) There exists a self-adjoint operator $B \in \mathcal{B}(\mathcal{H})$ such that $A = B^2$.

(iv) There exists an operator $B \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, for some Hilbert space $\mathcal{K}$, such that $A = B^*B$.

Moreover, the operator B in (ii) and in (iii) are unique, and equal to $\sqrt{A}$.

More generally, we may define arbitrary complex powers of a PSD operator A as

$$A^z := \sum_{a \in \mathrm{spec}(A),\, a>0} a^z P^A(a),$$

where $a^z := e^{z \ln a}$. Note that we only consider the strictly positive eigenvalues in the above definition, and $\ln a$ is the usual real natural logarithm of the positive number a. In particular, we have

$$A^0 = \sum_{a>0} P^A(a),$$

which is nothing but the projection onto the support of A, and

$$A^{-1} = \sum_{a>0} a^{-1} P^A(a)$$

is the *generalized inverse*, which satisfies

$$A^{-1}A = AA^{-1} = A^0.$$

Exercise 4.269. Let $A \in \mathcal{B}(\mathcal{H})_{\geq 0}$ be a PSD operator. Show that for any $z, w \in \mathbb{C}$,

$$A^{z+w} = A^z A^w.$$

Show that

$$A^0 = \lim_{t \searrow 0} A^t.$$

Remark 4.270. Note that we may define the absolute value of any normal operator A via functional calculus as

$$|A| = \sum_{a \in \operatorname{spec}(A)} |a| P^A(a).$$

For a general (not necessarily normal) operator this is not possible, as A does not admit a spectral decomposition. However, we may still use functional calculus (more precisely, the existence of a PSD square root) to define the absolute value of an arbitrary operator A as $|A| := \sqrt{A^* A}$. We come back to this important concept in Section 4.31.

Lemma 4.271. Let $A \in \mathcal{B}(\mathcal{H})_{\text{sa}}$ be such that $0 \leq T \leq \beta I$ for some real number β, and let $\mathcal{V} \subseteq \mathcal{H}$. Then we have the following equivalences:

$$\begin{aligned}
&\langle v, Tv \rangle = \beta \|v\|^2 \quad \forall v \in \operatorname{span}(\mathcal{V}) && (4.47)\\
\iff\ &\langle v, Tv \rangle = \beta \|v\|^2 \quad \forall v \in \mathcal{V} && (4.48)\\
\iff\ &Tv = \beta v \quad \forall v \in \operatorname{span}(\mathcal{V}) && (4.49)\\
\iff\ &P_{\operatorname{span}(\mathcal{V})} \leq P^T(\beta) && (4.50)\\
\iff\ &\beta P_{\operatorname{span}(\mathcal{V})} \leq T, && (4.51)
\end{aligned}$$

and

$$\begin{aligned}
&\langle v, Tv \rangle = 0 \quad \forall v \in \mathcal{V} && (4.52)\\
\iff\ &Tv = 0 \quad \forall v \in \operatorname{span}(\mathcal{V}) && (4.53)\\
\iff\ &T^0 \leq I - P_{\operatorname{span}(V)} && (4.54)\\
\iff\ &T \leq \beta(I - P_{\operatorname{span}(V)}). && (4.55)
\end{aligned}$$

Proof. The implication (4.47)$\Longrightarrow$(4.48) is obvious. Assume (4.48); then

$$\begin{aligned}
&\langle v, Tv \rangle = \beta \|v\|^2 = \langle v, \beta I v \rangle \quad \forall v \in \mathcal{V}\\
\iff\ &0 = \langle v, (\beta I - T)v \rangle = \left\| (\beta I - T)^{1/2} v \right\|^2 \quad \forall v \in \mathcal{V}\\
\iff\ &(\beta I - T)v = 0 \ \ \forall v \in \mathcal{V}, && (4.56)
\end{aligned}$$

which yields (4.49) by linearity. The equivalence (4.49)$\Longleftrightarrow$(4.50) is obvious, and (4.50)$\Longrightarrow$(4.51) follows by $\beta P^T(\beta) \leq \beta P^T(\beta) + \sum_{0<t<\beta} t P^T(t) = T$. Finally, assume (4.51); then for all $v \in \mathcal{V}$, $\beta \|v\|^2 = \left\langle v, \beta P_{\operatorname{span}(\mathcal{V})} v \right\rangle \leq \langle v, Tv \rangle \leq \langle v, \beta I v \rangle = \beta \|v\|^2$, i.e., (4.47) holds.

For the second set of equivalences, we have $\langle v, Tv \rangle = \left\| T^{1/2} v \right\|^2$, and hence $T^{1/2} v = 0 \iff Tv = 0$, which yields (4.52)$\Longleftrightarrow$(4.53). Moreover, (4.53) is equivalent to $\operatorname{ran} P_{\operatorname{span}(\mathcal{V})} \subseteq \ker T \iff P_{\operatorname{span}(V)} \leq (I - T^0) \iff I - P_{\operatorname{span}(\mathcal{V})} \geq T^0$, which is

(4.54). If (4.54) holds then $T \leq \beta T^0 \leq \beta(I - P_{\mathrm{span}(\mathcal{V})})$ follows immediately, while the latter condition yields $0 \leq \langle v, Tv\rangle \leq \beta \left\langle v, (I - P_{\mathrm{span}(\mathcal{V})})v\right\rangle = 0$ for all $v \in \mathrm{span}(\mathcal{V})$, showing the implications (4.54)$\Longrightarrow$(4.55)$\Longrightarrow$(4.52). □

Lemma 4.272. Let $A, T \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ such that $0 \leq T \leq I$. Then

$$-\operatorname{Tr} A_- \leq \operatorname{Tr} AT \leq \operatorname{Tr} A_+,$$

and the first inequality holds with equality if and only if $\{A < 0\} \leq T \leq \{A \leq 0\}$, and the second inequality holds with equality if and only if $\{A > 0\} \leq T \leq \{A \geq 0\}$.

Proof. Let $A = \sum_{i=1}^d a_i \, |e_i\rangle\langle e_i|$ be an eigen-decomposition of A. Then

$$\operatorname{Tr} AT = \sum_{i\colon a_i > 0} a_i \, \langle e_i, Te_i\rangle - \sum_{i\colon a_i < 0} (-a_i) \, \langle e_i, Te_i\rangle \leq \sum_{i\colon a_i > 0} a_i = \operatorname{Tr} A_+. \tag{4.57}$$

The rest of the statement follows immediately using Lemma **??** with $\beta = 1$. □

Corollary 4.273. For a self-adjoint operator $A \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr} A_- = \min\{\operatorname{Tr} AT : \, 0 \leq T \leq I\} = \operatorname{Tr} A_+ \leq \max\{\operatorname{Tr} AT : \, 0 \leq X \leq I\} = \operatorname{Tr} A_+,$$

Exercise 4.274. Consider the operators $A_1, A_2 \in \mathrm{Lin}(\mathbb{C}^2)$, which are given by their matrices in the standard orthonormal basis of $\mathbb{C}^2$ as

$$A_1 := \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}, \qquad A_2 := \begin{bmatrix} -2 & -i \\ i & -2 \end{bmatrix}.$$

For $k = 1, 2$, compute

$$(A_k)_+, \quad (A_k)_-, \quad |A_k|, \quad \sin(A_k), \quad e^{A_k}, \quad A^{2015}.$$

Solution: Hidden.

Exercise 4.275. (Continuity of the functional calculus)
Let $A, B \in \mathrm{Lin}(\mathcal{H})$. Show that

(i) For every $n \in \mathbb{N} \setminus \{0\}$,

$$\begin{aligned} A^{n+1} - B^{n+1} = \frac{1}{2}\big[&(A - B)(A^n + B^n) + (A^n + B^n)(A - B) \\ &+ A(A^{n-1} - B^{n-1})B + B(A^{n-1} - B^{n-1})A\big], \end{aligned} \tag{4.58}$$

with the convention $A^0 := B^0 := I$.

(ii) Let $\|.\|$ be any norm on $\mathrm{Lin}(\mathcal{H})$ that is submultiplicative, i.e., $\|XY\| \leq \|X\| \, \|Y\|$, $X, Y \in \mathrm{Lin}(\mathcal{H})$. Use mathematical induction and (4.58) to prove

$$\|A^n - B^n\| \leq \|A - B\| \sum_{k=0}^{n-1} \|A\|^k \, \|B\|^{n-1-k} \tag{4.59}$$

for every $n \in \mathbb{N} \setminus \{0\}$. Note that if $\|A\| \neq \|B\|$ then the above formula can be rewritten as

$$\frac{\|A^n - B^n\|}{\|A - B\|} \leq \frac{\|A\|^n - \|B\|^n}{\|A\| - \|B\|}.$$

(iii) Show that if $\lim_{k\to\infty} \|A_k - A\| = 0$ then $\lim_{k\to\infty} A_k^n = A^n$ for any fixed power $n \in \mathbb{N}$.

(iv) Let A and A_k, $k \in \mathbb{N}$, be self-adjoint operators such that $\lim_{k\to\infty} \|A_k - A\| = 0$. Let f be a continuous function on $[-\|A\| - 1, \|A\| + 1]$. Show that

$$\lim_{k\to\infty} \|f(A_k) - f(A)\| = 0.$$

Solution: Hidden.

Exercise 4.276. Show that if $M \in \mathcal{B}(\mathcal{H})$ is such that $0 \leq M \leq I$ then $M^2 \leq M$.

Solution: Hidden.

Exercise 4.277. (Every PSD matrix is a Gram matrix) Let $A \in \mathbb{C}^{d\times d}$ be PSD. Show that in any Hilbert space $\mathcal{K}$ with $\dim \mathcal{K} \geq \mathrm{rk}\, A$, there exist vectors $v_1, \ldots, v_d$ such that $A = G(\{v_i\}_{i=1}^d)$.

Solution: Hidden.

Exercise 4.278. (i) Let $A \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ be a self-adjoint operator. Show that $I - iA$ is invertible, and

$$U := \frac{I + iA}{I - iA} := \kappa(A) \tag{4.60}$$

is a unitary operator such that $-1 \notin \mathrm{spec}(U)$. Show that if $\|A\| < 1$ then

$$\kappa(A) = I + 2\sum_{n=1}^{+\infty} (iA)^n, \tag{4.61}$$

and all eigenvalues of $\kappa(A)$ have positive real parts.

(ii) Let $U \in \mathcal{B}(\mathcal{H})$ be a unitary operator such that $-1 \notin \operatorname{spec}(U)$, and define

$$A := i\frac{I-U}{I+U} := \kappa^{-1}(U). \tag{4.62}$$

Show that A is self-adjoint, and $\|A\| < 1$ if and only if $\mathfrak{Re}\, u > 0$ for all $u \in \operatorname{spec}(U)$.

(iii) Show that for a unitary operator U, $\mathfrak{Re}\, u > 0$ for all $u \in \operatorname{spec}(U)$ if and only if $\|I-U\| < \sqrt{2}$.

(iv) Show that the maps in (4.60) and (4.62) are inverses of each other, and they give a bijection between the set of self-adjoint operators $\mathcal{B}(\mathcal{H})_{\text{sa}}$ and the set of unitary operators on $\mathcal{H}$ with -1 not in their spectra.

(v) Show that any unitary operator U such that $\|I-U\| < \sqrt{2}$ can be expanded as

$$U = I + 2\sum_{n=1}^{+\infty} \left(i\kappa^{-1}(U)\right)^n.$$

Solution: Hidden.

Remark 4.279. In functional analysis, the *Cayley transform* of a self-adjoint operator A is defined as

$$\tilde{\kappa}(A) := \frac{A-iI}{A+iI} = -\frac{I+iA}{I-iA} = -\kappa(A),$$

i.e., $\kappa(A)$ defined in (4.60) is minus one times the Cayley transform of A. The inverse of the Cayley transformation $\tilde{\kappa}$ is given by

$$\tilde{\kappa}^{-1}(U) = i\frac{I+U}{I-U} = \left(-i\frac{I-U}{I+U}\right)^{-1} = -\left(\kappa^{-1}(U)\right)^{-1},$$

where the first identity works when $U \in \operatorname{ran}\tilde{\kappa}$, i.e., $1 \notin \operatorname{spec}(U)$, and for the second identity to hold, one has to assume that $-1 \notin \operatorname{spec}(U)$, i.e., $U \in \operatorname{ran}\kappa$. The advantage of κ to the Cayley transform is that the former yields a series expansion of unitaries close enough to I, where the latter close enough to $-I$.

4.25 Trace-class operators

Recall that for a linear operator on a finite-dimensional Hilbert space, its trace was defined as the sum of its diagonal elements in an orthonormal basis, and this quantity was independent of the basis chosen. If we want to define the trace of a bounded

operator on an infinite-dimensional Hilbert space in the same way then we run into some problems. Indeed, let $\mathcal{H}$ be separable, $\{e_n\}_{n\in\mathbb{N}}$ be an ONB in $\mathcal{H}$, and let A be diagonal in this ONB with $Ae_n = \frac{(-1)^n}{n} e_n$. Then the series $\sum_{n=1}^{+\infty} \langle e_n, Ae_n\rangle$ is convergent, but not absolutely convergent, and hence, by the Riemann series theorem, for any $c \in \overline{\mathbb{R}}$, there exists a bijection $\kappa : \mathbb{N} \to \mathbb{N}$ such that $\sum_{n=1}^{+\infty} \langle e_{\kappa(n)}, Ae_{\kappa(n)}\rangle = c$, and one can also find a κ such that the previous sum does not converge. Hence, whether or not the sum $\sum_{n=1}^{+\infty} \langle e_n, Ae_n\rangle$ converges for a given operator $A \in \mathcal{B}(\mathcal{H})$, and its value, may depend on the orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$. Our goal in this section is to characterize those operators for which this is not the case, and hence their trace can be unambigously defined. Our first observation shows that this is the case for any positive semi-definite operator, if we allow the trace to be $+\infty$.

Proposition 4.280. Let $\mathcal{H}$ be separable Hilbert space and $A \in \mathcal{B}(\mathcal{H})$ be positive semi-definite. Then for any two ONBs $\{e_n\}_{n\in\mathbb{N}}$ and $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$,

$$\operatorname{Tr} A := \sum_{n=1}^{+\infty} \langle e_n, Ae_n\rangle = \sum_{n=1}^{+\infty} \langle f_n, Af_n\rangle .$$

The quantity $\operatorname{Tr} A \in [0, +\infty]$ is the *trace* of A.

Proof. □

Definition 4.281. For any $A \in \mathcal{B}(\mathcal{H})$, let

$$\|A\|_1 := \operatorname{Tr} |A|$$

be its *trace-norm*. We say that $A \in \mathcal{B}(\mathcal{H})$ is *trace-class* if $\|A\|_1 = \operatorname{Tr}|A| < +\infty$. We denote the set of trace-class operators on $\mathcal{H}$ by $\mathcal{S}_1(\mathcal{H})$.

Next we show that $\|\ \|_1$ is a norm on the set of trace-class operators. To make it more easily distinguishable (and for other reasons that will become clear later) we will denote the usual operator norm of an $A \in \mathcal{B}(\mathcal{H})$ by $\|A\|_\infty$.

Lemma 4.282. For any $A, B \in \mathcal{B}(\mathcal{H})$ and $\lambda \in \mathbb{C}$, we have

(i) $\|\lambda A\|_1 = |\lambda| \, \|A\|_1$.

(ii) $\|A\|_1 \geq \|A\|_\infty \geq 0$, and $\|A\|_1 = 0 \iff A = 0$.

(iii) $\|A + B\|_1 \leq \|A\|_1 + \|B\|_1$.

In particular, $\|\ \|_1$ is a norm on $\mathcal{S}_1(\mathcal{H})$. Moreover,

(iv) $\|\ \|_1$ is *unitarily invariant*, i.e., for any unitaries U, V, $\|UAV\|_1 = \|A\|_1$.

(v) $\|A\|_1 = \|A^*\|_1$.

(vi) $\|AB\|_1 \leq \|A\|_\infty \|B\|_1$, $\|AB\|_1 \leq \|A\|_1 \|B\|_{+\infty}$.

Proof. The first two assertion is trivial. For the second, note first that $\||A|\|^2 = \||A|^2\| = \|A^*A\| = \|A\|^2$. For every $\varepsilon > 0$, let $x_\varepsilon \in \mathcal{H}$ be a unit vector such that

$$\langle x_\varepsilon, |A| x_\varepsilon \rangle > -\varepsilon + \sup\{\langle x, |A| x \rangle : \|x\| = 1\} = \||A|\| - \varepsilon = \|A\| - \varepsilon.$$

Then $\|A\| - \varepsilon \leq \operatorname{Tr} |A|$, from which the assertion follows.

The triangle inequality follows by polar decomposition (see Reed-Simon). □

Note that A is trace-class if and only if $|A|$ is trace-class, by definition. Let $P^{|A|}$ be the spectral PVM of $|A|$. If there existed an $\varepsilon > 0$ such that $\operatorname{ran} P^{|A|}([\varepsilon, +\infty))$ is infinite-dimensional then we could take an ONS $\{e_n\}_{n\in\mathbb{N}} \subseteq \operatorname{ran} P^{|A|}([\varepsilon, +\infty))$, and thus we would have

$$\operatorname{Tr} |A| \geq \sum_{n\in\mathbb{N}} \langle e_n, |A| e_n \rangle \geq \sum_{n\in\mathbb{N}} \varepsilon = +\infty,$$

a contradiction. Hence, we see that for every $\varepsilon > 0$, $\operatorname{ran} P^{|A|}([\varepsilon, +\infty))$ is finite-dimensional, and therefore we can find an ONB $\{e_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that

$$|A| = \sum_{n\in\mathbb{N}} a_n |e_n\rangle\langle e_n|, \qquad \text{where} \qquad \sum_{n\in\mathbb{N}} a_n = \operatorname{Tr} |A| < +\infty,$$

where the convergence of the first sum is in $\| \ \|_1$.

Lemma 4.283. For any $A \in \mathcal{B}(\mathcal{H})$ and any ONB $\{e_n\}_{n\in\mathbb{N}}$,

$$\sum_{n\in\mathbb{N}} |\langle e_n, A e_n \rangle| \leq \operatorname{Tr} |A|.$$

Proof. Let $A = U|A|$ be the polar decomposition of A. Then

$$\begin{aligned} |\langle e_n, A e_n \rangle| &= |\langle e_n, U|A| e_n \rangle| = |\langle |A|^{1/2} U^* e_n, |A|^{1/2} e_n \rangle| \leq \||A|^{1/2} U^* e_n\| \, \||A|^{1/2} e_n\| \\ &= \langle |A|^{1/2} U^* e_n, |A|^{1/2} U^* e_n \rangle^{1/2} \langle |A|^{1/2} e_n, |A|^{1/2} e_n \rangle \\ &= \langle e_n, U|A|U^* e_n \rangle^{1/2} \langle e_n, |A| e_n \rangle^{1/2}. \end{aligned}$$

Thus, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \sum_{n\in\mathbb{N}} |\langle e_n, A e_n \rangle| &\leq \left(\sum_{n\in\mathbb{N}} \langle e_n, U|A|U^* e_n \rangle \right)^{1/2} \left(\sum_{n\in\mathbb{N}} \langle e_n, |A| e_n \rangle \right)^{1/2} \\ &= (\operatorname{Tr} U|A|U^*)^{1/2} (\operatorname{Tr} |A|)^{1/2} = \operatorname{Tr} |A| < +\infty. \end{aligned}$$

□

Proposition 4.284. Let $A \in \mathcal{B}(\mathcal{H})$. T.f.a.e.:

(i) For any ONB $\{e_n\}_{n\in\mathbb{N}}$, $\sum_{n=1}^{+\infty} \langle e_n, Ae_n\rangle$ is convergent, and the value of the sum is independent of the ONB.

(ii) For any ONB $\{e_n\}_{n\in\mathbb{N}}$, $\sum_{n=1}^{+\infty} \langle e_n, Ae_n\rangle$ is convergent.

(iii) For any ONB $\{e_n\}_{n\in\mathbb{N}}$, $\sum_{n=1}^{+\infty} |\langle e_n, Ae_n\rangle| < +\infty$.

(iv) A is trace-class.

(v) A can be approximated in trace-norm by finite-rank operators.

(vi) There exist orthonormal systems $\{f_n\}_{n\in\mathbb{N}}$, $\{g_n\}_{n\in\mathbb{N}}$, and $(a_n)_{n\in\mathbb{N}} \subseteq [0, +\infty)$, $\sum_{n\in\mathbb{N}} a_n < +\infty$

$$A = \sum_{n=1}^{+\infty} a_n |g_n\rangle\langle f_n| ,$$

where the sum is convergent in trace-norm.

(vii) The previous statement holds with convergence in the weak operator topology.

Proof. (ii) $\Longleftrightarrow$ (iii) due to the Riemann series theorem. First, note that $|\langle e_n, (A + A^*)/2e_n\rangle| \leq |\langle e_n, Ae_n\rangle|$, and similarly, $|\langle e_n, (A - A^*)/(2i)e_n\rangle| \leq |\langle e_n, Ae_n\rangle|$. Hence, we can assume without loss of generality that A is self-adjoint.

$$\sum_{n=1}^{+\infty} |\langle e_n, Ae_n\rangle|$$

□

4.26 Uniformly convex spaces

As we have seen, the norm is a continuous function with respect to the topology induced by itself, or put more simply, for any sequence $(x_n)_{n\in\mathbb{N}}$ in a normed space,

$$\|x_n - x\| \to 0 \quad \Longrightarrow \quad \|x_n\| \to \|x\| .$$

The converse implication is obviously not true in general, but there are many important normed spaces in which a statement holds that looks at least formally like a weak converse of the above implication; see, e.g., (4.67) below. Moreover,, if the RHS above is appended by the weak convergence of the sequence then the implication does indeed become two-way. A sufficient condition for these is the so-called

uniform convexity of the space, which we will define below. All Hilbert spaces are uniformly convex, and uniformly convex Banach spaces possess many properties of Hilbert spaces, proving which is slightly different from, but not too much more difficult than in the Hilbert space case.

Definition 4.285. We say that a normed space X is *uniformly convex* (or that its norm is uniformly convex), if

$$\begin{aligned} \forall\, \varepsilon > 0 \;\; \exists\, \delta > 0 : \quad & \forall\, x, y \in X, \quad \|x\| = \|y\| = 1, \\ & \|x - y\| \geq \varepsilon \implies \left\| \frac{x+y}{2} \right\| < 1 - \delta. \end{aligned} \tag{4.63}$$

Geometrically, for unit vectors that are not too close to each other, the midpoint of their connecting line segment cannot be too far from the origin; moreover, this is quantified uniformly on the unit sphere (i.e., δ only depends on ε, and not on x and y).

Remark 4.286. For unit vectors x, y, $\|x - y\| \leq \|x\| + \|y\| \leq 2$, and hence condition 4.63 only needs to be verified for $\varepsilon \in (0, 2]$.

Remark 4.287. Note that the choice $\varepsilon := 2$ yields that

$$\|x_0\| = 1 = \|x\|\,, \quad \|x - x_0\| = 2 \implies x = -x_0, \tag{4.64}$$

i.e., the only unit vector x of distance 2 from a unit vector x_0 is its antipodal $-x_0$. In yet another geometric picture, the radius 2 sphere drawn around a unit vector x_0 only intersects the unit sphere with origin 0 at $-x_0$. The proof of (4.64) is simple:

$$\begin{aligned} \|x - x_0\| = 2 \implies & \left\| \frac{x - x_0}{2} \right\| \not< 1 - \delta \;\; \forall \delta > 0 \\ \implies & \|x + x_0\| < \varepsilon \;\; \forall \varepsilon \in (0, 2] \implies x = -x_0, \end{aligned}$$

where in the second implication we used the contrapositive of the implication in (4.63) with $y = -x_0$.

Remark 4.288. Yet another geometric picture behind uniform convexity is that the unit sphere has to be "curved", i.e., if x, y are two distinct elements of the unit sphere then the midpoint of the line segment connecting them is not on the unit sphere (but inside the open unit ball).

The above geometric pictures can used to solve the following:

Exercise 4.289. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space.

(i) Show that if there exist two disjoint sets $A, B \in \mathcal{A}$ of positive measure then $L^\infty(\mathcal{X}, \mathcal{A}, \mu)$ is not uniformly convex.

(ii) Show that if there exist two disjoint sets $A, B \in \mathcal{A}$ of finite positive measure then $L^1(\mathcal{X}, \mathcal{A}, \mu)$ is not uniformly convex.

Solution: Hidden.

Remark 4.290. It is instructive to draw a picture of the unit spheres of $l^\infty([2])$ and of $l^1([2])$, and study the relation of $(-1, 1)$, $(1, 1)$ in the first, and of $(-1, 0)$, $(0, 1)$ in the second case.

Example 4.291. By the parallelogram identity, if a norm is induced by an inner product then

$$\left\|\frac{x+y}{2}\right\| = \left(\frac{1}{2}\left\|x^2\right\| + \frac{1}{2}\|y\|^2 - \left\|\frac{x-y}{2}\right\|^2\right)^{1/2}\Bigg|_{\|x\|=\|y\|=1} = \left(1 - \frac{1}{4}\|x-y\|^2\right)^{1/2}.$$

This implies immediately that any inner product space is uniformly convex; indeed one may choose $\delta := 1 - \sqrt{1-\varepsilon^2/4}$ in (4.63).

As it turns out, the parallelogram identity can be extended to L^p spaces in the weaker form of an inequality:

Lemma 4.292. (Clarkson's inequalities)
Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and $f, g : \mathcal{X} \to \mathbb{K}$ be measurable functions on $\mathcal{X}$. Then

$$\left\|\frac{f+g}{2}\right\|_p^p + \left\|\frac{f-g}{2}\right\|_p^p \leq \frac{1}{2}\|f\|_p^p + \frac{1}{2}\|g\|_p^p, \qquad p \in [2, +\infty), \tag{4.65}$$

$$\left\|\frac{f+g}{2}\right\|_p^q + \left\|\frac{f-g}{2}\right\|_p^q \leq \left(\frac{1}{2}\|f\|_p^p + \frac{1}{2}\|g\|_p^p\right)^{1/(p-1)}, \qquad p \in (1, 2], \tag{4.66}$$

where in the second inequality, $q = p/(p-1)$ is the Hölder conjugate of p.

Proof. We only prove (4.65). Note that for any $a, b \in \mathbb{K}$,

$$\left|\frac{a+b}{2}\right|^p + \left|\frac{a-b}{2}\right|^p \leq \left(\left|\frac{a+b}{2}\right|^2 + \left|\frac{a-b}{2}\right|^2\right)^{p/2} = \left(\frac{1}{2}|a|^2 + \frac{1}{2}|b|^2\right)^{p/2} \leq \frac{1}{2}|a|^p + \frac{1}{2}|b|^p,$$

where the first inequality follows from the fact that $\mathrm{id}_{\mathbb{R}_{\geq 0}}^{p/2}$ is convex and takes the value 0 at 0, and therefore it is also super-additive. The second inequality follows from the convexity of $\mathrm{id}_{\mathbb{R}_{\geq 0}}^{p/2}$. Replacing a and b with $f(x)$ and $g(x)$, and integrating over $x \in \mathcal{X}$ yields (4.65). □

Clarkson's inequalities immediately imply the following:

Corollary 4.293. For any measure space $(\mathcal{X}, \mathcal{A}, \mu)$, and any $p \in (1, +\infty)$, $L^p(\mathcal{X}, \mathcal{A}, \mu)$ is uniformly convex.

Let us now move to the study of the general properties of uniformly convex spaces. We start with the following:

Lemma 4.294. Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be sequences of unit vectors in a uniformly convex space. Then

$$\lim_{n,m\to+\infty} \left\| \frac{x_n + y_m}{2} \right\| = 1 \quad \Longrightarrow \quad \lim_{n,m\to+\infty} \|x_n - y_m\| = 0.$$

Proof. Assume the contrary, i.e., that there exists some $\varepsilon > 0$ such that for every $N \in \mathbb{N}$ there exist $n, m \in \mathbb{N}$ for which $\|x_n - y_m\| \geq \varepsilon$. Let $\delta > 0$ be the constant corresponding to ε in the definition of uniform convexity; then for all such n, m, as above, $\left\| \frac{x_n+y_m}{2} \right\| < 1 - \delta$. This, however, contradicts the assumption that $\lim_{n,m\to+\infty} \left\| \frac{x_n+y_m}{2} \right\| = 1$. □

The following are easy consequences of Lemma 4.294:

Exercise 4.295. (i) Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be sequences in a uniformly convex space. Show that

$$\lim_{n\to+\infty} \|x_n\| = \lim_{n\to+\infty} \|y_n\| = \frac{1}{2} \lim_{n,m\to+\infty} \|x_n + y_m\| \quad \Longrightarrow \quad \lim_{n,m\to\infty} \|x_n - y_m\| = 0.$$

(ii) Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in a uniformly convex Banach space. Show that

$$\lim_{n\to+\infty} \|x_n\| = \frac{1}{2} \lim_{n,m\to+\infty} \|x_n + x_m\| \quad \Longrightarrow \quad \exists\, x : \lim_{n\to+\infty} \|x_n - x\| = 0.$$

(iii) Let x and x_n, $n \in \mathbb{N}$, be elements of a uniformly convex space. Show that

$$\|x_n\| \to \|x\|, \quad \|x_n + x\| \to \|x + x\| = 2\|x\| \quad \Longrightarrow \quad \|x_n - x\| \to 0. \tag{4.67}$$

Solution: Hidden.

Theorem 4.296. Let C be a non-empty closed convex set in a uniformly convex Banach space X. For any point $x \in X$, there exists a unique closest element to x in C.

Proof. Let $d := d(x, C) = \inf_{c \in C} \|x - c\|$. By the closedness of C, $d = 0 \iff x \in C$, and in this case x itself is the unique closest point. Hence, for the rest we assume that $d > 0$. By the definition of the infimum, there exists a sequence $(c_n)_{n \in \mathbb{N}} \subseteq C$ such that $\|x - c_n\| \to d$. Then for any $\varepsilon > 0$ and any n, m large enough,

$$d + \varepsilon \geq \frac{1}{2} \|x - c_n\| + \frac{1}{2} \|x - c_m\| \geq \left\| \frac{1}{2}(x - c_n) + \frac{1}{2}(x - c_m) \right\| = \left\| x - \frac{c_n + c_m}{2} \right\| \geq d,$$

where the first inequality is by the assumption $\|x - c_n\| \to d$, the second inequality is by the triangle inequality, and the last inequality follows from $(c_n + c_m)/2 \in C$ due to the convexity of C. Thus,

$$\begin{aligned} \lim_{n,m \to +\infty} \left\| \frac{1}{2}(x - c_n) + \frac{1}{2}(x - c_m) \right\| = d \implies 0 &= \lim_{n,m \to +\infty} \|x - c_n - (x - c_m)\| \\ &= \lim_{n,m \to +\infty} \|c_n - c_m\| , \end{aligned}$$

where the implication is due to Exercise 4.295. Thus, $(c_n)_{n \in \mathbb{N}}$ is a Cauchy sequence, and completeness of the space yields that it has a limit c. Moreover, since, C is closed, $c \in C$. Finally, $\|x - c\| = \lim_{n \to +\infty} \|x - c_n\| = d$.

Assume now that there exist $c, \tilde{c} \in C$, such that $\|x - c\| = \|x - \tilde{c}\| = d(x, C) = d$. If $c \neq \tilde{c}$ then there exists some $\varepsilon > 0$ such that $\left\| \frac{x-c}{d} - \frac{x-\tilde{c}}{d} \right\| = \frac{1}{d} \|c - \tilde{c}\| > \varepsilon$. By uniform convexity, for this $\varepsilon > 0$ there exists a $\delta > 0$ such that $\left\| \frac{1}{2} \frac{x-c}{d} + \frac{1}{2} \frac{x-\tilde{c}}{d} \right\| < 1 - \delta$, or equivalently, $\left\| x - \frac{c+\tilde{c}}{2} \right\| < (1 - \delta)d$. Since $(c + \tilde{c})/2 \in C$, this contradicts the definition of d. □

Exercise 4.297. Let X be a finite-dimensional normed space. Show that for any $x \in X$ and any closed convex set $C \subseteq X$, there exists a point $c \in C$ such that $\|x - c\| = d(x, C)$. Show an example where the closest point is not unique.

Solution: Hidden.

Another property in which uniformly convex spaces are like inner product spaces is the following characterization of weak convergence:

Proposition 4.298. Let X be a uniformly convex space, and $(x_n)_{n \in \mathbb{N}} \subseteq X$ be a sequence. Then

$$\lim_{n \to +\infty} \|x_n - x\| = 0 \iff \begin{cases} \forall\, \varphi \in X^* : \lim_{n \to +\infty} \varphi(x_n) = \varphi(x) \quad (\text{i.e., } x_n \xrightarrow{w} x), \\ \lim_{n \to +\infty} \|x_n\| = \|x\| . \end{cases}$$

Proof. The implication $\implies$ is obvious, and hence we only have to prove the converse implication. For that, we have

$$2 \|x\| = \|x + x\| \leq \liminf_{n \to +\infty} \|x_n + x\| \leq \|x\| + \lim_{n \to +\infty} \|x_n\| = 2 \|x\| ,$$

where the first inequality follows from $x_n \xrightarrow{w} x$. Hence, $\lim_{n\to+\infty} \|x_n - x\| = 0$, according to Exercise 4.295. □

Exercise 4.299. Let $\mathcal{X}$ be a set equipped with the counting measure on $\mathcal{P}(\mathcal{X})$. Show that the equivalence in Proposition 4.298 holds in $l^p(\mathcal{X})$ for every $p \in [1, +\infty)$. More precisely,

$$\lim_{n\to+\infty} \|f_n - f\|_p = 0 \iff \begin{cases} \forall\, g \in l^q(\mathcal{X}) : \lim_{n\to+\infty} \sum_{x\in\mathcal{X}} g(x) f_n(x) = \sum_{x\in\mathcal{X}} g(x) f(x), \\ \lim_{n\to+\infty} \|f_n\|_p = \|f\|_p , \end{cases}$$

$$\iff \begin{cases} \forall\, x \in \mathcal{X} : \lim_{n\to+\infty} f_n(x) = f(x), \\ \lim_{n\to+\infty} \|f_n\|_p = \|f\|_p . \end{cases}$$

(Hint: Use the Lebesgue dominated convergence theorem.)

Remark 4.300. Note that the in the above exercise, $l^1(\mathcal{X})$ is not uniformly convex, yet the equivalence still holds.

4.27 Operator algebras

Definition 4.301. Let $\mathcal{H}$ be a finite-dimensional Hilbert space.

- We say that a subset $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ is an *operator algebra* on $\mathcal{H}$ if it is a linear subspace of $\mathcal{B}(\mathcal{H})$ that is closed under the operator product, i.e.,

$$A, B \in \mathcal{A} \implies \lambda A + \eta B \in \mathcal{A}, \quad \lambda, \eta \in \mathbb{K}, \qquad AB \in \mathcal{A}.$$

- An algebra on $\mathcal{H}$ is called a *-algebra, if it is also closed under the adjoint, i.e., $A \in \mathcal{A} \implies A^* \in \mathcal{A}$.

- A *-algebra on $\mathcal{H}$ that contains $I_{\mathcal{H}}$ is called a *von Neumann algebra* on $\mathcal{H}$.

Example 4.302.

(i) $\mathcal{A} = \mathcal{B}(\mathcal{H})$ is clearly a von Neumann algebra on $\mathcal{H}$, and it is the largest one.

(ii) $\mathcal{A} = \mathbb{C}I := \{\lambda I : \lambda \in \mathbb{C}\}$ is also a von Neumann algebra on $\mathcal{H}$, and it is clearly the smallest one, as any von Neumann algebra has to contain all constant multiples of the identity by definition.

(iii) Let $(e_i)_{i=1}^d$ be an ONB in $\mathcal{H}$, and $\mathcal{A}$ be all the diagonal operators in this ONB, i.e.,

$$\mathcal{A} := \left\{ \sum_{i=1}^{d} a_i \, |e_i\rangle\langle e_i| : (a_i)_{i\in[d]} \in \mathbb{C}^d \right\} = \operatorname{span}\{|e_i\rangle\langle e_i| : i \in [d]\}.$$

It is easy to see that $\mathcal{A}$ is a von Neumann algebra, that we call a *diagonal von Neumann algebra*.

(iv) More generally, let $(P_i)_{i=1}^r$ be pairwise orthogonal projections such that $\sum_{i=1}^r P_i = I$. Then

$$\mathcal{A} := \left\{ \sum_{i=1}^r a_i P_i : (a_i)_{i\in[r]} \in \mathbb{C}^r \right\} = \operatorname{span}\{P_i : i \in [d]\}$$

is a von Neumann algebra.

All the above examples are special cases of the following:

Example 4.303. Let $E := \{e_{k,l,i} : k = 1,\dots,r,\ l = 1,\dots,m_k,\ i = 1,\dots,d_k\}$ be an ONS in $\mathcal{H}$. Then

$$\begin{aligned}\mathcal{A} &:= \operatorname{span}_{\mathbb{C}}\{|e_{k,l,i}\rangle\langle e_{k,l,j}| : k = 1,\dots,r,\ l = 1,\dots,m_k,\ i,j = 1,\dots,d_k\} \\ &= \left\{ \sum_{k=1}^r \sum_{l=1}^{m_k} \sum_{i,j=1}^{d_k} a_{k,i,j} |e_{k,l,i}\rangle\langle e_{k,l,j}| : a_{k,i,j} \in \mathbb{C} \right\}\end{aligned}$$

is easily seen to be a *-algebra on $\mathcal{H}$, and it is a von Neumann algebra if and only if E is an ONB.

Written in matrix formalism, $\mathcal{A}$ is the collection of all block-diagonal operators in a given ONS, where the number of different types of blocks is r, each block is a $d_k \times d_k$ matrix, which appears with a multiplicity m_k. For instance,

$$\mathcal{A} := \left\{ \left[\begin{array}{cc|cc|c|c|c|c} a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & a_{11} & a_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{21} & a_{22} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & b & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & b & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & b & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] : a_{ij}, b \in \mathbb{C} \right\} \tag{4.68}$$

is a *-algebra in $\mathcal{B}(\mathbb{C}^8)$, where the ONS consists of the first 7 canonical basis vectors of $\mathbb{C}^8$, and we have $r = 2$, $d_1 = 2$, $m_1 = 2$ and $d_2 = 1$, $m_2 = 3$. This is not a von Neumann algebra, as the bottom right entry of each element of $\mathcal{A}$ is 0, and hence $I \notin \mathcal{A}$. If we modify the above example by allowing an arbitrary number in the bottom right entry then we get a von Neumann algebra.

As it turns out, the above example gives the most general *-algebra on a finite-dimensional Hilbert space:

Theorem 4.304. Let $\mathcal{H}$ be a finite-dimensional Hilbert space. A subset $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ is a *-algebra if and only if there exists an ONS $\{e_{k,l,i} : k = 1, \ldots, r,\, l = 1, \ldots, m_k,\, i = 1, \ldots, d_k\}$ such that

$$\begin{aligned}\mathcal{A} &= \operatorname{span}_{\mathbb{C}}\{|e_{k,l,i}\rangle \langle e_{k,l,j}| : k = 1, \ldots, r,\, l = 1, \ldots, m_k,\, i, j = 1, \ldots, d_k\} \\ &= \left\{ \sum_{k=1}^{r} \sum_{l=1}^{m_k} \sum_{i,j=1}^{d_k} a_{k,i,j} |e_{k,l,i}\rangle \langle e_{k,l,j}| : a_{k,i,j} \in \mathbb{C} \right\}.\end{aligned}$$

It is a von Neumann algebra if and only if the ONS is an ONB.

Proof. We have discussed the "if" direction in Example 4.303. We omit the proof of the "only if" direction, as we will not need it in the rest. We direct the interested reader to []. □

Remark 4.305. Using tensor products (see Section **??**), the above structure theorem can be rewritten in the following way: For any *-algebra $\mathcal{A}$ on a finite-dimensional Hilbert space $\mathcal{H}$, there exist natural numbers r, $(d_k)_{k=1}^r$, $(m_k)_{k=1}^r$, and an isometry

$$V : \oplus_{k=1}^r \mathbb{C}^{m_k} \otimes \mathbb{C}^{d_k} \to \mathcal{H} \qquad \text{s.t.} \qquad \mathcal{A} = V \left[\oplus_{k=1}^r I_{\mathbb{C}^{m_k}} \otimes \mathcal{B}(\mathbb{C}^{d_k})\right] V^*.$$

When $\mathcal{A}$ is a von Neumann algebra (i.e., it contains I), then the above V is a unitary.

Definition 4.306. Let $\mathcal{A}_i \subseteq \mathcal{B}(\mathcal{H}_i)$ be von Neumann algebras. A map $\alpha : \mathcal{A}_1 \to \mathcal{A}_2$ is called a *von Neumann algebra morphism* if it is

- linear
- multiplicative: $\alpha(AB) = \alpha(A)\alpha(B)$, $A, B \in \mathcal{A}_1$;
- adjoint-preserving: $\alpha(A^*) = (\alpha(A))^*$, $A \in \mathcal{A}_1$;
- unital: $\alpha(I_{\mathcal{H}_1}) = I_{\mathcal{H}_2}$

A bijective von Neumann algebra morphism is called a *von Neumann algebra isomorphism*. An injective von Neumann algebra morphism is also called a *representation* of $\mathcal{A}$.

The structure theorem 4.304 immediately yields the following:

Theorem 4.307. Let $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ be a von Neumann algebra on a finite-dimensional Hilbert space $\mathcal{H}$. Then $\mathcal{A}$ is isomorphic to the direct sum of full matrix algebras, i.e., there exist natural number $r, d_1, \ldots, d_r$, and a von Neumann algebra isomorphism

$$\alpha : \mathcal{A} \to \oplus_{k=1}^r \mathcal{B}(\mathcal{H}_k),$$

where $\dim \mathcal{H}_i = d_i$.

Proof. An explicit isomorphism can be obtained by simply removing the multiplicities. In detail, consider an ONB as in Theorem 4.304, define $\mathcal{H}_k := \operatorname{span}\{e_{k,1,i} : i \in [d_k]\}$, and

$$\alpha : A = \sum_{k=1}^{r} \sum_{l=1}^{m_k} \sum_{i,j=1}^{d_k} a_{k,i,j} \left|e_{k,l,i}\right\rangle \left\langle e_{k,l,j}\right| \mapsto \bigoplus_{k=1}^{r} \sum_{i,j=1}^{d_k} a_{k,i,j} \left|e_{k,1,i}\right\rangle \left\langle e_{k,1,j}\right|,$$

where on the RHS we consider $e_{k,1,i}$ as an element of $\mathcal{H}_k$. It is easy to see that this is a von Neumann algebra isomorphism. □

Remark 4.308. Alternatively, using the tensor product form of the structure theorem in Remark 4.305, an isomorphism as in the statement of Theorem 4.307 can be given as

$$\alpha : V \left[\oplus_{k=1}^{r} I_{\mathbb{C}^{m_k}} \otimes A_k\right] V^* \mapsto \oplus_{k=1}^{r} A_k.$$

Example 4.309. Consider the von Neumann algebra

$$\mathcal{A} := \left\{ \left[\begin{array}{cc|cc|c|c|c} a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & a_{11} & a_{12} & 0 & 0 & 0 \\ 0 & 0 & a_{21} & a_{22} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & b & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & b & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & b \end{array}\right] : a_{ij}, b \in \mathbb{C} \right\}$$

Then its image under the isomorphism given in Theorem 4.307 is

$$\alpha(\mathcal{A}) = \left\{ \left[\begin{array}{cc|c} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ \hline 0 & 0 & b \end{array}\right] : a_{i,j}, b \in \mathbb{C} \right\}.$$

Definition 4.310. Any representation of a von Neumann algebra $\mathcal{A}$ as in Theorem 4.307 is called a *minimal representation* of $\mathcal{A}$.

Remark 4.311. It is not too difficult to see that the dimensions $d_1, \ldots, d_r$ in Theorem 4.307 are uniquely determined by $\mathcal{A}$, and hence the minimal representation is essentially unique.

Definition 4.312. We say that a von Neumann algebra $\mathcal{A}$ on a finite-dimensional Hilbert space is *multiplicity-free*, if in its representation as in Theorem 4.304, all multiplicities $m_k = 1$.

Exercise 4.313. Show that a von Neumann algebra $\mathcal{A}$ on a finite-dimensional Hilbert space $\mathcal{H}$ is multiplicity-free if and only if there exist pairwise orthogonal projections $P_1, \ldots, P_r$ on $\mathcal{H}$, with $\sum_{k=1}^r P_k = I$, such that

$$\mathcal{A} = \left\{ A \in \mathcal{B}(\mathcal{H}) : \; A = \sum_{k=1}^{r} P_k A P_k \right\},$$

i.e., $\mathcal{A}$ is the collection of all operators on $\mathcal{H}$ that are block-diagonal according to the decomposition $\mathcal{H} = \oplus_{k=1}^r \operatorname{ran} P_k$.

There is a more abstract notion of a *-algebra, which, however, we will not use here. We only mention that complex-valued functions on a finite set can be naturally considered as a von Neumann algebra, as we show in Example 4.314, (which is essentially a reformulation of example (iii) in Example 4.302) and Remark 4.315. This identification of a function algebra with an operator algebra is very useful in treating classical and quantum models in the same formalism.

Example 4.314. For a finite set Ω, consider $\mathcal{H}_\Omega := l^2(\Omega)$ defined in Example 4.72, and recall that $\delta_\omega = \mathbf{1}_{\{\omega\}}$, $\omega \in \Omega$, is an ONB in $\mathcal{H}_\Omega$.

To every function $f \in \mathbb{C}^\Omega$, we assign an operator on $l^2(\Omega)$, called the *multiplication operator* corresponding to f, defined as

$$\mathrm{M}_f : \, g \mapsto fg, \qquad g \in l^2(\Omega).$$

Note that

$$\mathrm{M}_f \, \mathbf{1}_{\{\omega\}} = f(\omega) \mathbf{1}_{\{\omega\}}, \qquad x \in \Omega,$$

and hence M_f is diagonal in the canonical basis. Putting it in a different way, if we order the elements of Ω in some way, so that $\Omega = \{\omega_1, \ldots, \omega_d\}$, then the matrix of M_f in the ONB $\{e_i := \mathbf{1}_{\{\omega_i\}}\}_{i=1}^d$ is

$$\begin{bmatrix} f(\omega_1) & & & \\ & f(\omega_2) & & \\ & & \ddots & \\ & & & f(\omega_d) \end{bmatrix}, \qquad \text{with zeroes outside the diagonal.}$$

Vice versa, every operator F that is diagonal in this basis defines a function $f(\omega_i) := \left\langle \mathbf{1}_{\{i\}}, F \mathbf{1}_{\{i\}} \right\rangle$, $i \in [d]$, such that the corresponding multiplication operator is exactly F. Hence,

$$\mathcal{A}_\Omega := \{ M_f : \, f \in \mathbb{C}^\Omega \} = \{ A \in \mathcal{B}(l^2(\Omega) : \, A \text{ is diagonal in the canonical basis} \},$$

and the latter is a von Neumann algebra by Example 4.302 (iii).

Remark 4.315. It is easy to see that

$$\|f\|_\infty := \max_{\omega\in\Omega} |f(\omega)|$$

is a norm on $\mathbb{C}^\Omega$, called the *supremum norm*, and we use the notation $l^\infty(\Omega)$ for the function space $\mathbb{C}^\Omega$ equipped with $\|.\|_\infty$.

Clearly, $l^\infty(\Omega)$ is an algebra with the usual point-wise linear operations and product of functions, the constant one function 1 is a multiplicative unit in it, and we can define an adjoint operation on it by $f^* := \overline{f}$, the point-wise conjugate of the function.

Moreover, is easy to see that

$$M_{\lambda f+\eta g} = \lambda M_f + \eta M_g, \qquad M_{fg} = M_f M_g, \qquad M_{\mathbf{1}} = I, \qquad M_{\overline{f}} = M_f^*, \qquad \|M_f\| = \|f\|_\infty .$$

Thus, according to the above, $f \mapsto M_f$ is an algebra morphism that preserves the adjoint and the unit, and also the norm, and therefore we may identify the function algebra $l^\infty(\Omega)$ with $\mathcal{A}_\Omega$ in Example 4.314.

It is easy to see that if $(\mathcal{A}_i)_{i\in\mathcal{I}}$ are von Neumann algebras on $\mathcal{H}$ then so is $\cap_{i\in\mathcal{I}}\mathcal{A}_i$. In particular, for any $\mathcal{B} \subseteq \mathcal{B}(\mathcal{H})$, there exists a smallest von Neumann algebra containing $\mathcal{B}$, given as

$$\mathrm{vN}(\mathcal{B}) := \cap\{\mathcal{A} \subseteq \mathcal{B}(\mathcal{H}) : \mathcal{A} \text{ is a von Neumann algebra, and } \mathcal{A} \supseteq \mathcal{B}\}.$$

This is called the *von Neumann algebra generated by* $\mathcal{B}$. For a single operator $A \in \mathcal{B}(\mathcal{H})$, we use the notation

$$\mathrm{vN}(A) := \mathrm{vN}(\{A\}) = \cap\{\mathcal{A} \subseteq \mathcal{B}(\mathcal{H}) : \mathcal{A} \text{ is a von Neumann algebra, and } \mathcal{A} \ni A\}.$$

Exercise 4.316. Let $A \in \mathcal{B}(\mathcal{H})$ be a normal operator.

(i) Show that if $\mathcal{A}$ is a von Neumann algebra containing A then $f(A) \in \mathcal{A}$ for any function defined at least on $\mathrm{spec}(A)$. Conclude that every spectral projection of A is in $\mathcal{A}$.

(Hint: Use the Lagrange interpolation polynomials.)

(ii) Show that

$$\mathrm{vN}(A) = \left\{ \sum_{a\in\mathrm{spec}(A)} \lambda(a)P^A(a) : \lambda \in \mathbb{C}^{\mathrm{spec}(A)} \right\} = \mathrm{span}\{P^A(a) : a \in \mathrm{spec}(A)\}.$$

(Hint: Use Example (iv).)

(iii) Show that the functional calculus for A gives a von Neumann algebra isomorphism between $l^\infty(\mathrm{spec}(A)) \equiv \mathcal{A}_{\mathrm{spec}(A)}$ and $\mathrm{vN}(A)$.

(Hint: Use Exercise 4.257.)

Score: 3+5=8 points.

Exercise 4.317. (i) Let $\mathcal{H}$ be a finite-dimensional Hilbert space and $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ be an *algebra*, i.e., a linear subspace that is closed under the multiplication of operators, and assume that it contains I. Show that if $A \in \mathcal{A}$ is normal, then for any complex-valued function f defined on $\mathrm{spec}(A)$, $f(A) \in \mathcal{A}$. Conclude that all spectral projections of A are in $\mathcal{A}$, and that $A^* \in \mathcal{A}$ (even tough we did not assume that $\mathcal{A}$ is closed under the adjoint).

(ii) Let $\mathcal{A} := \left\{ \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} : a, b, c \in \mathbb{C} \right\} \subseteq \mathcal{B}(\mathbb{C}^2)$. Show that $\mathcal{A}$ is a unital algebra but not a *-algebra. Describe all the projections and all the normal elements in $\mathcal{A}$.

Solution: Hidden.

Definition 4.318. For a von Neumann algebra $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})_+$, let

- $\mathcal{A}_{\mathrm{sa}} := \{A \in \mathcal{A} : A = A^*\}$ be the set of self-adjoint elements in $\mathcal{A}$.
- $\mathcal{A}_+ := \mathcal{A}_{\geq 0} := \{A \in \mathcal{A} : A \geq 0\}$ be the set of PSD elements in $\mathcal{A}$.
- $\mathcal{A}_{++} := \mathcal{A}_{>0} := \{A \in \mathcal{A} : A > 0\}$ be the set of positive definite elements in $\mathcal{A}$.
- $\mathcal{A}_{[0,I]} := \{A \in \mathcal{A} : 0 \leq A \leq I\}$ be the set of *tests* in $\mathcal{A}$.
- $\mathcal{S}(A) := \{\varrho \in \mathcal{A}_+ : \mathrm{Tr}\, \varrho = 1\}$ be the set of *density operators*, or *states* in $\mathcal{A}$.

Exercise 4.319. Let $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ be a von Neumann algebra. Show that

$$\mathcal{A}_{\mathrm{sa}} = \mathrm{span}_{\mathbb{R}}\, \mathcal{A}_+ = \mathrm{span}_{\mathbb{R}}\, \mathcal{A}_{[0,I]} = \mathrm{span}_{\mathbb{R}}\, \mathcal{S}(A), \tag{4.69}$$

$$\mathcal{A} = \mathrm{span}_{\mathbb{C}}\, \mathcal{A}_{\mathrm{sa}} = \mathrm{span}_{\mathbb{C}}\, \mathcal{A}_+ = \mathrm{span}_{\mathbb{C}}\, \mathcal{A}_{[0,I]} = \mathrm{span}_{\mathbb{C}}\, \mathcal{S}(A). \tag{4.70}$$

Show that if $\mathcal{A}$ is multiplicity-free then we also have

$$\mathcal{A}_{\mathrm{sa}} = \mathrm{span}_{\mathbb{R}}\{|\psi\rangle\langle\psi| \in \mathcal{A} : \|\psi\| = 1\},$$
$$\mathcal{A} = \mathrm{span}_{\mathbb{C}}\{|\psi\rangle\langle\psi| \in \mathcal{A} : \|\psi\| = 1\}.$$

(Hint: Use Corollary 4.267.)

Score: 10 points.

According to the above exercise, any linear relation about the elements of a von Neumann algebra can be proved by proving it for every PSD operator, every density operator, etc. in $\mathcal{A}$. For instance, we have

$$A \perp \mathcal{A} \iff \operatorname{Tr} \varrho A = 0 \ \ \forall \varrho \in \mathcal{S}(\mathcal{A}).$$

Exercise 4.320. Let $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ be a von Neumann algebra on a finite-dimensional Hilbert space $\mathcal{H}$, and let $X \in \mathcal{A}$. Show the following:

(i) $X \in \mathcal{A}_{\text{sa}} \iff \forall A \in \mathcal{A}_{sa} : \quad \operatorname{Tr} XA \in \mathbb{R}.$

$\iff \forall A \in \mathcal{A}_{\geq 0} : \quad \operatorname{Tr} XA \in \mathbb{R}.$

(ii) $X \in \mathcal{A}_{\geq 0} \iff \forall A \in \mathcal{A}_{\geq 0} : \quad \operatorname{Tr} XA \in \mathbb{R}_{\geq 0}.$

(Hint: Use the previous point and the spectral decomposition.)

Solution: Hidden.

Remark 4.321. (ii) of Exercise 4.320 shows that the positive cone in a von Neumann algebra is self-dual; see also Exercise 4.189 and Section **??**.

4.28 Super-operators: Basic notions

Definition 4.322. A linear map that maps operators on a Hilbert space into operators on a (possibly different) Hilbert space is called a *super-operator*. More precisely, if $\mathcal{A}_1 \subseteq \mathcal{B}(\mathcal{H}_1)$ and $\mathcal{A}_2 \subseteq \mathcal{B}(\mathcal{H}_2)$ are von Neumann algebras then any element of $\mathcal{B}(\mathcal{A}_1, \mathcal{A}_2)$ is called a super-operator.

Remark 4.323. Note that any von Neumann algebra on a finite-dimensional Hilbert space is itself a finite-dimensional Hilbert space w.r.t. Hilbert-Schmidt inner product. In particular, any super-operator $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ admits an adjoint $\Phi^* : \mathcal{A}_2 \to \mathcal{A}_1$, determined by the relation

$$\langle A_2, \Phi(A_1) \rangle_{HS} = \langle \Phi^*(A_2), A_1 \rangle , \qquad A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2.$$

Definition 4.324. We say that a super-operator $\Phi \in \mathcal{B}(\mathcal{A}_1, \mathcal{A}_2)$ is

- *positive*, or *positivity-preserving* (in notation: $\Phi \geq 0$), if

$$A \in (\mathcal{A}_1)_{\geq 0} \quad \Longrightarrow \quad \Phi(A) \in (\mathcal{A}_2)_{\geq 0};$$

- *unital*, if

$$\Phi(I_{\mathcal{H}_1}) = I_{\mathcal{H}_2};$$

- *trace-preserving*, if

$$\operatorname{Tr}\Phi(A) = \operatorname{Tr} A, \qquad A \in \mathcal{A}_1.$$

Remark 4.325. It is important to note that we have two different notions of positivity for super-operators, that are also denoted the same way: the one defined above, and positive semi-definiteness when the von Neumann algebra $\mathcal{A}$ is considered as a finite-dimensional Hilbert space w.r.t. the Hilbert-Schmidt inner product. Note, however, that these two notions of positivity are unrelated. First of all, positive semi-definiteness may only be defined when the super-operator maps the algebra $\mathcal{A}$ into itself, while for the concept of a positivity-preserving super-operators, there is no such restriction. However, even when $\mathcal{A} = \mathcal{B}$, the two concepts of positivity are unrelated; see Exercise 4.329 below.

Unless otherwise stated, "positivity" for a super-operator Φ, and correspondingly, the notation $\Phi \geq 0$, will always mean the positivity-preserving property defined above, as that is the concept that is interesting for super-operators. If we nevertheless want to stress that we mean "positivity" in the sense of Definition 4.324, we might use the term "positivity-preserving".

Exercise 4.326. Let $\mathcal{A}_i \in \mathcal{B}(\mathcal{H}_i)$, $i = 1, 2$ be von Neumann algebras on the finite-dimensional Hilbert spaces $\mathcal{H}_i$, and let $\Phi \in \mathcal{B}(\mathcal{A}_1, \mathcal{A}_2)$. Show that

(i) $\Phi \geq 0 \quad \iff \quad \Phi^* \geq 0$

(ii) Φ is trace-preserving $\iff$ Φ^* is unital.

(iii) Φ is unital $\iff$ Φ^* is trace-preserving.

Exercise 4.327. Let $\alpha : \mathcal{A}_1 \to \mathcal{A}_2$ be a von Neumann algebra morphism.

(i) Show that α is positive and unital, and conclude that α^* is positive and trace-preserving.

(ii) Show that for any $M \in \mathrm{POVM}(\mathcal{A}_1, \mathcal{X})$, $\alpha(M) := (\alpha(M_x))_{x \in \mathcal{X}} \in \mathrm{POVM}(\mathcal{A}_2, \mathcal{X})$, and for any $\varrho \in \mathcal{S}(\mathcal{A}_2)$, $\alpha^*(\varrho) \in \mathcal{S}(\mathcal{A}_1)$.

Exercise 4.328. Let $\mathrm{Lin}(\mathcal{H}, \mathcal{K})$ be equipped with the Hilbert-Schmidt inner product, and for every $A \in \mathrm{Lin}(\mathcal{K})$ and $B \in \mathrm{Lin}(\mathcal{H})$, define the left multiplication L_A and the right multiplication R_B as

$$L_A : X \mapsto AX, \qquad R_B : X \mapsto XB, \qquad X \in \mathrm{Lin}(\mathcal{H}, \mathcal{K}).$$

Show the following:

(i) $L_{A_1} = L_{A_2} \iff A_1 = A_2$ and $R_{B_1} = R_{B_2} \iff B_1 = B_2$.

(ii) $(L_A)^* = L_{A^*}$, $(R_B)^* = R_{B^*}$.

(iii) L_A is normal/unitary/self-adjoint/positive/projection if and only if so is A, and the same holds for R_B and B.

(iv) If A is normal with spectral PVM P^A then L_A is normal with spectral PVM $P^{L_A} = L_{P^A}$. Formulate and prove the same statement for R_B.

(v) Let $\mathcal{K} = \mathcal{H}$, and for positive definite $A, B \in \mathrm{Lin}(\mathcal{H})$, define the *relative modular operator*

$$\Delta_{A/B} := L_A R_{B^{-1}}.$$

Show that $\Delta_{A/B}$ is a positive operator on $\mathrm{Lin}(\mathcal{H})$, and find its spectral decomposition.

Exercise 4.329. Let $\dim \mathcal{H} \geq 2$. Show examples of linear maps from $\mathcal{B}(\mathcal{H})$ to $\mathcal{B}(\mathcal{H})$ that are positive semi-definite but not positivity-preserving, and the other way around. (See Definition **??** and Remark 4.325.)

Solution: Hidden.

4.29 Positive operators revisited

Recall that an operator $A \in \mathrm{Lin}(\mathcal{H})$ is positive semidefinite (or simply positive), if $\langle x, Ax\rangle \geq 0$ for all $x \in \mathcal{H}$ (see Section 4.19).

By the functional calculus (Section 4.24), we can substitute positive operators into any complex-valued function that is defined on the positive half-line $[0, +\infty)$. Moreover, we will use the following convention throughout the text: for a positive operator A, we define all its real powers on its support only. That is, if $P(.)$ is the spectral PVM of a positive operator $A \in \mathrm{Lin}(\mathcal{H})_+$ then for every $r \in \mathbb{R}$ we define

$$A^r := \sum_{a>0} a^r P(a). \tag{4.71}$$

In particular,

$$A^0 = \sum_{a>0} P(a)$$

is the support projection of A, i.e., it is the projection onto the support of A, where

$$\mathrm{supp}\, A := (\ker A)^\perp = \mathrm{ran}\, A.$$

Another important special case is A^{-1}, that stands for the *generalized inverse* of A, satisfying

$$A^{-1}A = AA^{-1} = A^0.$$

Exercise 4.330. Show that for any $A \in \mathrm{Lin}(\mathcal{H})_+$,

$$A^0 = \lim_{t \searrow 0} A^t.$$

Exercise 4.331. (i) Let $A \in \mathcal{B}(\mathcal{H})_+$ be a PSD operator. Show that for any $x \in \mathcal{H}$,

$$Ax = 0 \iff \langle x, Ax\rangle = 0 \iff A^{1/2}x = 0.$$

(ii) Let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ be PSD operators. Show that

$$\ker(A_1 + \ldots + A_r) = \cap_{i=1}^r \ker A_i, \tag{4.72}$$
$$\mathrm{supp}(A_1 + \ldots + A_r) = \mathrm{span}\{\cup_{i=1}^r \mathrm{supp}\, A_i\}. \tag{4.73}$$

(iii) Conclude that if $A = \sum_{i=1}^r |v_i\rangle\langle v_i|$ then

$$\mathrm{supp}\, A = \mathrm{span}\{v_i : i = 1, \ldots, r\}.$$

Score: 2+8+2=12 points.

Solution:

(i) If $Ax = 0$ then $0 = \langle x, Ax\rangle = \left\langle A^{1/2}x, A^{1/2}x\right\rangle = \left\|A^{1/2}x\right\|^2$, and hence $A^{1/2}x = 0$. Conversely, if $A^{1/2}x = 0$ then $0 = A^{1/2}(A^{1/2}x) = Ax$.

(ii) By the previous point,

$$\begin{aligned} & x \in \ker(A_1 + \ldots + A_r) \\ & \iff \quad 0 = \langle x, (A_1 + \ldots + A_r)x\rangle = \langle x, A_1x\rangle + \ldots + \langle x, A_rx\rangle \\ & \iff \quad \langle x, A_ix\rangle = 0 \ \ \forall i \quad \iff \quad x \in \ker(A_i) \ \ \forall i, \end{aligned}$$

showing (4.72). From this we have

$$\begin{aligned} \mathrm{supp}(A_1 + \ldots + A_r) &= (\ker(A_1 + \ldots + A_r))^\perp \\ &= (\cap_{i=1}^r \ker(A_i))^\perp \supseteq (\ker(A_i))^\perp = \mathrm{supp}(A_i) \end{aligned}$$

for all i, and hence $\mathrm{supp}(A_1 + \ldots + A_r) \supseteq \cup_{i=1}^r \mathrm{supp}(A_i)$. Since $\mathrm{supp}(A_1 + \ldots + A_r)$ is a subspace, this also implies $\mathrm{supp}(A_1 + \ldots + A_r) \supseteq \mathrm{span}\{\cup_{i=1}^r \mathrm{supp}(A_i)\}$. In the converse direction, if $y \in \mathrm{supp}(A_1 + \ldots + A_r) = \mathrm{ran}(A_1 + \ldots + A_r)$ then there exists an $x \in \mathcal{H}$ such that $y = (A_1 + \ldots + A_r)x = A_1x + \ldots + A_rx \in \mathrm{span}\{\cup_{i=1}^r \mathrm{supp}(A_i)\}$.

(iii) Immediate from the previous point, as $\operatorname{ran}(|v\rangle\langle v|) = \mathbb{C}v$.

Exercise 4.332. (i) Show that for any $A \in \mathcal{B}(\mathcal{H})_+$ and $c \in (0, +\infty)$, $(cA)^0 = A^0$.

(ii) Show that for $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$A \leq B \implies \operatorname{supp} A \subseteq \operatorname{supp} B \iff A^0 \leq B^0.$$

(iii) Show that if $A \in \mathcal{B}(\mathcal{H})_+$ is PSD and P is a projection such that $cA \leq P$ for some $c \in (0, +\infty)$ then $AP = PA = A$.

(iv) Show that if $A \in \mathcal{B}(\mathcal{H})$ is self-adjoint and $P_{\operatorname{supp} A}$ is the projection onto its support then

$$-\|A\| P_{\operatorname{supp} A} \leq \lambda_{\min}(A) P_{\operatorname{supp} A} \leq A \leq \lambda_{\max}(A) P_{\operatorname{supp} A} \leq \|A\| P_{\operatorname{supp} A},$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the minimal and the maximal non-zero eigenvalues of A, respectively, when $A \neq 0$, and both are set to be 0 when $A = 0$.

(v) Show that if $A \in \mathcal{B}(\mathcal{H})_+$ is PSD and P is a projection then

$$P = A^0 \iff \exists c, d \in (0, +\infty) \text{ s.t. } cP \leq A \leq dP.$$

(vi) Show that if $B \in \mathcal{B}(\mathcal{H})$ is self-adjoint and $M \in \mathcal{B}(\mathcal{H})_+$ is PSD are such that $\operatorname{supp} B \subseteq \operatorname{supp} M$ then there exists an $\varepsilon_0 \in (0, +\infty)$ such that $M \pm \varepsilon B \geq 0$ for all $\varepsilon \in [0, \varepsilon_0)$.

Solution: Hidden.

Exercise 4.333. Let $A, M \in \operatorname{Lin}(\mathcal{H})_+$ be PSD operators such that $M \leq I$. Show that $\operatorname{Tr} AM \leq \operatorname{Tr} A$, and the following are equivalent:

(i) $\operatorname{Tr} AM = \operatorname{Tr} A$.

(ii) $\operatorname{supp} A \subseteq \{x \in \mathcal{H} : Mx = x\}$, i.e., the support of A is in the fixed-point space of M.

(iii) $P^M(1) \geq A^0$, where $P^M(1)$ is the spectral projection of M corresponding to the eigenvalue 1.

Solution: Hidden.

Exercise 4.334. Show that for an operator $A \in \operatorname{Lin}(\mathcal{H})$, the following are equivalent:

(i) A is PSD.

(ii) There exists a PSD $B \in \mathrm{Lin}(\mathcal{H})$ such that $A = B^2$.

(iii) There exists a self-adjoint $B \in \mathrm{Lin}(\mathcal{H})$ such that $A = B^2$.

(iv) There exists a $B \in \mathrm{Lin}(\mathcal{H})$ such that $A = B^*B$.

(v) There exist $B_1, \ldots, B_r \in \mathrm{Lin}(\mathcal{H})$ with some $r \in \mathbb{N}$ such that $A = \sum_{k=1}^r B_k^* B_k$.

Remark 4.335. It is this last characterization above how positivity is defined in more abstract settings (e.g., for C^*-algebras).

Exercise 4.336. Show that the product of two PSD operators is PSD if and only if they commute.

Solution: Hidden.

Definition 4.337. Let C be a cone in a Hilbert space $\mathcal{H}$. The *dual cone* of C is defined as

$$\widehat{C} := \{y \in \mathcal{H} : \langle y, x\rangle \geq 0 \ \ \forall x \in C\}.$$

We say that a cone C is *self-dual* if $\widehat{C} = C$.

Exercise 4.338. Show that for any $A \in \mathrm{Lin}(\mathcal{H})$,

$$A \in \mathrm{Lin}(\mathcal{H})_+ \iff \langle B, A\rangle_{HS} = \mathrm{Tr}\, BA \geq 0, \quad B \in \mathrm{Lin}(\mathcal{H})_+.$$

This property is called the *self-duality* of the cone $\mathrm{Lin}(\mathcal{H})_+$.

Solution: Hidden. □

Exercise 4.339. Let $A, B \in \mathcal{B}(\mathcal{H})_+$ be PSD operators. Show that the following are equivalent:

(i) $A \perp_{HS} B$, i.e., $\langle A, B\rangle_{HS} = 0$.

(ii) $\mathrm{Tr}\, AB = 0$.

(iii) $AB = 0$.

(iv) $\mathrm{ran}\, A \perp \mathrm{ran}\, B$.

(v) $A^0 B^0 = 0$.

(vi) $P^A(a) \perp P^B(b)$ for all $a \in \mathrm{spec}(A) \setminus \{0\}$ and $b \in \mathrm{spec}(B) \setminus \{0\}$.

(vii) Any of (i)–(iv) holds for $f(A)$ and $g(B)$ in place of A and B where f and g are arbitrary functions that contain $\operatorname{spec}(A)$ and $\operatorname{spec}(B)$ in their domains, respectively, and take the value 0 at 0 if they are defined there.

(viii) Any of (i)–(iv) holds for $f(A)$ and $g(B)$ in place of A and B where f and g are some functions as in (vii), and take strictly positive values on $\operatorname{spec}(A) \setminus \{0\}$ and $\operatorname{spec}(B) \setminus \{0\}$, respectively.

Score: 12 points.

Solution: (i)$\iff$(ii) by the definition of the Hilbert-Schmidt inner product, and $AB = 0 \iff \operatorname{ran} B \subseteq \ker A = (\operatorname{ran} A)^\perp$, due to Exercise 4.165, implying (iii)$\iff$(iv). The implication (iii)$\implies$(ii) is trivial.

The implication (ii)$\implies$(iii) can be proved as follows. Let $A = \sum_{i=1}^d a_i \, |e_i\rangle\langle e_i|$ and $B = \sum_{j=1}^d b_j \, |f_j\rangle\langle f_j|$ be eigen-decompositions of A and B, respectively. Assume

$$0 = \operatorname{Tr} AB = \operatorname{Tr} \sum_{i:\, a_i \neq 0} \sum_{j:\, b_j \neq 0} a_i b_j \, |e_i\rangle\langle e_i| \, |f_j\rangle\langle f_j| = \sum_{i:\, a_i \neq 0} \sum_{j:\, b_j \neq 0} a_i b_j |\, \langle e_i, f_j \rangle \,|^2.$$

Then $e_i \perp f_j$ for all i, j such that $a_i, b_j > 0$, which is equivalent to $\operatorname{ran} A \perp \operatorname{ran} B$.

Definition 4.340. Let $A, B \in \operatorname{Lin}(\mathcal{H})_+$ be PSD operators. If any (and hence all) of the conditions in Exercise 4.339 are satisfied then we say that A and B are orthogonal to each other.

Example 4.341. Let $A, B \in \operatorname{Lin}(\mathcal{H})_+$ be non-zero PSD operators. Show that the following are equivalent:

(i) A and B are not orthogonal.

(ii) There exists an eigenvector ψ of A with non-zero eigenvalue such that $\langle \psi, B\psi \rangle > 0$.

(iii) There exists a vector ψ such that $\langle \psi, A\psi \rangle > 0$, $\langle \psi, B\psi \rangle > 0$.

Exercise 4.342. Let $A, B \in \operatorname{Lin}(\mathcal{H})_+$ be PSD operators. Show that the following are equivalent:

(i) $\operatorname{supp} A \cap \operatorname{supp} B = \{0\}$.

(ii) If $C \in \operatorname{Lin}(\mathcal{H})_+$ is such that $C \leq A$ and $C \leq B$ then $C = 0$.

4.30 More on the PSD order

Recall that for $A, B \in \mathcal{B}(\mathcal{H})$, $A \leq B$ means that $B - A \geq 0$, i.e., $\langle x, (B-A)x\rangle \geq 0$ for all $x \in \mathcal{H}$. This is called the positive semidefinite (PSD) order on self-adjoint operators, and we will use the notation $A \leq_{\text{PSD}} B$ when we want to emphasize this.

This order is the extension of the pointwise ordering of real-valued functions on a finite set Ω, defined as $f \leq g$ if $f(\omega) \leq g(\omega)$ for all $\omega \in \Omega$, in the following sense: If $(|\omega\rangle)_{\omega\in\Omega}$ is an ONB in $\mathcal{H}$ then $\sum_\omega f(\omega)\,|\omega\rangle\langle\omega| \leq_{\text{PSD}} \sum_\omega g(\omega)\,|\omega\rangle\langle\omega|$ if and only if $f \leq g$ pointwise.

Obviously, for any finite set of functions $f_1, \ldots, f_r \in \mathbb{R}^\Omega$, there is a unique smallest upper bound (maximum) and a unique largest lower bound (minimum), given as

$$(\max_i f_i)(\omega) = \max_i f_i(\omega), \qquad (\min_i f_i)(\omega) = \min_i f_i(\omega), \qquad \omega \in \Omega,$$

respectively. This means that $\mathbb{R}^\Omega$ with the pointwise order has the lattice property:

Definition 4.343. A partial order on a set has the *lattice property* if any finite subset has a (unique) smallest upper bound and a largest lower bound.

The following exercise shows that $\mathcal{B}(\mathcal{H})_{\text{sa}}$ is not a lattice w.r.t the PSD order, i.e., a finite set of self-adjoint operator need not have a maximum or a minimum in general. Surprisingly, this can happen even in the most extreme case, i.e., when the set consists of two commuting operators. This may first seem to contradict the above considerations about functions and diagonal operators. The resolution of this apparent contradiction is that (in the case of the maximum) the set of upper bounds is defined among all self-adjoint operators, not only among those diagonal in the same basis, and, as a result, the upper bounds may not be comparable.

Exercise 4.344. Let $A_1 := \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and $A_2 := \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. Show that

$$\mathcal{U}(A_1, A_2) := \{Y \in \mathcal{B}(\mathcal{H}) : A_1 \leq Y,\, A_2 \leq Y\},$$

has no minimum. (Hint: Find an explicit parametrization of $U(A_1, A_2)$.)

Solution: Hidden.

However, as the following theorem shows, it is possible to define a maximum (minimum) for any finite set of operators by using a hybrid ordering: the set of upper (lower) bounds is defined w.r.t. the PSD order, while the minimal (maximal) element among the upper (lower) bounds is chosen according to the trace order.

The latter is defined as $A \leq_{\text{Tr}} B$ if $\operatorname{Tr} A \leq \operatorname{Tr} B$. Note that this is not a partial order but a preorder, as it is not antisymmetric; indeed, $\operatorname{Tr} A \leq \operatorname{Tr} B$ and $\operatorname{Tr} A \geq \operatorname{Tr} B$

only implies $\operatorname{Tr} A = \operatorname{Tr} B$ and not $A = B$. The trace order makes $\mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ a preordered vectors space, as $A \leq_{\mathrm{Tr}} B$ clearly implies that $\lambda A \leq_{\mathrm{Tr}} \lambda A$ for any $\lambda > 0$, and $A + C \leq_{\mathrm{Tr}} B + C$ for any $C \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$. (See section **??** for more detailes on these notions.) In fact, the same would hold if we extended the trace order to all (not just self-adjoint) operators.

The trace maximum of a finite set of PSD operators plays an important role in quantum state disrcimination; see Section **??** for details.

The proof of the following theorem uses Lemma 4.349, which we give after the theorem, for convenience.

Theorem 4.345. Let $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ be an observable algebra, and $\{A_1, \dots, A_r\} \subset \mathcal{A}_{\mathrm{sa}}$ be a non-empty finite set of self-adjoint operators.

(i) In the set of (PSD) upper bounds $\mathcal{U}(A_1, \dots, A_r) := \{Y \in \mathcal{A} : Y \geq_{PSD} A_1, \dots, A_r\}$ there is a unique element with minimal trace, which we denote by $\max_{\mathrm{Tr}}\{A_1, \dots, A_r\}$.

(ii) In the set of (PSD) lower bounds $\mathcal{L}(A_1, \dots, A_r) := \{Y \in \mathcal{A} : Y \leq_{PSD} A_1, \dots, A_r\}$ there is a unique element with maximal trace, which we denote by $\min_{\mathrm{Tr}}\{A_1, \dots, A_r\}$.

Proof. We only prove (i) as the proof of (ii) is completely similar (alternatively, can be obtained by replacing all A_i with $-A_i$).

Note that $\mathcal{U}(A_1 + cI, \dots, A_r + cI) = \mathcal{U}(A_1, \dots, A_r) + cI$, and the trace is linear, so we can assume without loss of generality that all A_i are PSD. Let $m := \inf\{\operatorname{Tr} Y : Y \geq A_1, \dots, A_r\}$ and $m' > m$. Then $\inf\{\operatorname{Tr} Y : Y \geq A_1, \dots, A_r\} = \inf \mathcal{U}'(A_1, \dots, A_r)$, where $\mathcal{U}'(A_1, \dots, A_r) := \{\operatorname{Tr} Y : Y \geq A_1, \dots, A_r, \|Y\|_1 = \operatorname{Tr} Y \leq m'\}$ is a compact set. Continuity of the trace then guarantees the existence of a $Y \in \mathcal{U}(A_1, \dots, A_r)$ with minimal trace.

Let us assume that there are two distinct elements Y_1 and Y_2 in $\mathcal{U}(A_1, \dots, A_r)$ with minimal trace. Let $\overline{Y} = (Y_1 + Y_2)/2$ and $\Delta = (Y_1 - Y_2)/2$. Then $Y_1 = \overline{Y} + \Delta$ and $Y_2 = \overline{Y} - \Delta$, and $Y_1, Y_2 \geq A_i$ implies $\overline{Y} - A_i \geq \pm\Delta$. Hence, by Lemma 4.349, there exists a constant $c_i > 0$ such that $Y_m - A_i \geq c_i|\Delta|$ for every $i = 1, \dots, r$. Taking $c := \min_i c_i$, we have $\overline{Y} - c|\Delta| \geq A_i$, $i = 1, \dots, r$. Thus, $\overline{Y} - c|\Delta| \in \mathcal{U}(A_1, \dots, A_r)$, but $\operatorname{Tr}(\overline{Y} - c|\Delta|) = \operatorname{Tr} \overline{Y} - c \operatorname{Tr} |\Delta| < \operatorname{Tr} \overline{Y} = \operatorname{Tr} Y_i$, $i = 1, 2$, contradicting our original assumption. □

Remark 4.346. Note that the trace minimum and maximum does not depend on the algebra $\mathcal{A}$, in the following sense. Let $\mathcal{A} = \{A \in \mathcal{B}(\mathcal{H}) : \sum_{k=1}^r P_K A P_k = A\}$ for some projections P_k with $\sum_{k=1}^r P_k = I$. If we take, for instance, the upper bounds in the whole of $\mathcal{B}(\mathcal{H})$ instead of just $\mathcal{A}$, then we may get a larger set, and thus the minimum of the trace functional on this larger set is at most as large as over the set of upper bounds in $\mathcal{A}$. However, if Y is any upper bound in $\mathcal{B}(\mathcal{H})$, i.e., $A_i \leq Y$ for all i, then $A_i = \sum_{k=1}^r P_k A_i P_k \leq \sum_{k=1}^r P_k Y P_k \in \mathcal{A}$, and $\operatorname{Tr} \sum_{k=1}^r P_k Y P_k = \operatorname{Tr} A$. Hence, for any upper bound in $\mathcal{B}(\mathcal{H})$, there is an upper bound in $\mathcal{A}$ with the same trace. Uniqueness

of the upper bound with minimal trace then yields $\max_{\mathrm{Tr}}\{A_1,\dots,A_r\} \in \mathcal{A}$. In particular, if $\mathcal{A}$ is a diagonal algebra then the trace maximum is just diagonal-wise maximum of the operators, i.e., the same as when we identify the operators with functions. Analogous considerations apply to $\min_{\mathrm{Tr}}\{A_1,\dots,A_r\}$.

Remark 4.347. Note that in general $\max_{\mathrm{Tr}}\{A_1,\dots,A_r\} \notin \{A_1,\dots,A_r\}$, not even for commuting operators and similarly for $\min_{\mathrm{Tr}}\{A_1,\dots,A_r\}$.

Exercise 4.348. Show that $\max_{\mathrm{Tr}}\{A_1,\dots,A_r\} \in \{A_1,\dots,A_r\}$ if and only if there exists a $j \in [r]$ such that $A_j \geq A_i$ for all $i \in [r]$. State and prove the analogous statement for the trace minimum.

Solution: Hidden.

Lemma 4.349. Let $D, T \in \mathcal{B}(\mathcal{H})$ be self-adjoint operators such that $D \geq \pm T$. Then there exists a positive constant $c \in (0, +\infty)$ such that $D \geq c|T|$.

Proof. First, $D \geq \pm T$ implies $D \geq (T + (-T))/2 = 0$, i.e., D is PSD. Let $\mathcal{H}_1$ denote the support of D, and decompose $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$. Then D and T can be written in the corresponding block forms as $D = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix}$ and $T = \begin{bmatrix} T_{11} & T_{12} \\ T_{12}^* & T22 \end{bmatrix}$, and positive semidefiniteness of $D \pm T$ implies $0 \geq T_{22} \geq 0$. Using again that $D + T \geq 0$, we finally obtain that $T_{12} = 0$, too.

The smallest eigenvalue of T_{11} is $\lambda := \left\|T_{11}^{-1}\right\|^{-1}$, which is strictly larger than 0, and hence we have $|T_{11}| \leq \|T\| I \leq (\|T\|/\lambda) D_{11}$. Thus, the assertion holds with $c := \|T\|/\lambda$. □

Note that for diagonal D and T, $D \geq \pm T$ implies $D \geq |T|$. As the following exercise shows, this is no longer true for operators, and hence $c < 1$ in Lemma 4.349 in general. On the other hand, $D \geq \pm T$ implies $D \geq_{\mathrm{Tr}} |T|$.

Recall that for a self-adjoint operator X, $\{X \geq c\}$ stands for the spectral projection $P^X([c, +\infty))$.

Exercise 4.350. Let $A, B \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ be self-adjoint operators such that

$$A \geq B \quad \text{and} \quad A \geq -B.$$

(i) Show that

$$\{B \geq 0\}A\{B \geq 0\} + \{B < 0\}A\{B < 0\} \geq |B|, \tag{4.74}$$

and hence

$$\operatorname{Tr} A \geq \operatorname{Tr}|B|, \quad \text{i.e.,} \quad A \geq_{\mathrm{Tr}} |B|. \tag{4.75}$$

(ii) Show examples of A, B as above for which $A \not\geq |B|$.

Solution: Hidden.

4.31 Polar decomposition and the singular value decomposition

We have seen in Section 4.23 that every normal operator has a particularly simple form, as it can be decomposed as a linear combination of projections. We have also seen (Exercise 4.246) that normality is also necessary for such a decomposition. There are, however, decompositions that work for any linear operator on a Hilbert space. The following Theorem 4.355 shows that any operator can be decomposed as a positive operator (its absolute value) followed by a partial isometry; this is called the *polar decomposition*. This turns out to be a very useful technical tool in many computations with operators; see, e.g., Section 4.32 on the trace-norm. Moreover, the polar decomposition combined with the spectral decomposition yields a canonical form of Hilbert space operators, called the *singular value decomposition*, that is an extension of the spectral decomposition to non-normal operators.

First, as a motivation, note that any complex number $a \in \mathbb{C}$ can be decomposed as $a = v|a|$, where $|a| \in \mathbb{R}_+$ and $|v| = 1$, and this decomposition to the product of a non-negative number and a complex number of modulus one is unique unless $a = 0$. Next, consider a normal operator $A \in \mathcal{B}(\mathcal{H})$, with an eigen-decomposition

$$A = \sum_{i=1}^{d} a_i \, |e_i\rangle\langle e_i| = \sum_{i=1}^{d} v_i |a_i| \, |e_i\rangle\langle e_i| = \left(\sum_{i=1}^{d} v_i \, |e_i\rangle\langle e_i| \right) \left(\sum_{i=1}^{d} |a_i| \, |e_i\rangle\langle e_i| \right) = V|A|,$$

or in matrix form,

$$A = \underbrace{\begin{bmatrix} v_1 & & & \\ & v_2 & & \\ & & \ddots & \\ & & & v_d \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} |a_1| & & & \\ & |a_2| & & \\ & & \ddots & \\ & & & |a_d| \end{bmatrix}}_{|A|}.$$

Here, V need not be unique if $a_i = 0$ for some i, but we can always choose it to be a partial isometry by setting $v_i = 0$ when $a_i = 0$, or to a unitary by setting $|v_i| = 1$ for all i.

To generalize this decomposition to arbitrary linear operators, we first need to define the absolute value:

Definition 4.351. For $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, let

$$|A| := \sqrt{A^* A}. \tag{4.76}$$

Remark 4.352. Note that $|A| \in \mathcal{B}(\mathcal{H})$, i.e., A and $|A|$ act on the same Hilbert space $\mathcal{H}$, but if $\mathcal{K} \neq \mathcal{H}$ then they map into different Hilbert spaces.

Note that for normal operators we already have a notion of absolute value via functional calculus, and it is easy to see that the two definitions coincide:

Exercise 4.353. Let $A \in \mathcal{B}(\mathcal{H})$ be a normal operator with spectral decomposition $A = \sum_{a \in \operatorname{spec}(A)} a P^A(a)$. Show that $\sqrt{A^*A} = \sum_{a \in \operatorname{spec}(A)} |a| P^A(a)$.

Of course, there may be other ways to extend the absolute value from normal to general operators. The above definition is further motivated by the following:

Exercise 4.354. Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Show that

$$\| |A|x \| = \|Ax\|, \qquad x \in \mathcal{H}, \tag{4.77}$$

and if $B \in \mathcal{B}(\mathcal{H})_+$ is such that $\|Bx\| = \|Ax\|$, $x \in \mathcal{H}$, then $B = |A|$. Conclude that

$$\|A\| = \| |A| \|, \qquad \ker |A| = \ker A, \qquad \operatorname{ran} |A| = (\ker A)^\perp = \operatorname{ran} A^*. \tag{4.78}$$

Solution: Hidden.

Now we can generalize the above decomposition of normal operators to arbitrary linear operators, possibly mapping between different Hilbert spaces:

Theorem 4.355. (Polar decomposition) Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. There exists a partial isometry $V \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ such that

$$A = V|A| \qquad \text{and} \qquad V^*V = |A|^0 = P_{\operatorname{ran}|A|}. \tag{4.79}$$

Moreover, the pair $(V, |A|)$ is unique in the sense that if S is a positive semidefinite operator and $\tilde{V}$ is a partial isometry such that $A = \tilde{V}S$ and $(\tilde{V})^*\tilde{V} = S^0$ then $S = |A|$ and $\tilde{V} = V$.

Proof. Let us define $Vx := 0$ on $\ker |A|$ and $V|A|x := Ax$ on $\operatorname{ran} |A|$. Note that if $|A|x_1 = |A|x_2$ then $\|Ax_1 - Ax_2\| = \|A(x_1 - x_2)\| = \| |A|(x_1 - x_2)\| = 0$ by (4.77), and hence V is well-defined. It is clear that V is linear on $\ker |A|$ and also on $\operatorname{ran} |A|$, and hence it defines a linear map on the whole of $\mathcal{H}$ by $V(y_1 + y_2) := Vy_1 + Vy_2$, $y_1 \in \ker |A|$, $y_2 \in \operatorname{ran} |A|$. Moreover, (4.77) also implies that V is isometric on $(\ker V)^\perp = \operatorname{ran} |A|$, completing the proof of existence.

To see uniqueness, let $A = \tilde{V}S$ be a decomposition with the given properties. Then $A^*A = S^*(\tilde{V})^*\tilde{V}S = S|S|^0S = S^2$, and hence, $S = |A|$. As $\tilde{V}$ is 0 on the orthocomplement of $(\operatorname{ran} |A|)^\perp$, we have $V = A|A|^{-1} = \tilde{V}|A||A|^{-1} = \tilde{V}|A|^0 = \tilde{V}$. $\square$

Definition 4.356. The unique partial isometry in (4.79) is called the *polar isometry* of A. (Note that it is only an isometry on $\operatorname{ran} |A|$.)

Remark 4.357. For the partial isometry in (4.79) we have

$$V^*V = |A|^0 = P_{\operatorname{ran}|A|} = P_{\operatorname{ran}A^*} \quad \text{and} \quad VV^* = P_{\operatorname{ran}V} = P_{\operatorname{ran}A}.$$

Moreover,

$$\ker V = \ker|A| = \ker A.$$

Remark 4.358. The above proof works for operators between infinite-dimensional spaces as well, with the following slight modification. Since $\mathcal{H} = \ker|A| \oplus \overline{\operatorname{ran}|A|}$, we need to invoke the boundedness of V on $\operatorname{ran}|A|$ to see that it uniquely extends to $\overline{\operatorname{ran}|A|}$, and the extension is automatically an isometry, and then proceed as in the proof of Theorem 4.355.

Remark 4.359. Note that (in the finite-dimensional case) we can express V as $V := A|A|^{-1}$, where the inverse stands for the generalized inverse, and it may be verified by a direct computation that this is a partial isometry with the required properties. Indeed, $V^*V = |A|^{-1}A^*A|A|^{-1} = |A|^0$ is a projection, and hence V is a partial isometry by Lemma 4.212, and we have $V|A| = A|A|^0$. Hence, we have to show that $A|A|^0 = A$, or equivalently, that $A(I - |A|^0) = 0$. This is indeed true, as for any $x \in \mathcal{H}$,

$$\left\|A(I-|A|^0)x\right\|^2 = \left\langle A(I-|A|^0)x, A(I-|A|^0)x\right\rangle = \left\langle A^*A(I-|A|^0)x, (I-|A|^0)x\right\rangle,$$

and $A^*A(I - |A|^0) = |A|^2(I - |A|^0) = 0$.

Remark 4.360. Since $V^*V = |A|^0$, V is an isometry if and only if $\{0\} = \ker|A| = \ker A$, i.e., if A is invertible. If $A \in \mathcal{B}(\mathcal{H})$ and $\dim\mathcal{K} \geq \dim\mathcal{H}$ then V can be modified on $\ker V$ to an isometry, and if $\dim\mathcal{K} = \dim\mathcal{H} < +\infty$ then it can be modified to a unitary U such that $A = U|A|$. Note, however, that this unitary is not unique if A is not invertible.

A few important consequences of Theorem 4.355 can be easily deduced as follows. Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and $A = V|A|$ be its polar decomposition as in (4.79). Then

$$V^*A = V^*V|A| = P_{\operatorname{ran}|A|}|A| = |A|$$

i.e., the polar decomposition can be "inverted" in the above form, and it is often in this form that we use the polar decomposition.

Next, observe that (4.79) yields

$$A^* = |A|V^*, \tag{4.80}$$

and hence

$$(A^*)^*A^* = AA^* = (V|A|)(|A|V^*) = V|A|V^*V|A|V^* = (V|A|V^*)^2,$$

from which

$$|A^*| = \sqrt{AA^*} = V|A|V^*, \tag{4.81}$$

where the first identity is due to the definition (4.76). Multiplying (4.81) from the left by V^*, we get

$$V^*|A^*| = V^*V|A|V^* = P_{\operatorname{ran}|A|}|A|V^* = |A|V^* = A^* \tag{4.82}$$

where the last identity follows from (4.80). Since

$$(V^*)^*V^* = VV^* = P_{\operatorname{ran}A} = P_{(\ker A^*)^\perp} = P_{(\ker|A^*|)^\perp} = P_{\operatorname{ran}|A^*|},$$

we see that V^* is the polar isometry of A^*, and thus (4.82), gives the polar decomposition of A^*, i.e., $A^* = V^*|A^*|$. That is, the polar decomposition of A (i.e., its absolute value and its polar isometry) then we can obtain the polar decomposition of A^*.

Exercise 4.361. Show that any element in the (operator norm) unit ball of $\mathcal{B}(\mathcal{H},\mathcal{K})$ can be written as the equal weight convex combination of
a) two isometries, if $\dim\mathcal{H} \leq \dim\mathcal{K}$;
b) two unitaries, if $\dim\mathcal{H} = \dim\mathcal{K}$;
c) two surjective partial isometries, if $\dim\mathcal{H} \geq \dim\mathcal{K}$.
(Hint: Use polar decomposition, and apply the functions $f_\pm(t) := t \pm \sqrt{1-t^2}$ to the absolute value.)

Solution: Hidden.

Combining the polar decomposition and the spectral decomposition of PSD operators yields a decomposition for arbitrary operators, called the singular value decomposition.

Corollary 4.362. (Singular value decomposition) For any linear operator $A \in \mathcal{B}(\mathcal{H},\mathcal{K})$ between finite-dimensional Hilbert spaces, there exist orthonormal systems $\{e_1,\ldots,e_r\}$, $\{f_1,\ldots,f_r\}$ in $\mathcal{H}$ and $\mathcal{K}$, respectively, and positive numbers $s_1,\ldots,s_r$, such that

$$A = \sum_{k=1}^{r} s_k\,|f_k\rangle\langle e_k|\,. \tag{4.83}$$

Proof. Since $|A|$ is normal, Corollary 4.237 yields that $|A|$ can be written as $|A| = \sum_{k=1}^{r} s_k \, |e_k\rangle\langle e_k|$, where $\{e_1, \dots, e_r\}$ is an orthonormal system and $s_1, \dots, s_r$ are positive numbers. Let V be the isometry from the polar decomposition theorem. Since it is an isometry on $(\ker A)^\perp = (\ker |A|)^\perp = \operatorname{span}\{e_1, \dots, e_r\}$, we can see that $f_k := Ve_k$, $k = 1, \dots, r$, is again an orthonormal system, and

$$A = V|A| = V \sum_{k=1}^{r} s_k \, |e_k\rangle\langle e_k| = \sum_{k=1}^{r} s_k V \, |e_k\rangle\langle e_k| = \sum_{k=1}^{r} s_k \, |f_k\rangle\langle e_k| \,. \qquad \square$$

Definition 4.363. We call any triple (S, F, E) a *singular value decomposition* of an operator $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ if $S = (s_1, \dots, s_r) \subseteq [0, +\infty)^r$, $F = (f_1, \dots, f_r)$ is an ONS in $\mathcal{K}$, $E = (e_1, \dots, e_r)$ is an ONS in $\mathcal{H}$, and

$$A = \sum_{k=1}^{r} s_k \, |f_k\rangle\langle e_k| \,.$$

For simplicity, we will write that "$A = \sum_{k=1}^{r} s_k \, |f_k\rangle\langle e_k|$ is a singular value decomposition of A" (even though it is mathematically not completely precise).

By Corollary 4.362, every linear operator between finite-dimensional Hilbert spaces has a singular value decomposition. As it turns out, the set of non-zero s_k, counted with multiplicities, is the same in any such decomposition, while the corresponding ONSs need not be unique.

Exercise 4.364. Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ and assume that A can be written as

$$A = \sum_{k=1}^{m} s_k \, |f_k\rangle \, \langle e_k| \,,$$

where $(s_k)_{k=1}^{r} \in [0, +\infty)^m$, $(e_k)_{k=1}^{m}$ is an ONS in $\mathcal{H}$ and $(f_k)_{k=1}^{m}$ is an ONS in $\mathcal{K}$. Show that $\operatorname{rk}(A) = \#\{k : s_k > 0\}$, the non-zero s_k counted with multiplicities give exactly the non-zero eigenvalues of $|A|$, counted with multiplicities, and

$$\operatorname{ran} A = \operatorname{span}\{e_k : s_k > 0\}, \qquad (\ker A)^\perp = \operatorname{span}\{f_k : s_k > 0\}.$$

There are different conventions to define the singular values of an operator $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$.

Definition 4.365. Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then the sequence of *singular values* $s_1(A), \dots, s_r(A)$ might mean any of the following:

- The non-zero eigenvalues of $|A|$, counted with multiplicities, ordered in any way; in this case $r = \operatorname{rk}(A)$.

- The same as above, appended with zeros so that $r = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}$.
- All the eigenvalues of $|A|$, counted with multiplicites, ordered in any way; in this case, $r = \dim \mathcal{H}$.
- Any of the above, with the singular values ordered in decreasing order.

Remark 4.366. With any of the above conventions, the sequence of non-zero singular values is well-defined up to the ordering of the singular values, and the set of (non-zero) singular values is the same as the set of (non-zero) eigenvalues of $|A|$, without multiplicities.

The different conventions above may be convenient in different problems, while in most cases it is irrelevant which convention is used. We will always emphasize when we work with a particular convention.

Note that the number of non-zero eigenvalues (counted with multiplicities) is always well-defined, but it has the disadvantage that with this convention, different operators mapping between the same spaces could have a different number of singular values.

In Exercise 4.367 it is natural to choose $r = \mathrm{rk}(A)$ or $r = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}$, so that the number of singular values is the same for A and for A^*, and both conventions work equally well. On the other hand, in Exercise 4.368 it is more convenient to choose $r = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}$, so that we take into account part of the kernel of A as well.

Exercise 4.367. Show that the non-zero eigenvalues of $|A|$ and $|A^*|$, counted with multiplicities, are the same, and hence the non-zero singular values of A and A^* are the same. Show that $A = \sum_{k=1}^{r} s_k \, |f_k\rangle \langle e_k|$ is a singular value decomposition of A if and only if $A^* = \sum_{k=1}^{r} s_k \, |e_k\rangle \langle f_k|$ is a singular value decomposition of A^*.

Exercise 4.368. Let us count the number of singular values with multiplicities as $\min\{\dim \mathcal{H}, \dim \mathcal{K}\}$. Show that a non-negative number $s \in [0, +\infty)$ is a singular value of A if and only if there exist non-zero vectors $x \in \mathcal{H}$ and $y \in \mathcal{K}$ such that

$$Ax = sy, \qquad A^* y = sx. \tag{4.84}$$

Show that the same characterization may not hold with the convention where the number of singular values (with multiplicities) is defined to be $\dim \mathcal{H}$.

Solution: Hidden.

Obviosuly, if $A \in \mathcal{B}(\mathcal{H})$ is normal, then any eigen-decomposition $A = \sum_{k=1}^{r} a_k \, |e_k\rangle\langle e_k|$ as in (4.43) is a singular value decomposition. One might think that for normal operators (at least with non-degenerate eigenvalues), any singular-value decomposition $A = \sum_{k=1}^{r'} a'_k \, |f'_k\rangle\langle e'_k|$ is an eigen-decomposition, i.e., $f'_k = e'_k$. This, however, is not the case.

Exercise 4.369. Let $|0\rangle, |1\rangle$ be an ONS in a Hilbert space $\mathcal{H}$, and let

$$X := |1\rangle\langle 0| + |0\rangle\langle 1| . \tag{4.85}$$

Find the unique eigen-decomposition of X and compare with (4.85).

Exercise 4.370. Find the polar decomposition and all singular value decompositions of $A := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

Exercise 4.371. Let P, Q be projections and $PQ = \sum_{i=1}^{r} \lambda_i |e_i\rangle\langle f_i|$ be a singular value decomposition of PQ.

(i) Show that $Pe_i = e_i$, $Pf_i = \lambda_i e_i$, $Qf_i = f_i$, $Qe_i = \lambda_i f_i$.

(ii) Show that $\langle e_i, f_j\rangle = \lambda_i \delta_{i,j}$.

(iii) Show that the projections $P_0 := P - \sum_{i=1}^{r} |e_i\rangle\langle e_i|$, $P_i := P_{\mathrm{span}\{e_i, f_i\}}$, $i = 1, \ldots, r$, and $P_{r+1} := Q - \sum_{i=1}^{r} |f_i\rangle\langle f_i|$ are pairwise orthogonal, and $\sum_{k=0}^{r+1} P_k = P \vee Q$.

(iv) Show that $P_0 \perp Q$ and $P_{r+1} \perp P$.

(v) Assume that we make the PVM measurement $(P_k)_{k=0}^{r+1}$ on a system which is either in state ϱ or in state σ, where $\varrho^0 = P$ and $\sigma^0 = Q$. Show that if the outcome is 0 then we know that the system was in state ϱ, and the post-measurement state has support P_0; if the outcome is 1 then we know that the system was in state σ, and the post-measurement state has support Q_0; if the outcome is $k \in [r]$ then the post-measurement state is $|e_k\rangle\langle e_k|$ if the original state was ϱ, and $|f_k\rangle\langle f_k|$ if the original state was σ.

Solution: Hidden.

4.32 Schatten p-norms and the trace norm

Definition 4.372. For $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ and $p \in [1, +\infty)$, we define the *Schatten p-norm* (or simply p-norm) of A as

$$\|A\|_p := (\mathrm{Tr}\,|A|^p)^{1/p} = \left(\sum_{i=1}^{r} s_i(A)^p \right)^{1/p},$$

where $s_1(A), \ldots, s_r(A)$ are the singular values of A. For $p = 1$,

$$\|A\|_1 = \mathrm{Tr}\,|A| = \sum_{i=1}^{r} s_i(A)$$

is called the *trace-norm* of A.

Remark 4.373. The case $p = 2$ yields the Hilbert-Schmidt norm

$$\|A\|_2 := (\operatorname{Tr}|A|^2)^{1/2} = (\operatorname{Tr} A^*A)^{1/2} = \langle A, A\rangle_{HS}^{1/2}.$$

Exercise 4.374. Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Show that $p \mapsto \|A\|_p$ is monotone decreasing in p, and

$$\|A\|_\infty := \lim_{p\to+\infty} \|A\|_p = \inf_{p\in[1,+\infty)} \|A\|_p = \max_{i\in[r]} s_i(A),$$

where the $(s_i(A))_{i=1}^r$ are the singular values of A. Show that $\|A\|_\infty$ is exactly the operator norm of A.

Solution: Hidden.

Exercise 4.375. Show that for any $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and any $p, q \in [1, +\infty]$, $p < q$,

$$\|A\|_q \le \|A\|_p \le \|A\|_q (\operatorname{Tr}|A|^0)^{1-\frac{p}{q}} \le \|A\|_q (\dim \mathcal{H})^{1-\frac{p}{q}}.$$

(Hint: Use the Hölder inequality for real vectors.)

We have already seen that the operator norm $\|.\|_\infty$ and the Hilbert-Schmidt norm give norms on $\mathcal{B}(\mathcal{H}, \mathcal{K})$. It is clear that all Schatten norms are strictly positive and positive homogeneous, i.e.,

$$\begin{aligned} \|A\|_p &\ge 0 && \|\lambda A\|_p = |\lambda| \|A\|_p, \qquad \lambda \in \mathbb{C}, \\ &= 0 \iff A = 0 \end{aligned}$$

so one only needs to prove the triangle inequality to show that the Schatten norms are indeed norms. This can be obtained from the following inequality:

Theorem 4.376. Let $A_i \in \mathcal{B}(\mathcal{H}_{i-1}, \mathcal{H}_i)$, $i = 1, \dots, n$, and let $r, p_1, \dots, p_n \in (0, +\infty)$ so that $\frac{1}{p_1} + \dots + \frac{1}{p_n} = \frac{1}{r}$. Then

$$\|A_n \cdot \ldots \cdot A_1\|_r \le \|A_n\|_{p_n} \cdot \ldots \cdot \|A_1\|_{p_1}.$$

The proof of the above inequality is quite involved, and we only do it in the following special case:

Lemma 4.377. Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then,

$$\|AB\|_1 \le \|B\|_\infty \|A\|_1, \quad B \in \mathcal{B}(\mathcal{L}, \mathcal{H}), \qquad \|BA\|_1 \le \|B\|_\infty \|A\|_1, \quad B \in \mathcal{B}(\mathcal{K}, \mathcal{L}).$$

Moreover,

$$\begin{aligned} \|A\|_1 &= \max\{|\operatorname{Tr} BA| \,:\, B \in \mathcal{B}(\mathcal{K}, \mathcal{H}),\, \|B\|_\infty \le 1\} \\ &= \max\{|\operatorname{Tr} VA| \,:\, V \in \mathcal{B}(\mathcal{K}, \mathcal{H}),\, \text{is a partial isometry}\}. \end{aligned} \tag{4.86}$$

Proof. Let $A = \sum_{k=1}^{r} s_k \, |f_k\rangle\langle e_k|$ be a singular-value decomposition of A and let $AB = V|AB|$ be the polar decomposition of AB. Then $|AB| = V^*AB$ and hence,

$$\begin{aligned}\|AB\|_1 = \operatorname{Tr}|AB| = \operatorname{Tr} V^*AB = \operatorname{Tr} BV^*A = \operatorname{Tr}\sum_{k=1}^{r} s_k \, |BV^*f_k\rangle\langle e_k| \\ = \sum_{k=1}^{r} s_k \, \langle e_k, BV^*f_k\rangle \leq \Big(\max_k |\, \langle e_k, BV^*f_k\rangle \,|\Big) \sum_{k=1}^{r} s_k \leq \|BV^*\| \operatorname{Tr}|A| \\ \leq \|B\|_\infty \|A\|_1\end{aligned}$$

The inequality $\|BA\|_1 \leq \|B\|_\infty \|A\|_1$ follows the same way.

As a consequence,

$$\begin{aligned}\|A\|_1 &\geq \sup\{|\operatorname{Tr} BA| \, : \, B \in \mathcal{B}(\mathcal{K},\mathcal{H}),\, \|B\|_\infty \leq 1\} \\ &\geq \sup\{|\operatorname{Tr} VA| \, : \, V \in\in \mathcal{B}(\mathcal{K},\mathcal{H}) \text{ is a partial isometry}\}\end{aligned}$$

where the second inequality is obvious. On the other hand, by choosing V from the polar decomposition $A = V|A|$, we see that both suprema are attained at V^* and are equal to $\operatorname{Tr}|A| = \|A\|_1$. □

Proposition 4.378. $\|\ \|_1$ defines a norm on $\mathcal{B}(\mathcal{H})$, which we call the *trace-norm*.

Proof. The properties $\|A\|_1 \geq 0$ and $\|\lambda A\|_1 = |\lambda| \, \|A\|_1$, $\lambda \in \mathbb{C}$, are obvious. If $A = \sum_{k=1}^{r} s_k \, |f_k\rangle\langle e_k|$ is a singular-value decomposition of A then $\|A\|_1 = \sum_{k=1}^{r} s_k$, hence if $\|A\|_1 = 0$ then $A = 0$. Finally, let $A, B \in \mathcal{B}(\mathcal{H},\mathcal{K})$ and let $A+B = V|A+B|$ be the polar decomposition of $A+B$. Then

$$\|A+B\|_1 = \operatorname{Tr}|A+B| = |\operatorname{Tr} V^*(A+B)| \leq |\operatorname{Tr} V^*A| + |\operatorname{Tr} V^*B| \leq \|A\|_1 + \|B\|_1 ,$$

where the last inequality is due to (4.86). □

Corollary 4.379. For any finite-dimensional Hilbert space $\mathcal{H}$,

$$T(A\,,\,B) := \frac{1}{2}\|A-B\|_1$$

defines a metric on $\mathcal{B}(\mathcal{H})$, which we call the *trace-norm distance* on $\mathcal{B}(\mathcal{H})$.

Exercise 4.380. Let $A, B \in \mathcal{B}(\mathcal{H})$ be such that $\operatorname{Tr} A = \operatorname{Tr} B$. Show that

$$\frac{1}{2}\|A-B\|_1 = \max_{0\leq T\leq I} \operatorname{Tr} T(A-B),$$

and the equality is reached for any T such that $\{A-B>0\} \leq T \leq \{A-B\geq 0\}$.

Since the trace-norm distance comes from a norm, it obviously satisfies the following joint convexity property: if $A_1, \ldots, A_r, B_1, \ldots, B_r \in \mathrm{Lin}(\mathcal{H})$ and $p_1, \ldots, p_r \geq 0$ are such that $p_1 + \ldots + p_r = 1$ then

$$T\left(\sum_{i=1}^{r} p_i A_i, \sum_{i=1}^{r} p_i B_i\right) \leq \sum_{i=1}^{r} p_i T(A_i, B_i).$$

Exercise 4.381. Show that

$$|\operatorname{Tr} A| \leq \operatorname{Tr}|A|, \qquad \operatorname{Tr}|A| = \operatorname{Tr}|A^*|.$$

Solution: Hidden.

Proposition 4.382. Let $\{p_1, \ldots, p_r\}$ and $\{q_1, \ldots, q_r\}$ be probability distributions and $\{A_1, \ldots, A_r\}$ and $\{B_1, \ldots, B_r\}$ be operators on $\mathcal{B}(\mathcal{H})$. Then,

$$\begin{aligned} T\left(\sum_{k=1}^{r} p_k A_k\,,\, \sum_{k=1}^{r} q_k B_k\right) \leq \min\{&(\max_k \|B_k\|) T\left(p\,,\, q\right) + \sum_{k=1}^{r} p_k T\left(A_k\,,\, B_k\right), \\ &(\max_k \|A_k\|) T\left(p\,,\, q\right) + \sum_{k=1}^{r} q_k T\left(A_k\,,\, B_k\right)\}. \end{aligned} \tag{4.87}$$

In particular,

$$T\left(\sum_{k=1}^{r} p_k A_k\,,\, \sum_{k=1}^{r} p_k B_k\right) \leq \sum_{k=1}^{r} p_k T\left(A_k\,,\, B_k\right). \tag{4.88}$$

Proof. We have

$$\begin{aligned} T\left(\sum_{k=1}^{r} p_k A_k\,,\, \sum_{k=1}^{r} q_k B_k\right) &= \left\| \sum_{k=1}^{r} p_k A_k - \sum_{k=1}^{r} q_k B_k \right\|_1 \\ &= \left\| \sum_{k=1}^{r} p_k (A_k - B_k) - \sum_{k=1}^{r} (q_k - p_k) B_k \right\|_1 \\ &\leq \sum_{k=1}^{r} p_k \left\| A_k - B_k \right\|_1 + \max_k \|B_k\| \sum_{k=1}^{r} |q_k - p_k|, \end{aligned}$$

and $T\left(\sum_{k=1}^{r} p_k A_k\,,\, \sum_{k=1}^{r} q_k B_k\right) \leq (\max_k \|A_k\|) T\left(p\,,\, q\right) + \sum_{k=1}^{r} q_k T\left(A_k\,,\, B_k\right)$ follows the same way. The choice $p_k := q_k$, $k = 1, \ldots, r$ yields (4.88). □

We refer to inequality (4.87) as the *strong convexity of the trace-norm distance*, and inequality (4.88) as the *joint convexity of the trace-norm distance*.

Exercise 4.383. Let $A, B \in \mathcal{B}(\mathcal{H})$ be such that $\operatorname{Tr} A = \operatorname{Tr} B$. Show that

$$\frac{1}{2}\|A-B\|_1 = \max_{0\le T\le I} \operatorname{Tr} T(A-B),$$

and the equality is reached for any T such that $\{A-B>0\} \le T \le \{A-B\ge 0\}$.

Show that if we also have $A, B \ge 0$ then

$$\frac{1}{2}\|A-B\|_1 \le \operatorname{Tr} A - \max\{\lambda_{\min}(A), \lambda_{\min}(B)\}. \tag{4.89}$$

As a consequence,

$$\frac{\operatorname{Tr} A + \operatorname{Tr} B}{2} - \frac{1}{2}\|A-B\|_1 \ge \max\{\lambda_{\min}(A), \lambda_{\min}(B)\}.$$

Proof. We have

$$0 = \operatorname{Tr}(A-B) = \operatorname{Tr}(A-B)_+ - \operatorname{Tr}(A-B)_-,$$

from which

$$\operatorname{Tr}|A-B| = \operatorname{Tr}(A-B)_+ + \operatorname{Tr}(A-B)_- = 2\operatorname{Tr}(A-B)_+ = 2\max_{0\le T\le I} \operatorname{Tr} T(A-B).$$

Assume now that $A, B \ge 0$. Since $B \ge \lambda_{\min}(B) I$, we have $A - B \le A - \lambda_{\min}(B) I$, and hence,

$$\operatorname{Tr}(A-B)\{A-B\ge 0\} \le \operatorname{Tr} A\{A-B\ge 0\} - \lambda_{\min}(B)\operatorname{Tr}\{A-B\ge 0\}.$$

Note that $\operatorname{Tr}(A-B)=0$ yields that either $A=B$, or $\{A-B\ge 0\}\neq 0$. In the first case, the inequality in (4.89) is obvious. In the second case, $\operatorname{Tr}\{A-B\ge 0\} \ge 1$, and hence we can continue the above inequality as

$$\operatorname{Tr}(A-B)\{A-B\ge 0\} \le \operatorname{Tr} A - \lambda_{\min}(B) = \operatorname{Tr} B - \lambda_{\min}(B).$$

Switching the role of A and B, we finally get (4.89). □

Exercise 4.384. Show that if $A, B \in \mathcal{B}(\mathcal{H})_+$ then

$$\min_{0\le T\le I}\{\operatorname{Tr} TA + \operatorname{Tr}(I-T)B\} = \frac{\operatorname{Tr} A + \operatorname{Tr} B}{2} - \frac{1}{2}\|A-B\|_1.$$

Conclude that

$$\frac{\operatorname{Tr} A + \operatorname{Tr} B}{2} - \frac{1}{2}\|A-B\|_1 \ge \min\{\lambda_{\min}(A), \lambda_{\min}(B)\}\dim\mathcal{H}. \tag{4.90}$$

Proof. We have

$$\begin{aligned}\operatorname{Tr} TA + \operatorname{Tr}(I-T)B &= \operatorname{Tr} B - \operatorname{Tr} T(A-B) \geq \operatorname{Tr} B - \operatorname{Tr}(A-B)_+ \\ &= \operatorname{Tr} B - \frac{1}{2}\operatorname{Tr}(A-B) - \frac{1}{2}\operatorname{Tr}|A-B| = \frac{\operatorname{Tr} A + \operatorname{Tr} B}{2} - \frac{1}{2}\|A-B\|_1,\end{aligned}$$

and equality holds for $T = \{A - B \geq 0\}$. Let $\lambda_{\min} := \min\{\lambda_{\min}(A), \lambda_{\min}(B)\}$. Then $A \geq \lambda_{\min} I$ and $B \geq \lambda_{\min} I$ yields that

$$\operatorname{Tr} TA + \operatorname{Tr}(I-T)B \geq \lambda_{\min} \operatorname{Tr} T + \lambda_{\min} \operatorname{Tr}(I-T) = \lambda_{\min} \dim \mathcal{H},$$

and we arrive at (4.90) □

Exercise 4.385. Let $x, y \in \mathcal{H}$. Express the eigenvalues of $|x\rangle\langle x| - |y\rangle\langle y|$, and express $\||x\rangle\langle x| - |y\rangle\langle y|\|_p$, $1 \leq p \leq +\infty$, in terms of $\|x\|$, $\|y\|$ and $\langle x, y\rangle$. In particular, show that

$$\||x\rangle\langle x| - |y\rangle\langle y|\|_1 = \sqrt{(\|x\|^2 + \|y\|^2) - 4|\langle x, y\rangle|^2}.$$

Show that if ψ_1, ψ_2 are state vectors in some Hilbert space then

$$\||\psi_1\rangle\langle\psi_1| - |\psi_2\rangle\langle\psi_2|\|_1 = 2\sqrt{1 - |\langle\psi_1, \psi_2\rangle|^2} = 2\sqrt{1 - F\left(|\psi_1\rangle\langle\psi_1|, |\psi_2\rangle\langle\psi_2|\right)^2}.$$

(Hint: Use that if $x, y \neq 0$ then the problem is essentially 2-dimensional, and use Gram-Schmidt orthogonalization.)

Solution: Hidden.

Exercise 4.386. Show that if $A, B \in \mathcal{B}(\mathcal{H})_+$ then

$$\min_{0 \leq T \leq I} \{\operatorname{Tr} TA + \operatorname{Tr}(I-T)B\} = \frac{\operatorname{Tr} A + \operatorname{Tr} B}{2} - \frac{1}{2}\|A-B\|_1.$$

Solution: Hidden.

We use the convention $0^p := 0$, $p \in \mathbb{R} \setminus \{0\}$. Accordingly, powers of a positive semidefinite operator are only taken on its support, and defined to be zero on the orthocomplement of the support.

Lemma 4.387. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Then

$$\operatorname{Tr} A^\alpha B^{1-\alpha} \leq (\operatorname{Tr} A)^\alpha (\operatorname{Tr} B)^{1-\alpha}, \qquad \alpha \in [0, 1], \tag{4.91}$$

and if $A^0 \leq B^0$ then

$$\operatorname{Tr} A^\alpha B^{1-\alpha} \geq (\operatorname{Tr} A)^\alpha (\operatorname{Tr} B)^{1-\alpha}, \qquad \alpha \in (1, +\infty). \tag{4.92}$$

Proof. Let $A = \sum_a aP_a$ and $B = \sum_b bQ_b$ be the spectral decompositions of A and B, respectively, and let $c(a,b) := a \operatorname{Tr} P_a Q_b$, $d(a,b) := b \operatorname{Tr} P_a Q_b$. Then

$$\operatorname{Tr} A^\alpha B^{1-\alpha} = \sum_{a,b} a^\alpha b^{1-\alpha} \operatorname{Tr} P_a Q_b = \sum_{a,b} c(a,b)^\alpha d(a,b)^{1-\alpha}.$$

Let $\alpha \in (0,1)$ and $p := 1/\alpha$, $q = 1/(1-\alpha)$, such that $1/p + 1/q = 1$. Hölder's inequality then yields

$$\begin{aligned}\sum_{a,b} c(a,b)^\alpha d(a,b)^{1-\alpha} &\le \left(\sum_{a,b} (c(a,b)^\alpha)^p\right)^{1/p} \left(\sum_{a,b} (d(a,b)^\alpha)^q\right)^{1/q} \\ &= \left(\sum_{a,b} c(a,b)\right)^\alpha \left(\sum_{a,b} d(a,b)\right)^{1-\alpha} \\ &= (\operatorname{Tr} A)^\alpha (\operatorname{Tr} B)^{1-\alpha},\end{aligned}$$

proving (4.91) for $\alpha \in (0,1)$, and the cases $\alpha = 0$ and $\alpha = 1$ are trivial.

Let $\alpha \in (1, +\infty)$, and $p := 1/\alpha$, $q = 1/(1-\alpha)$, such that $1/p + 1/q = 1$. Note that $p > 0$, $q < 0$, and $A^0 \le B^0$ implies that if $d(a,b) = 0$ for some a, b then also $c(a,b) = 0$. Hence, the reverse Hölder inequality yields that

$$\begin{aligned}\sum_{a,b} c(a,b)^\alpha d(a,b)^{1-\alpha} &\ge \left(\sum_{a,b} (c(a,b)^\alpha)^p\right)^{1/p} \left(\sum_{a,b} (d(a,b)^\alpha)^q\right)^{1/q} \\ &= (\operatorname{Tr} A)^\alpha (\operatorname{Tr} B)^{1-\alpha},\end{aligned}$$

proving (4.92). □

Remark 4.388. Lemma 4.387 is a special case of the generalized log-sum inequality for (classical) f-divergences.

Remark 4.389. Inequality (4.92) does not hold in general without the support condition. For instance, if A, B are orthogonal and both non-zero then $\operatorname{Tr} A^\alpha B^{1-\alpha} = 0$ but $(\operatorname{Tr} A)^\alpha (\operatorname{Tr} B)^{1-\alpha} > 0$.

Proposition 4.390. For any $C \in \mathcal{B}(\mathcal{H})_+$ and $p \in \mathbb{R} \setminus \{0\}$, we have

$$\|C\|_p := (\operatorname{Tr} C^p)^{\frac{1}{p}} = \begin{cases} \sup\limits_{\sigma \in \mathcal{S}(\mathcal{H})} \operatorname{Tr} C\sigma^{1-\frac{1}{p}}, & p \in [1, +\infty), \\ \inf\limits_{\sigma \in \mathcal{S}(\mathcal{H}),\, C^0 \le \sigma^0} \operatorname{Tr} C\sigma^{1-\frac{1}{p}}, & p \in (0,1), \\ \inf\limits_{\sigma \in \mathcal{S}(\mathcal{H}),\, C^0 \ge \sigma^0} \operatorname{Tr} C\sigma^{1-\frac{1}{p}}, & p \in (-\infty, 0). \end{cases}$$

Moreover, all the optimizations can be restricted to $\sigma^0 = C^0$.

Proof. Let $p \in [1, +\infty)$, $\alpha := 1/p \in (0, 1)$, and $A := C^p = C^{1/\alpha}$. Then for any $B \in \mathcal{B}(\mathcal{H})_+$, (4.91) yields

$$\operatorname{Tr} CB^{1-\alpha} = \operatorname{Tr} A^\alpha B^{1-\alpha} \leq (\operatorname{Tr} A)^\alpha (\operatorname{Tr} B)^{1-\alpha} = (\operatorname{Tr} C^p)^{\frac{1}{p}} (\operatorname{Tr} B)^{1-\alpha}, \tag{4.93}$$

and hence,

$$(\operatorname{Tr} C^p)^{\frac{1}{p}} \geq \operatorname{Tr} C(B/\operatorname{Tr} B)^{1-\alpha} = \operatorname{Tr} C(B/\operatorname{Tr} B)^{1-\frac{1}{p}}. \tag{4.94}$$

This yields that

$$(\operatorname{Tr} C^p)^{\frac{1}{p}} \geq \sup_{\sigma \in \mathcal{S}(\mathcal{H})} \operatorname{Tr} C\sigma^{1-\frac{1}{p}},$$

and equality holds with the choice $\sigma := C^p/\operatorname{Tr} C^p$.

Next, let $p \in (0, 1)$, $\alpha := 1/p \in (1, +\infty)$, and $A := C^p = C^{1/\alpha}$. Then for any $B \in \mathcal{B}(\mathcal{H})_+$ such that $A^0 \leq B^0$, the inequalities in (4.93) and (4.94) hold in the opposite direction, due to (4.92), and hence

$$(\operatorname{Tr} C^p)^{\frac{1}{p}} \leq \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \operatorname{Tr} C\sigma^{1-\frac{1}{p}}.$$

Equality can again be attained by $\sigma := C^p/\operatorname{Tr} C^p$.

Finally, let $p \in (-\infty, 0)$, $\alpha := 1 - 1/p \in (1, +\infty)$, and $B := C^p = C^{1/(1-\alpha)}$. Then for any $A \in \mathcal{B}(\mathcal{H})_+$ such that $A^0 \leq C^0 = B^0$, we have

$$\operatorname{Tr} A^\alpha C = \operatorname{Tr} A^\alpha B^{1-\alpha} \geq (\operatorname{Tr} A)^\alpha (\operatorname{Tr} B)^{1-\alpha} = (\operatorname{Tr} A)^\alpha (\operatorname{Tr} C^p)^{\frac{1}{p}},$$

and hence,

$$(\operatorname{Tr} C^p)^{\frac{1}{p}} \leq \operatorname{Tr} C(A/\operatorname{Tr} A)^\alpha = \operatorname{Tr} C(A/\operatorname{Tr} A)^{1-\frac{1}{p}}.$$

This yields

$$(\operatorname{Tr} C^p)^{\frac{1}{p}} \leq \inf_{\sigma \in \mathcal{S}(\mathcal{H}),\, C^0 \geq \sigma^0} \operatorname{Tr} C\sigma^{1-\frac{1}{p}},$$

and equality can be attained by choosing $\sigma := C^{1/(1-\alpha)}/\operatorname{Tr} C^{1/(1-\alpha)}$. □

It is easy to obtain another variational formula for $\operatorname{Tr} C^p$ from Proposition 4.390:

Proposition 4.391. For any $C \in \mathcal{B}(\mathcal{H})_+$ and $p \in \mathbb{R} \setminus \{0\}$, we have

$$\operatorname{Tr} C^p = \begin{cases} \sup\limits_{X \in \mathcal{B}(\mathcal{H})_+} \left\{ p \operatorname{Tr} CX^{1-\frac{1}{p}} - (p-1) \operatorname{Tr} X \right\}, & p \in [1, +\infty), \\ \inf\limits_{X \in \mathcal{B}(\mathcal{H})_+,\, C^0 \leq X^0} \left\{ p \operatorname{Tr} CX^{1-\frac{1}{p}} - (p-1) \operatorname{Tr} X \right\}, & p \in (0, 1), \\ \sup\limits_{X \in \mathcal{B}(\mathcal{H})_+,\, C^0 \geq X^0} \left\{ p \operatorname{Tr} CX^{1-\frac{1}{p}} - (p-1) \operatorname{Tr} X \right\}, & p \in (-\infty, 0). \end{cases}$$

Moreover, all the optimizations can be restricted to $\sigma^0 = C^0$.

Proof. We give the proof of the case $p \in [1, +\infty)$; the proof for the other cases are completely similar. Then

$$\sup_{X \in \mathcal{B}(\mathcal{H})_+} \left\{ p \operatorname{Tr} CX^{1-\frac{1}{p}} - (p-1) \operatorname{Tr} X \right\} = \sup_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\lambda > 0} f(\lambda),$$

where

$$f(\lambda) := p\lambda^{1-\frac{1}{p}} \operatorname{Tr} C\sigma^{1-\frac{1}{p}} - (p-1)\lambda$$

is concave. We have $f'(\lambda) = (p-1)\lambda^{-\frac{1}{p}} \operatorname{Tr} C\sigma^{1-\frac{1}{p}} - (p-1)$, which is zero if and only if $\lambda = \lambda_p = (\operatorname{Tr} C\sigma^{1-\frac{1}{p}})^p$, and

$$f(\lambda_p) = p(\operatorname{Tr} C\sigma^{1-\frac{1}{p}})^{p-1} \operatorname{Tr} C\sigma^{1-\frac{1}{p}} - (p-1)(\operatorname{Tr} C\sigma^{1-\frac{1}{p}})^p = (\operatorname{Tr} C\sigma^{1-\frac{1}{p}})^p.$$

Hence,

$$\sup_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\lambda > 0} f(\lambda) = \sup_{\sigma \in \mathcal{S}(\mathcal{H})} (\operatorname{Tr} C\sigma^{1-\frac{1}{p}})^p = \operatorname{Tr} C^p,$$

where the last identity is due to Proposition 4.390. □

For any linear map $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$, let

$$\|\Phi\|_{p/q} := \sup\{\|\Phi(A)\|_p \,:\, \|A\|_q \leq 1\}.$$

We have the following:

Proposition 4.392. For any linear map $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$,

$$\|\Phi\|_{1/1} \leq \|\Phi^*\|_{\infty/\infty}.$$

Proof. Let $A \in \mathcal{B}(\mathcal{H})$ and let $\Phi(A) = V|\Phi(A)|$ be its polar decomposition. Then,

$$\|\Phi(A)\|_1 = \operatorname{Tr} V^*\Phi(A) = \operatorname{Tr} \Phi^*(V^*)A \leq \|\Phi^*(V^*)\|_\infty \|A\|_1 \leq \|\Phi^*\|_{\infty/\infty} \|A\|_1. \quad \square$$

Corollary 4.393. Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a positive and trace-preserving map. Then

$$\|\Phi(A)\|_1 \leq \|A\|_1, \qquad A \in \mathcal{B}(\mathcal{H}).$$

Proof. By the Russo-Dye theorem, $\|\Phi\|_{\infty/\infty} = 1$ and hence the assertion follows from Proposition 4.392. □

Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a linear map. The norm of Φ induced by the trace norm is given by

$$\|\Phi\|_{1/1} := \sup\{\|\Phi(X)\|_1 \,:\, \|X\|_1 \leq 1\}. \tag{4.95}$$

Lemma 4.394. Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a linear map. Then

$$\|\Phi\|_{1/1} = \sup\{\|\Phi(|x\rangle\langle x|)\|_1 \,:\, \|x\| = 1\}.$$

Proof. Let $X \in \mathcal{B}(\mathcal{H})$ be such that $\|X\|_1 \leq 1$, and let $X = \sum_{i=1}^r \lambda_i |f_i\rangle\langle e_i|$ be a singular value decomposition. Since $0 \leq \lambda_i$ and $\sum_{i=1}^r \lambda_i \leq 1$, we have

$$\|\Phi(X)\|_1 = \left\| \sum_{i=1}^r \lambda_i \Phi(|f_i\rangle\langle e_i|) \right\|_1 \leq \sum_{i=1}^r \lambda_i \|\Phi(|f_i\rangle\langle e_i|)\|_1 \leq \max_{1\leq i\leq r} \|\Phi(|f_i\rangle\langle e_i|)\|_1 .$$

Hence, in the optimization in (4.95), it is enough to consider operators of rank 1. Obviously, we can also restrict to operators with unit trace norm. Now, let X be a rank 1 operator with unit trace norm; then there are unit vectors $x, y \in \mathcal{H}$ such that $X = |y\rangle\langle x| = \frac{1}{4}\sum_{k=0}^3 i^k \left|x + i^k y\right\rangle\left\langle x + i^k y\right|$. Hence,

$$\|\Phi(X)\|_1 \leq \frac{1}{4}\sum_{k=0}^3 \left\|\Phi(\left|x + i^k y\right\rangle\left\langle x + i^k y\right|)\right\| \leq \max_{0\leq k\leq 3} \left\|\Phi(\left|x + i^k y\right\rangle\left\langle x + i^k y\right|)\right\| ,$$

which proves the assertion. □

Corollary 4.395. Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a positive trace-preserving map. Then $\|\Phi\|_{1/1} = 1$.

Proof. Obvious from Lemma 4.394 □

Definition 4.396. Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a linear map. For each $n \in \mathbb{N}$, define

$$\|\Phi\|_{1/1}^{(n)} := \|\mathrm{id}_n \otimes \Phi\|_{1/1} ,$$

where id_n is the identity map on $\mathcal{B}(\mathcal{H}_n)$ for some n-dimensional Hilbert space $\mathcal{H}_n$ (obviously, the value of $\|\Phi\|_{1/1}^{(n)}$ does not depend on the choice of $\mathcal{H}_n$). It is easy to see that $\|(.)\|_{1/1}^{(n)}$ is a norm on $\mathcal{B}(\mathcal{B}(\mathcal{H}), \mathcal{B}(\mathcal{K}))$ for every $N \in \mathbb{N}$.

The *diamond norm* of Φ is defined as

$$\|\Phi\|_\diamond := \sup_{n\in\mathbb{N}} \|\Phi\|_{1/1}^{(n)} .$$

Note that $\|\Phi\|_\diamond$ might be infinite for a general linear map.

4.33 Anti-linear operators

Definition 4.397. Let $\mathcal{H}, \mathcal{K}$ be finite-dimensional Hilbert spaces. A map $A : \mathcal{H} \to \mathcal{K}$ is *anti-linear* (or *conjugate linear*) if it is

(i) *additive:* $A(x+y) = Ax + Ay, \quad x, y \in \mathcal{H}, \quad$ and

(ii) *conjugate homogeneous:* $A(\lambda x) = \overline{\lambda} Ax, \ x \in \mathcal{H}, \lambda \in \mathbb{C}.$

The adjoint of an anti-linear operator is defined somewhat differently from the adjoint of a linear operator:

Lemma 4.398. Let $A : \mathcal{H} \to \mathcal{K}$ be an anti-linear operator. There exists a unique anti-linear operator $A^* : \mathcal{K} \to \mathcal{H}$ such that

$$\langle y, Ax \rangle = \langle x, A^* y \rangle , \qquad x \in \mathcal{H}, \ y \in \mathcal{K}. \tag{4.96}$$

Proof. For every $y \in \mathcal{K}$, the map $x \mapsto \langle Ax, y \rangle$ is a linear functional on a finite-dimensional Hilbert space. Hence, by the Riesz representation theorem (Exercise **??**), there exists a unique vector in $\mathcal{H}$, that we denote by $A^* y$, such that

$$\langle A^* y, x \rangle = \langle Ax, y \rangle .$$

Taking the conjugate of both sides yields (4.96). Anti-linearity of the map $y \mapsto A^* y$ follows from the uniqueness of $A^* y$; we leave it as an exercise. □

Recall that for linear operators, the map $A \mapsto A^*$ is a conjugate linear involution. For anti-linear maps, we have the following:

Exercise 4.399. Show that for anti-linear operators, the map $A \mapsto A^*$ is a *linear* involution, i.e.,

$$(A+B)^* = A^* + B^*, \qquad (\lambda A)^* = \lambda A^*,$$

and

$$(A^*)^* = A$$

for any anti-linear operators A, B on the same Hilbert space, and any $\lambda \in \mathbb{C}$.

Solution: Hidden.

Recall that for linear operators $A_1, \ldots, A_n$ such that the product $A_1 \cdot \ldots \cdot A_n$ makes sense, we have the identity

$$(A_1 \cdot \ldots \cdot A_n)^* = A_n^* \cdot \ldots \cdot A_1^*. \tag{4.97}$$

This identity is still true for any *mixed* product of linear and anti-linear operators:

Exercise 4.400. Show that (4.97) is true for any product of linear and anti-linear operators.

Solution: Hidden.

Exercise 4.401. Show that for an anti-linear operator $J : \mathcal{H} \to \mathcal{K}$, the following are equivalent:

(i) $J^*J = I$

(ii) $\|Jx\|^2 = \|x\|^2, \ \ x \in \mathcal{H}.$

(iii) $\langle Jx, Jy\rangle = \langle y, x\rangle, \ \ x, y \in \mathcal{H}.$

Solution: Hidden.

Definition 4.402. An anti-linear operator $J : \mathcal{H} \to \mathcal{K}$, satisfying any (and hence all) of the above (i)–(iii), is called an *anti-isometry*. Obviously, anti-isometries are injective. A surjective anti-isometry is called an *anti-unitary*.

Exercise 4.403. Show that an anti-linear operator $J : \mathcal{H} \to \mathcal{K}$ is an anti-unitary if and only if

$$J^*J = I_{\mathcal{H}} \qquad \text{and} \qquad JJ^* = I_{\mathcal{K}}.$$

One has to be careful with manipulations involving anti-linear operators, as the following examples show:

Exercise 4.404. Show that the identities

$$|Ax\rangle = A\,|x\rangle\,, \qquad \langle Ax| = \langle x|\,A^*$$

do not hold for an anti-linear operator A and a vector x, unless $Ax = 0$. On the other hand, if A and B are both anti-linear operators then

$$B\,|y\rangle\langle x|\,A = |By\rangle\langle A^*x| \tag{4.98}$$

for every x, y for which the above expression makes sense.

Solution: Hidden.

The trace of an anti-linear operator cannot be defined in a basis-independent way, as the following example shows:

Example 4.405. Let $e_1, \dots, e_d$ be a basis in a Hilbert space $\mathcal{H}$, and let A be the unique anti-linear extension of the map $e_k \mapsto e_k$, i.e.,

$$Ax := \sum_{k=1}^{d} \overline{\langle e_k, x\rangle} e_k.$$

Define $\tilde{e}_k := ie_k$, where $i = \sqrt{-1}$ is the imaginary unit. Then

$$\sum_{k=1}^{d} \langle e_k, Ae_k\rangle = \sum_{k=1}^{d} 1 = d,$$
$$\sum_{k=1}^{d} \langle \tilde{e}_k, A\tilde{e}_k\rangle = \sum_{k=1}^{d} \langle ie_k, A(ie_k)\rangle = \sum_{k=1}^{d} \langle ie_k, (-i)Ae_k\rangle = -\sum_{k=1}^{d} \langle e_k, Ae_k\rangle = -d.$$

4.34 The conjugate Hilbert space

Let $\mathcal{H}$ be a complex Hilbert space. We will introduce new algebraic operations and a new inner product on $\mathcal{H}$ with which it again becomes a Hilbert space. We will call this Hilbert space the *conjugate Hilbert space* of $\mathcal{H}$ and we will denote it by $\overline{\mathcal{H}}$. We will use the notation $\overline{x}$ for a vector $x \in \mathcal{H}$ when we consider it as an element of $\overline{\mathcal{H}}$ instead of $\mathcal{H}$. Note that $\overline{\mathcal{H}} = \mathcal{H}$ as sets and $\overline{x} = x$ for every $x \in \mathcal{H}$.

The addition in $\overline{\mathcal{H}}$ is the same as in $\mathcal{H}$, i.e., $\overline{x} + \overline{y} = \overline{x+y}$ for all $x, y \in \mathcal{H}$. For a scalar $\lambda \in \mathbb{C}$ and a vector $x \in \mathcal{H}$, we define their new product as

$$\lambda\overline{x} := \overline{\lambda}x = \overline{\overline{\lambda}x}.$$

(Verify that this is indeed a vector space!) Moreover, we define an inner product on $\overline{\mathcal{H}}$ as

$$\langle \overline{x}, \overline{y}\rangle_{\overline{\mathcal{H}}} := \langle y, x\rangle_{\mathcal{H}} = \overline{\langle x, y\rangle_{\mathcal{H}}}.$$

(We will usually omit the subscripts $_{\mathcal{H}}$ and $_{\overline{\mathcal{H}}}$). One can easily see that $\overline{\mathcal{H}}$ with this inner product is indeed a Hilbert space, and, moreover, that $\overline{x} \mapsto \langle x|$ yields a natural isomorphism between $\overline{\mathcal{H}}$ and the dual of $\mathcal{H}$. In particular, $\mathcal{H}$ and $\overline{\mathcal{H}}$ are also isomorphic as Hilbert spaces. Obviously, if $\{e_i\}_{i\in I}$ is a basis for $\mathcal{H}$ then $\{\overline{e}_i\}_{i\in I}$ is a basis for $\overline{\mathcal{H}}$. Note that

$$\overline{(\overline{\mathcal{H}})} = \mathcal{H}.$$

Note that if A is a linear operator from $\mathcal{H}$ to $\mathcal{K}$ then A is also linear when considered as an operator from $\overline{\mathcal{H}}$ to $\overline{\mathcal{K}}$; we will denote this operator as $\overline{A}$. Indeed,

$$\overline{A}(\overline{x}+\overline{y}) = A(x+y) = Ax+Ay = \overline{A}\overline{x}+\overline{A}\overline{y}, \qquad \overline{A}(\lambda\overline{x}) = A(\overline{\lambda}x) = \overline{\lambda}Ax = \lambda\overline{Ax} = \lambda\overline{A}\overline{x}.$$

Exercise 4.406. Show that:

(i) The assignment $A \mapsto \overline{A}$ is antilinear, i.e.,

$$\overline{A+B} = \overline{A} + \overline{B}, \qquad \overline{\lambda A} = \overline{\lambda}\,\overline{A},$$

for all $A, B \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ and $\lambda \in \mathbb{C}$.

(ii) For every $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$,

$$\overline{A^*} = (\overline{A})^*.$$

(iii) For every $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, $x \in \mathcal{H}$ and $y \in \mathcal{K}$,

$$\langle \overline{y}, \overline{A}\overline{x} \rangle = \overline{\langle y, Ax \rangle}, \quad \text{and, in particular,} \quad \operatorname{Tr}_{\overline{\mathcal{H}}} \overline{A} = \overline{\operatorname{Tr}_{\mathcal{H}} A}.$$

(iv) For every $x \in \mathcal{H}, y \in \mathcal{K}$,

$$\overline{|y\rangle\langle x|} = |\overline{y}\rangle\langle\overline{x}|\,.$$

In particular,

$$\overline{A} = \overline{\sum_{ij} \langle f_j, Ae_i\rangle\, |f_j\rangle\langle e_i|} = \sum_{ij} \overline{\langle f_j, Ae_i\rangle}\, |\overline{f_j}\rangle\langle\overline{e_i}| = \sum_{ij} \langle \overline{f}_j, \overline{A}\overline{e}_i\rangle\, |\overline{f}_j\rangle\langle\overline{e}_i|$$

Now, if $\mathcal{H}$ is a Hilbert space then $\mathcal{B}(\mathcal{H})$ is also a Hilbert space when equipped with the Hilbert-Schmidt inner product $\langle A, B\rangle_{HS} := \operatorname{Tr} A^*B$, $A, B \in \mathcal{B}(\mathcal{H})$. As we have seen above, every $A \in \mathcal{B}(\mathcal{H})$ can be considered as a bounded linear operator $\overline{A}$ on $\overline{\mathcal{H}}$, and the identity $\lambda\overline{A} = \overline{\overline{\lambda}A}$ holds. Moroever, by Exercise 4.406,

$$\langle \overline{A}, \overline{B}\rangle_{HS} = \operatorname{Tr}_{\overline{\mathcal{H}}} \overline{A}^*\overline{B} = \operatorname{Tr}_{\overline{\mathcal{H}}} \overline{(A^*)}\,\overline{B} = \operatorname{Tr}_{\overline{\mathcal{H}}} \overline{A^*B} = \overline{\operatorname{Tr}_{\mathcal{H}} A^*B} = \overline{\langle A, B\rangle}_{HS}.$$

Hence, $\mathcal{B}(\overline{\mathcal{H}})$ can be naturally identified with $\overline{\mathcal{B}(\mathcal{H})}$.

4.35 Distances on operators

The following is one possible way to symmetrize it.

Definition 4.407. *Hilbert's projective metric* d_H is defined on $\mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+$ as

$$d_H(\varrho\|\sigma) := \begin{cases} D^*_\infty(\varrho\|\sigma) + D^*_\infty(\sigma\|\varrho), & \varrho, \sigma \neq 0, \\ 0, & \varrho = \sigma = 0. \end{cases}$$

Exercise 4.408. Show that d_H is constant on rays, i.e., $d_H(\lambda\varrho\|\eta\sigma) = d_H(\varrho\|\sigma)$ for any $\lambda, \eta \in (0, +\infty)$. In particular, $d_H(\varrho\|\sigma) = 0$ if and only if $\varrho = \lambda\sigma$ for some $\lambda \in (0, +\infty)$. Show that d_H defines a metric on the rays of positive definite operators, the points of which are of the form $[\varrho] := \{\lambda\varrho : \lambda \in (0, +\infty)\}$, $\varrho \in \mathcal{B}(\mathcal{H})_{++}$.

5 Quantum probabilistic models

5.1 Quantum states and measurements

Now we turn to the mathematical description of quantum probabilistic models of physical phenomena. Any such model is uniquely specified by a Hilbert space $\mathcal{H}$, that we call the Hilbert space of the system that we want to model. The possible states of the system in such a model are given by all density operators on the Hilbert space:

Definition 5.1. An operator $\varrho \in \mathcal{B}(\mathcal{H})$ is a *density operator*, or a *state* if it is positive semi-definite, and has unit trace:

$$\varrho \geq 0, \qquad \operatorname{Tr} \varrho = 1.$$

We denote the set of density operators on $\mathcal{H}$ by $\mathcal{S}(\mathcal{H})$, and call it the *state space* of the Hilbert space $\mathcal{H}$.

It is clear that if $\varrho_1, \ldots, \varrho_r \in \mathcal{S}(\mathcal{H})$ are states, and $p_1, \ldots, p_r$ is a probability distribution, then

$$\sum_{i=1}^{r} p_i \varrho_i \geq 0, \qquad \text{and} \qquad \operatorname{Tr} \sum_{i=1}^{r} p_i \varrho_i = \sum_{i=1}^{r} p_i \underbrace{\operatorname{Tr} \varrho_i}_{=1} = 1,$$

i.e., $\sum_{i=1}^{r} p_i \varrho_i$ is again a state. Thus, we have the following:

Proposition 5.2. For any Hilbert space $\mathcal{H}$, the state space $\mathcal{S}(\mathcal{H})$ is a convex set.

Example 5.3. It is easy to see that if $\psi \in \mathcal{H}$ is a unit vector then $|\psi\rangle\langle\psi|$, the projection onto $\mathbb{C}\psi$, is a density operator. Such a density operator is called a *vector state* or *pure state*.

By Example 5.3 and Proposition 5.2, if $|\psi_1\rangle\langle\psi_1|, \ldots, |\psi_n\rangle\langle\psi_n|$ are pure states, and $p_1, \ldots, p_r$ is a probability distribution, then

$$\varrho := \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|$$

is again a state. As it turns out, any state can be given in such a form, if we also allow limits:

Theorem 5.4. For any density operator $\varrho \in \mathcal{S}(\mathcal{H})$, there exists an ONB $(e_j)_{j\in J}$ and a probability distribution $(p_j)_{j\in J}$ such that

$$\varrho = \sum_{j\in J} p_j |e_j\rangle\langle e_j| . \tag{5.99}$$

Note that (5.99) implies immediately that the e_j are eigen-vectors of ϱ, with corresponding eigenvalues p_j. Thus, (5.99) gives an eigen-decomposition of ϱ. When the Hilbert space is finite-dimensional, the above theorem follows immediately from the eigen-decomposition theorem for self-adjoint (or, more generally, normal) matrices. The infinite-dimensional case can be seen as a generalization of it, but its proof is beyond the scope of these notes. One thing that should be clarified, though, is the precise meaning of the infinite sum in (5.99) when the Hilbert space is infinite-dimensional. In this case we may choose $J = \mathbb{N}$, and interpret (5.99) as

$$\varrho = \sum_{j=1}^{+\infty} p_j \left|e_j\right\rangle\!\left\langle e_j\right| := \lim_{n\to+\infty} \sum_{j=1}^{n} p_j \left|e_j\right\rangle\!\left\langle e_j\right| .$$

Here, the sum is clearly absolute convergent in operator norm, as

$$\sum_{j=1}^{+\infty} \left\| p_j \left|e_j\right\rangle\!\left\langle e_j\right| \right\| = \sum_{j=1}^{+\infty} p_j \left\| \left|e_j\right\rangle\!\left\langle e_j\right| \right\| = \sum_{j=1}^{+\infty} p_j = 1 < +\infty.$$

In fact, it is also convergent in the stronger trace-norm, by the same computation:

$$\sum_{j=1}^{+\infty} \left\| p_j \left|e_j\right\rangle\!\left\langle e_j\right| \right\|_1 = \sum_{j=1}^{+\infty} p_j \left\| \left|e_j\right\rangle\!\left\langle e_j\right| \right\|_1 = \sum_{j=1}^{+\infty} p_j = 1 < +\infty.$$

Next, we give the mathematical description of measurements in a quantum model:

Definition 5.5. Let $\mathcal{H}$ be a Hilbert-space and $(\mathcal{X}, \mathcal{A})$ be a measurable space. We say that $M : \mathcal{A} \to \mathcal{B}(\mathcal{H})_+$ is a *positive operator-valued measure (POVM)*, if for any sequence $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ of pairwise disjoint $\mathcal{A}$-measurable sets (i.e., $A_n \cap_{n\neq m} A_m = \emptyset$), and any $\psi \in \mathcal{H}$,

$$\left\langle \psi, M\left(\cup_{n\in\mathbb{N}} A_n\right)\psi \right\rangle = \sum_{n=1}^{+\infty} \left\langle \psi, M(A_n)\psi \right\rangle , \qquad (\sigma\text{-additivity}) \tag{5.100}$$

and

$$M(\mathcal{X}) = I \qquad \text{(normalization)} \tag{5.101}$$

We denote the set of such POVMs by $\mathrm{POVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$.

Definition 5.6. We say that a POVM $P \in \mathrm{POVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ is *projective*, or that it is a *projection-valued measure (PVM)*, if $P(A)$ is a projection for all $A \in \mathcal{A}$, i.e., $P(A)^* = P(A) = P(A)^2$. We denote the set of all such PVMs by $\mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$.

PVMs are also called *sharp measurements*.

Remark 5.7. In standard quantum mechanics, one usually only works with projective measurements, i.e., PVMs, while general POVMs play an important role in the mathematical description of open quantum systems, and in quantum information theory. In the introductory treatment of the subject in the present notes, we will also mainly restrict our attention to projective measurements, while we formulate some basic statements for general POVMs when it does not mean any extra effort.

Remark 5.8. A POVM describes a measurement in which the possible outcomes are elements of the set $\mathcal{X}$. This may be $\{\text{up}, \text{down}\}$ when measuring the spin of an electron along some axis, but most often the measurement outcomes are real numbers, or vectors with real components, e.g., when measuring the position of a particle moving in space. For mathematical convenience, we may also consider measurements with complex outcomes. Unless otherwise stated, we will always use the Borel σ-algebra in all these cases, and use the short-hand notation

$$\text{POVM}(\mathcal{H}, \mathbb{K}^n) \qquad \text{instead of} \qquad \text{POVM}(\mathcal{H}, \mathbb{K}^n, \mathcal{B}(\mathbb{K}^n)),$$

where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, and similarly $\text{PVM}(\mathcal{H}, \mathbb{K})$ instead of $\text{PVM}(\mathcal{H}, \mathbb{K}, \mathcal{B}(\mathbb{K}))$. We call a POVM $M \in \text{POVM}(\mathcal{H}, \mathbb{K})$ a *real-valued* or a *complex-valued* POVM/measurement, depending on whether $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

Finally, we need to specify how a state $\varrho \in \mathcal{S}(\mathcal{H})$ and a measurement $M \in \text{POVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ give rise to the measurement statistics. This is given by

$$\mathbb{P}_{\varrho,M}(A) := \operatorname{Tr} \varrho M(A), \qquad A \in \mathcal{A}, \tag{5.102}$$

and is called the *Born rule* in quantum mechanics.

One obvious thing to verify is that (5.102) does indeed define a probability measure on $(\mathcal{X}, A)$. Let us introduce for any vector $\psi \in \mathcal{H}$ and POVM $M \in \text{POVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ the set function

$$M_\psi(.) := \langle \psi, M(.)\psi \rangle \tag{5.103}$$

on $\mathcal{A}$. Then

$$M_\psi(A) = \langle \psi, M(A)\psi \rangle \geq 0, \qquad A \in \mathcal{A},$$

by the assumption that $M(A)$ is positive semi-definite. Moreover, the σ-additivity condition (5.100) expresses exactly that M_ψ is a (positive) measure on the measurable space $(\mathcal{X}, \mathcal{A})$. Finally, the normalization condition (5.101) is equivalent to

$$M_\psi(\mathcal{X}) = \|\psi\|^2,$$

i.e., M_ψ is a finite measure, and it is a probability measure if and only if $\|\psi\| = 1$.

Now, if $\varrho = |\psi\rangle\langle\psi|$ is a pure state then

$$\mathbb{P}_{\psi,M}(A) := \mathbb{P}_{|\psi\rangle\langle\psi|,M}(A) = \operatorname{Tr} |\psi\rangle\langle\psi| \, M(A) = \langle\psi, M(A)\psi\rangle = M_\psi(A),$$

i.e., the probability measure assigned by the Born rule to ϱ and M is exactly M_ψ, which, as we have seen above, is indeed a probability distribution. For a general density operator ϱ, we have a spectral decomposition $\varrho = \sum_{j\in J} p_j \, |e_j\rangle\langle e_j|$, and

$$\mathbb{P}_{\varrho,M}(A) = \operatorname{Tr} \varrho M(A) = \sum_{j\in J} p_j \underbrace{\operatorname{Tr} |e_j\rangle\langle e_j| \, M(A)}_{=\langle e_j, M(A)e_j\rangle} = \underbrace{\sum_{j\in J} p_j M_{e_j}(A)}_{=:M_\varrho(A)},$$

and it is straightforward to verify that M_ϱ is a probability measure on $\mathcal{A}$.

Remark 5.9. There is one very important thing to note about the terminology here. Recall that in a classical model $(\Omega, \mathcal{F})$, states of the system are given by probability distributions on $\mathcal{F}$, and a sharp measurement is represented by a measurable function $f : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ (also called a random variable), that transforms any state of the system into a probability distribution on the outcome space $(\mathcal{X}, \mathcal{A})$, the image of a state $\varrho \in \mathcal{S}(\Omega, \mathcal{F})$ being $\mathbb{P}_{\varrho,f} := f_*\varrho = \varrho \circ f^{-1}$.

By the above, the mathematical object used to describe a measurement in a quantum model is called a positive operator-valued measure, and its defining properties are indeed the same as what we used in the definition of a scalar-valued measure; in fact, if the Hilbert space is 1-dimensional then a POVM $M \in \mathrm{POVM}(\mathbb{C}, \mathcal{X}, \mathcal{A})$ is nothing else than a probability measure on $\mathcal{A}$, and every probability measure on $\mathcal{A}$ is a POVM. The similarities go even further: one can define a notion of integral of complex-valued measurable functions w.r.t. (projective) POVMs, as we will see in Section 5.2.

Nevertheless, one should keep in mind that, in comparison with classical models, POVMs play the role of the random variables, and not the role of the measures. Indeed, we can view any POVM M as a transformation of the states of the quantum system into probability distributions on the outcome space, under which the image of a state ϱ is $\mathbb{P}_{\varrho,M}(.) := \operatorname{Tr} \varrho M(.)$.

Hence, we have the following correspondence between classical and quantum models:

	classical	quantum
defining object	$(\Omega, \mathcal{F})$ measurable space	$\mathcal{H}$ Hilbert space
states	probability distributions on $\mathcal{F}$	density operators on $\mathcal{H}$
real-valued sharp measurements	real-valued measurable functions on $\mathcal{X}$	real-valued PVMs
outcome probabilities	$\mathbb{P}_{\varrho,f}(.) = (\varrho \circ f^{-1})(.)$	$\mathbb{P}_{\varrho,M}(.) = \operatorname{Tr} \varrho M(.)$

Example 5.10. (Position measurement of a $1D$ particle)

In quantum mechanics, a particle moving freely along a line is modeled by the Hilbert space $\mathcal{H} = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$, and λ is the Lebesgue-measure. For any Borel set $A \in \mathcal{B}(\mathbb{R})$, let

$$Q(A) := M_{\mathbf{1}_A} : f \mapsto \mathbf{1}_A f, \qquad f \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda).$$

be the multiplication operator by the characteristic function of A. Then $Q(\mathbb{R}) = M_{\mathbf{1}_\mathbb{R}} = I$, and for any $\psi \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$,

$$Q_\psi(A) = \langle \psi, M(A)\psi \rangle = \int_\mathbb{R} \overline{\psi}(x) \mathbf{1}_A(x) \psi(x)\, dx = \int_A |\psi(x)|^2\, dx.$$

Thus, for any sequence $(A_n)_{n\in\mathbb{N}}$ of pairwise disjoint Borel sets,

$$\begin{aligned}
\sum_{n=1}^{+\infty} Q_\psi(A_n) &= \lim_{n\to+\infty} \sum_{n=1}^{N} Q_\psi(A_n) = \lim_{n\to+\infty} \int_\mathbb{R} |\psi(x)|^2 \sum_{n=1}^{N} \mathbf{1}_{A_n}(x)\, dx \\
&= \int_\mathbb{R} |\psi(x)|^2 \sum_{n=1}^{+\infty} \mathbf{1}_{A_n}(x)\, dx = \int_\mathbb{R} |\psi(x)|^2 \mathbf{1}_{\cup_{n\in\mathbb{N}} A_n}(x)\, dx \\
&= Q_\psi\left(\cup_{n=1}^{+\infty} A_n\right),
\end{aligned}$$

where the third equality follows from the monotone convergence theorem. Hence, the σ-additivity condition (5.100) is satisfied, and therefore Q is a POVM in $\mathcal{H}$ with outcome space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Moreover, $Q(A)^2 = M^2_{\mathbf{1}_A} = M_{\mathbf{1}_A^2} = M_{\mathbf{1}_A} = Q(A)$ implies that it is a projective POVM.

By the above, if the system is in a pure state $\varrho = |\psi\rangle\langle\psi|$, where ψ is called the *wave function* of the system in quantum mechanics, the probability that the position of the particle is between some numbers a and b is given by

$$Q_\psi([a,b]) = \int_{[a,b]} |\psi(x)|^2 \, dx,$$

a familiar formula from quantum mechanics.

Example 5.11. In more generality, for any measure space $(\mathcal{X}, \mathcal{A}, \mu)$, we may consider the *multiplication PVM* $M \in \mathrm{PVM}(L^2(\mathcal{X}, \mathcal{A}, \mu), \mathcal{X}, \mathcal{A})$ as $M(A) := M_{\mathbf{1}_A}$, where the latter is the multiplication operator with the characteristic function of $A \in \mathcal{A}$. For any $\psi \in L^2(\mathcal{X}, \mathcal{A}, \mu)$,

$$M_\psi(A) = \langle \psi, M(A)\psi \rangle = \int_{\mathcal{X}} \overline{\psi} \mathbf{1}_A \psi \, d\mu = \int_A |\psi|^2 \, d\mu = (|\psi|^2 \mu)(A),$$

and the same argument as in the previous example yields that M is a PVM.

Example 5.12. Gibbs states of a harmonic oscillator.

Example 5.13. Spin measurement.

As we have seen in Example 5.12, the only measurement outcomes that we may obtain with non-zero probability in any state ϱ are of the form $\hbar\omega(n+1/2)$ for some $n \in \mathbb{N}$. This example motivates to introduce the following useful concept:

Definition 5.14. Consider a real- or complex-valued POVM $M \in \mathrm{POVM}(\mathcal{H}, \mathbb{K})$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. A point $z \in \mathbb{K}$ is in the *support* of M if for any $\varepsilon > 0$, $M(\{w \in \mathbb{K} : |w - z| < \varepsilon\}) \neq 0$. The collection of all such points is the support of M, denoted by $\operatorname{supp} M$.

A POVM is *finitely supported* if $|\operatorname{supp} M| < +\infty$.

Example 5.15. The support of the position measurement in Example 5.10 is $\mathbb{R}$, the support of the energy measurement in Example 5.12 is $\{\hbar\omega(n + 1/2) : n \in \mathbb{N}\}$. We may also consider the spin measurement in Example 5.13 as a real-valued measurement with support $\{\pm 1\}$; in particular, it is finitely supported.

Exercise 5.16. Show that the support is the smallest closed set $K \subseteq \mathbb{K}$ such that $M(K) = I$, or equivalently, the complement of the largest open set $G \subseteq \mathbb{K}$ with $M(G) = 0$. Show that for any $A \in \mathcal{B}(\mathbb{K})$,

$$M(A) = M(A \cap \operatorname{supp} M).$$

Exercise 5.17. Show that M is finitely supported if and only if there exists a finite set $S \subseteq \mathbb{K}$ such that $M(\{x\}) \neq 0$ for all $x \in S$, and $\sum_{x \in S} M(\{x\}) = I$. Show that in this case $S = \operatorname{supp} M$, and for any $A \in \mathcal{B}(\mathbb{K})$,

$$M(A) = \sum_{x \in S \cap A} M(\{x\}).$$

Vice versa, show that any finite collection of PSD operators $(M_x)_{x \in S}$, where $S \subseteq \mathbb{K}$ is a finite set, and $\sum_{x \in S} M_x = I$, defines a POVM M via

$$M(A) := \sum_{x \in A} M_x,$$

and $\operatorname{supp} M \subseteq S$.

Remark 5.18. Alternatively, we may define a finite-outcome real or complex measurement as a collection of operators $(M_x)_{x \in \mathbb{K}}$, such that $|\{x \in \mathbb{K} : M_x \neq 0\}| < +\infty$, and $\sum_{x \in \mathbb{K}} M_x = I$. Here, the sum is well-defined as only finitely many terms are non-zero.

Imagine now that we are not interested in the outcomes of some measurement $M \in \mathrm{POVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ directly, but in some function f of the measurement outcomes, where $f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ is a measurable function into another measurable space $(\mathcal{Y}, \mathcal{B})$. The measurement statistics of this function of the measurement outcomes in any state $\varrho \in \mathcal{S}(\mathcal{H})$ of the system is then given by

$$\operatorname{Tr} \varrho M\left(\{x \in \mathcal{X} : f(x) \in B\}\right) = \operatorname{Tr} \varrho (M \circ f^{-1})(B) = \mathbb{P}_{\varrho, M \circ f^{-1}}(B), \qquad B \in \mathcal{B}.$$

It is straightforward to verify that $M \circ f^{-1}$ is again a POVM, and we make the following

Definition 5.19. In the above setting, the POVM

$$f_* M := M \circ f^{-1} \in \mathrm{POVM}(\mathcal{H}, \mathcal{Y}, \mathcal{B})$$

is called the *classical post-processing of the POVM M by the function f*.

Remark 5.20. Alternatively, we say that $f_* M$ is "the f of the POVM M"; e.g., if M is real-valued, and $f = \mathrm{id}_{\mathbb{R}}^2 : x \mapsto x^2$, $x \in \mathbb{R}$, then we say that $(\mathrm{id}_{\mathbb{R}}^2)_* M$ is the square of the POVM M. Note that this is different from taking the square of the measurement operators, i.e., $M(A)^2$, $A \in \mathcal{A}$. For one thing, this is a set function on a different σ-algebra unless $(\mathcal{X}, \mathcal{A}) = (\mathcal{Y}, \mathcal{B})$, and, moreover, it is not a POVM unless M is projective.

In yet another terminology, $f_* M$ is called the *push-forward* of the POVM M by the function f.

Example 5.21. Imagine that in the setting of Example 5.10, we are not interested in the position of the particle, but only in its distance from the origin, i.e., the absolute value of its position. This can be described by choosing $f(.) = |.|$, and POVM

$$\begin{aligned}(Q \circ |.|^{-1})(A) &= Q\left((A \cap \mathbb{R}_+) \cup (-(A \cap \mathbb{R}_+))\right) = M_{\mathbf{1}_{(A\cap\mathbb{R}_+)\cup(-(A\cap\mathbb{R}_+))}} \\ &= Q(A \cap \mathbb{R}_+) + M(-(A \cap \mathbb{R}_+)) = M_{\mathbf{1}_{A\cap\mathbb{R}_+}} + M_{\mathbf{1}_{-(A\cap\mathbb{R}_+)}},\end{aligned}$$

where $-(A \cap \mathbb{R}_+) := \{-x : x \in A \cap \mathbb{R}_+\}$. Hence, for any wave function ψ, the probability of finding the particle in a Borel set $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}\mathbb{P}_{|\psi\rangle\langle\psi|,Q\circ|.|^{-1}}(A) &= \int_{A\cap\mathbb{R}_+} |\psi(x)|^2\,dx + \int_{-(A\cap\mathbb{R}_+)} |\psi(x)|^2\,dx \\ &= \int_{A\cap\mathbb{R}_+} \left(|\psi(x)|^2 + |\psi(-x)|^2\right) dx.\end{aligned}$$

Example 5.22. As a generalization of the previous example, we may consider the multiplication PVM $M \in \mathrm{POVM}(L^2(\mathcal{X}, \mathcal{A}, \mu), \mathcal{X}, \mathcal{A})$, $M(A) := M_{\mathbf{1}_A}$, $A \in \mathcal{A}$, from Example 5.11. For any measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}$,

$$(f_* M)(B) = M(f^{-1}(B)) = M_{\mathbf{1}_{f^{-1}(B)}}, \qquad B \in \mathcal{B}.$$

Hence, for any $\psi \in L^2(\mathcal{X}, \mathcal{A}, \mu)$,

$$\mathbb{P}_{|\psi\rangle\langle\psi|,f_*M}(B) = \int_{f^{-1}(B)} |\psi(x)|^2\,d\mu(x).$$

We close this section with some general observations about POVMs.

Remark 5.23. Note that σ-additivity implies, by choosing $A_n = \emptyset$ for all $n \in \mathbb{N}$, that

$$\langle\psi, M(\emptyset)\psi\rangle = \sum_{n=1}^{+\infty} \langle\psi, M(\emptyset)\psi\rangle\,,$$

whence $\langle\psi, M(\emptyset)\psi\rangle = 0$. Since this holds for all $\psi \in \mathcal{H}$, we have

$$M(\emptyset) = 0.$$

This in turn shows that σ-additivity implies finite additivity: If $A_1, \ldots, A_m \in \mathcal{A}$, are pairwise disjoint, then

$$\begin{aligned}\langle\psi, M(A_1 \cup \ldots \cup A_m)\psi\rangle &= \left\langle\psi, M\left((A_1 \cup \ldots \cup A_m) \cup \left(\cup_{n=m+1}^{+\infty}\emptyset\right)\right)\psi\right\rangle \\ &= \sum_{n=1}^{m} \langle\psi, M(A_n)\psi\rangle + \sum_{n=m+1}^{+\infty} \underbrace{\langle\psi, M(\emptyset\rangle\,\psi}_{=0} \\ &= \sum_{n=1}^{m} \langle\psi, M(A_n)\psi\rangle = \left\langle\psi, \left(\sum_{n=1}^{m} M(A_n)\right)\psi\right\rangle.\end{aligned}$$

Since this holds for every $\psi \in \mathcal{H}$, we finally see (due to the polarization identity) that

$$M(A_1 \cup \ldots \cup A_m) = M(A_1) + \ldots + M(A_m).$$

In particular, for any $A \in \mathcal{A}$,

$$I = M(\mathcal{X}) = M(A \cup (\mathcal{X} \setminus A)) = M(A) + M(\mathcal{X} \setminus A),$$

and hence

$$M(\mathcal{X} \setminus A) = I - M(A).$$

It follows immediately that a PVM M is *monotone*, in the sense that if $A, B \in \mathcal{A}$ and $A \supseteq B$ then

$$M(A) = M((A \setminus B) \cup B) = M(A \setminus B) + M(B) \geq M(B).$$

Remark 5.24. Note that if $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ of pairwise disjoint $\mathcal{A}$-measurable sets then $S_N := \sum_{n=1}^{N} M(A_n)$ is a monotone increasing sequence of PSD operators that is bounded as

$$\sum_{n=1}^{N} M(A_n) = M(\mathcal{X}) - M\left(\mathcal{X} \setminus \cup_{n=1}^{N} A_n\right) \leq M(\mathcal{X}) = I.$$

Hence, it is convergent in the strong, and therefore also in the weak, operator topologies, and (5.100) tells that

$$M\left(\cup_{n\in\mathbb{N}} A_n\right) = (\text{wo}) \sum_{n=1}^{+\infty} M(A_n) = (\text{so}) \sum_{n=1}^{+\infty} M(A_n),$$

where the latter can be equivalently written as

$$M\left(\cup_{n\in\mathbb{N}} A_n\right)\psi = \sum_{n=1}^{+\infty} M(A_n)\psi, \qquad \psi \in \mathcal{H}, \tag{5.104}$$

with the sum on the RHS converging in the norm of the Hilbert space.

It is easy to see that S_N does not converge in operator norm in general. Indeed, if M is a projective POVM and $M(A_n) \neq 0$ for all $n \in \mathbb{N}$, then $M(\cup_{n\in\mathbb{N}} A_n) - M(\cup_{n=1}^{N} A_n)$ is a non-zero projection, and thus

$$\left\| M(\cup_{n\in\mathbb{N}} A_n) - M(\cup_{n=1}^{N} A_n) \right\| = 1.$$

Such an example is easily constructed e.g., by taking the position measurement $M = Q$ from example 5.10, and $A_n = [n, n+1)$, $n \in \mathbb{N}$.

The following property of the distribution M_ψ will be important later:

Exercise 5.25. Show the following parallelogram rule for the distributions induced by a POVM $M \in \mathrm{POVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$: For any $\psi_1, \psi_2 \in \mathcal{H}$,

$$M_{\psi_1+\psi_2} + M_{\psi_1-\psi_2} = 2M_{\psi_1} + 2M_{\psi_2}. \tag{5.105}$$

Conclude that

$$M_{\psi_1+\psi_2} \leq 2M_{\psi_1} + 2M_{\psi_2}. \tag{5.106}$$

(Both (5.105) and (5.106) should be interpreted by substitution of an arbitrary $A \in \mathcal{A}$.

Solution:

By definition, for any $A \in \mathcal{A}$,

$$\begin{aligned} M_{\psi_1 \pm \psi_2}(A) &= \langle \psi_1 + \psi_2, M(A)(\psi_1 + \psi_2) \rangle \\ &= \underbrace{\langle \psi_1, M(A)\psi_1 \rangle}_{=M_{\psi_1}(A)} + \underbrace{\langle \psi_2, M(A)\psi_2 \rangle}_{=M_{\psi_2}(A)} \pm \langle \psi_1, M(A)\psi_2 \rangle \pm \langle \psi_2, M(A)\psi_1 \rangle , \end{aligned}$$

from which (5.105) follows immediately. (5.106) is obvious from (5.105), as $M_{\psi_1-\psi_2}(A) \geq 0$ for all $A \in \mathcal{A}$.

Remark 5.26. As a generalization of (5.103), we may define for any two vectors $\psi_1, \psi_2 \in \mathcal{H}$,

$$M_{\psi_1,\psi_2}(.) := \langle \psi_1, M(.)\psi_2 \rangle . \tag{5.107}$$

The polarization formula implies that for any sequence $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ of pairwise disjoint $\mathcal{A}$-measurable sets,

$$\begin{aligned} M_{\psi_1,\psi_2}\left(\cup_{n\in\mathbb{N}} A_n\right) &= \langle \psi_1, M\left(\cup_{n\in\mathbb{N}} A_n\right) \psi_2 \rangle \\ &= \frac{1}{4} \sum_{k=0}^{3} i^k \left\langle i^k\psi_1 + \psi_2, M\left(\cup_{n\in\mathbb{N}} A_n\right) (i^k\psi_1 + \psi_2) \right\rangle \\ &= \frac{1}{4} \sum_{k=0}^{3} i^k \sum_{n=1}^{+\infty} \left\langle i^k\psi_1 + \psi_2, M\left(A_n\right) (i^k\psi_1 + \psi_2) \right\rangle \\ &= \sum_{n=1}^{+\infty} \frac{1}{4} \sum_{k=0}^{3} i^k \left\langle i^k\psi_1 + \psi_2, M\left(A_n\right) (i^k\psi_1 + \psi_2) \right\rangle \\ &= \sum_{n=1}^{+\infty} \langle \psi_1, M(A_n)\psi_2 \rangle \\ &= \sum_{n=1}^{+\infty} M_{\psi_1,\psi_2}(A_n), \end{aligned}$$

i.e., M_{ψ_1,ψ_2} is a *complex measure*, and

$$M_{\psi_1,\psi_2} = \frac{1}{4}\sum_{k=0}^{3} i^k M_{i^k\psi_1+\psi_2},$$

where we used the short-hand notation

$$M_{\psi,\psi} := M_\psi.$$

Exercise 5.27. Show that for any $\psi_1, \psi_2 \in \mathcal{H}$ and $\lambda, \eta \in \mathbb{C}$,

$$M_{\lambda\psi_1,\eta\psi_2} = \overline{\lambda}\eta M_{\psi_1,\psi_2}.$$

Conclude that for any $\psi \in \mathcal{H}$ and $\lambda \in \mathbb{C}$,

$$M_{\lambda\psi} = |\lambda|^2 M_\psi. \tag{5.108}$$

The following simple characterization of projective POVMs and its consequences will play an important role later.

Lemma 5.28. Let $P \in \mathrm{POVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ be a POVM. The following are equivalent:

(i) P is projective.

(ii) $P(A) \perp P(B)$ for every $A, B \in \mathcal{A}$ such that $A \cap B = \emptyset$.

(iii) $P(A) \perp P(\mathcal{X} \setminus A)$ for every $A \in \mathcal{A}$.

Proof. (i) (i)$\Longrightarrow$(ii): Assume that $P(A) \not\perp P(B)$, i.e., that there exists a non-zero vector $\psi \in \operatorname{ran} P(A)$ such that $P(B)\psi \neq 0$, and hence $0 < \|P(B)\psi\|^2 = \langle P(B)\psi, P(B)\psi\rangle = \langle \psi, P(B)\psi\rangle$, where we used the projectivity of P in the last step. We have $P(A) + P(B) = I - P(\mathcal{X} \setminus (A \cup B)) \leq I$, and hence

$$\|\psi\|^2 = \langle\psi, I\psi\rangle \geq \underbrace{\langle\psi, P(A)\psi\rangle}_{=\|\psi\|^2} + \underbrace{\langle\psi, P(B)\psi\rangle}_{>0} > \|\psi\|^2,$$

a contradiction.

(ii)$\Longrightarrow$(iii) is trivial.

(iii)$\Longrightarrow$(i): We have $I = P(A) + P(\mathcal{X} \setminus A)$, and hence

$$I = I^2 = P(A)^2 + P(\mathcal{X} \setminus A)^2 + \underbrace{P(A)P(\mathcal{X} \setminus A)}_{=0} + \underbrace{P(\mathcal{X} \setminus A)P(A)}_{=0}.$$

Therefore,

$$\begin{aligned} 0 = I - I &= P(A) + P(\mathcal{X} \setminus A) - \big(P(A)^2 + P(\mathcal{X} \setminus A)^2\big) \\ &= \underbrace{\big[P(A) - P(A)^2\big]}_{\geq 0} + \underbrace{\big[P(\mathcal{X} \setminus A) - P(\mathcal{X} \setminus A)^2\big]}_{\geq 0} \geq 0, \end{aligned}$$

whence $P(A) - P(A)^2 = 0$.

□

Exercise 5.29. Let $P \in \mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ be a PVM. Show that for any $A, B \in \mathcal{A}$,

$$P(A \cap B) = P(A)P(B), \tag{5.109}$$
$$P(A \cup B) = P(A) + P(B) - P(A)P(B). \tag{5.110}$$

(Hint: Use Lemma 5.28.)

Solution:

We have $P(A) = P(A \setminus B) + P(A \cap B)$, and $P(B) = P(B \setminus A) + P(A \cap B)$, and hence

$$\begin{aligned} P(A)P(B) &= \underbrace{P(A \setminus B)P(B \setminus A)}_{=0} + \underbrace{P(A \setminus B)P(A \cap B)}_{=0} + \underbrace{P(A \cap B)P(B \setminus A)}_{=0} + P(A \cap B)^2 \\ &= P(A \cap B), \end{aligned}$$

where the nullity of the products follow from the respective sets being disjoint, according to Lemma 5.28. This proves (5.109), and (5.110) follows as

$$\begin{aligned} P(A \cup B) &= P(A \setminus (A \cap B)) + P(B \setminus (A \cap B)) + P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B), \end{aligned}$$

where in the last step we used (5.109).

The intersection property (5.109) yields immediately the following:

Lemma 5.30. For any $A \in \mathcal{A}$ and $\psi \in \mathcal{H}$,

$$P_{P(A)\psi} = P_\psi\big|_A, \quad \text{i.e.,} \quad P_{P(A)\psi}(B) = P_\psi(A \cap B), \quad B \in \mathcal{A}.$$

Proof.

$$\begin{aligned} P_{P(A)\psi}(B) &= \langle P(A)\psi, P(B)P(A)\psi\rangle = \langle \psi, P(B)P(A)P(B)\psi\rangle \\ &= \langle \psi, P(A \cap B)\psi\rangle = P_\psi(A \cap B), \end{aligned}$$

where the last equality follows from (5.109). □

Remark 5.31. The multiplicativity property (5.109) might seem quite surprising first, when compared to scalar-valued probability measures. Note, however, that scalar-valued probability measures correspond to POVMs on the one-dimensional Hilbert space $\mathbb{C}$, and projectivity of such a probability measure μ means that $\mu(A)^2 = \mu(A)$, and hence $\mu(A) = 0$ or $\mu(A) = 1$ for all $A \in \mathcal{A}$. In particular, if all singletons $\{x\}$, $x \in \mathcal{X}$, are measurable, then a probability measure μ on $\mathcal{A}$ is projective if and only if it is a Dirac measure $\mu = \delta_x$ concentrated at some point x_0, in which case $\delta_{x_0}(A \cap B) = \delta_{x_0}(A)\delta_{x_0}(B)$ holds trivially.

Thus, projective measures are very simple when the Hilbert space is 1-dimensional, and it is only in higher dimensions that we get a rich theory of projective measures.

5.2 Observables as operators

Let us consider a real-valued measurement on a Hilbert space $\mathcal{H}$, described by some POVM $M \in \mathrm{POVM}(\mathcal{H}, \mathbb{R})$. Then we may define the expectation value of the measurement outcomes in a state ϱ as

$$\mathbb{E}_\varrho(M) := \int_{\mathbb{R}} x \, dM_\varrho(x) = \int_{\mathbb{R}} x \operatorname{Tr} \varrho M(dx),$$

provided that the integral exists. Linearity of the trace then suggests to rewrite this as

$$\mathbb{E}_\varrho(M) = \operatorname{Tr} \varrho \left(\int_{\mathbb{R}} x M(dx) \right) = \operatorname{Tr} \varrho \hat{M},$$

where we introduce the formal operator

$$\hat{M} := \int_{\mathbb{R}} x M(dx). \tag{5.111}$$

In particular, for a pure state $|\psi\rangle\langle\psi|$ this would give

$$\mathbb{E}_{|\psi\rangle\langle\psi|, M} = \left\langle \psi, \hat{M}\psi \right\rangle, \tag{5.112}$$

the familiar formula from quantum mechanics for the expectation value of a physical observable described by a (self-adjoint) operator $\hat{M}$.

Our goal in this section is to give a precise mathematical meaning to the formal expression in (5.111), and thereby make the connection between the formalism introduced in Section 5.1, and the usual formalism of standard physics textbooks on quantum mechanics.

We start with the simple but instructive example of a finite-outcome real measurement $M \in \mathrm{POVM}(\mathcal{H}, \mathbb{R})$, given by $(M_x)_{x \in \mathbb{R}}$, where $M_x \in \mathcal{B}(\mathcal{H})_+$ for all $x \in \mathcal{X}$,

$M_x = 0$ for all but finitely many $x \in \mathbb{R}$, and $\sum_{x\in\mathbb{R}} M_x = I$. Given a state $\varrho \in \mathcal{S}(\mathcal{H})$, the expectation value of the measurement outcomes in the state ϱ is then given by

$$\mathbb{E}_\varrho(M) = \sum_{x\in\mathbb{R}} x\, \mathbb{P}_{\varrho,M}(x) = \sum_{x\in\mathbb{R}} x \operatorname{Tr} \varrho M_x = \operatorname{Tr} \varrho \sum_{x\in\mathbb{R}} x M_x = \operatorname{Tr} \varrho \widehat{M}, \tag{5.113}$$

where

$$\widehat{M} := \sum_{x\in\mathbb{R}} x M_x \tag{5.114}$$

is a bounded self-adjoint operator on $\mathcal{H}$, i.e., $\widehat{M} \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$.

Next, we consider the *variance* of the outcomes of a finite-outcome real-valued measurement M in a state ϱ, given by

$$\begin{aligned}\mathbb{V}_\varrho(M) &:= \sum_{x\in\mathbb{R}} (x - \mathbb{E}_\varrho(M))^2\, \mathbb{P}_{\varrho,M}(x) = \sum_{x\in\mathbb{R}} x^2\, \mathbb{P}_{M,\varrho}(x) - (\mathbb{E}_\varrho(M))^2 \\ &= \sum_{x\in\mathbb{R}} x^2 \operatorname{Tr} \varrho M_x - (\mathbb{E}_\varrho(M))^2 = \operatorname{Tr} \varrho \sum_{x\in\mathbb{R}} x^2 M_x - \left(\operatorname{Tr} \varrho \widehat{M}\right)^2.\end{aligned}$$

We can also consider higher moments, defined for all $k \in \mathbb{N}$ as

$$m_k(\varrho, M) := \sum_{x\in\mathbb{R}} x^k\, \mathbb{P}_{\varrho,M}(x) = \operatorname{Tr} \varrho \sum_{x\in\mathbb{R}} x^k M_x.$$

Now we can make the following observation: If M is projective then $M_x M_y = 0$ for all $x \neq y$, and hence $\widehat{M}^2 = \sum_{x\in\mathbb{R}} x^2 M_x$. Thus,

$$\mathbb{V}_\varrho(M) = \operatorname{Tr} \varrho \widehat{M}^2 - (\operatorname{Tr} \varrho \widehat{M})^2. \tag{5.115}$$

For a pure state $\varrho = |\psi\rangle\langle\psi|$, we get

$$\mathbb{V}_{|\psi\rangle\langle\psi|}(M) = \left\langle \psi, \widehat{M}^2 \psi \right\rangle - \left\langle \psi, \widehat{M}\psi \right\rangle^2, \tag{5.116}$$

again a familiar expression from standard quantum mechanics. More generally, for all $k \in \mathbb{N}$,

$$m_k(\varrho, M) = \operatorname{Tr} \varrho \widehat{M}^k. \tag{5.117}$$

That is, all moments, and hence, the complete measurement statistics of M in any state ϱ is completely determined by the self-adjoint operator $\widehat{M}$. In fact, we can uniquely recover the original PVM M from $\widehat{M}$, as the following exercise shows:

Exercise 5.32. Let $M = (M_x)_{x\in\mathbb{R}} \in \mathrm{PVM}(\mathcal{H}, \mathbb{R})$ be a finite-outcome real-valued projective measurement, and $\widehat{M} := \sum_{x\in\mathbb{R}} xM_x$. Show that

$$M_x \neq 0 \iff x \text{ is an eigenvalue of } \widehat{M},$$

and for all such x, M_x is the projection onto the eigen-subspace corresponding to x.

Remark 5.33. It is easy to see that if M is not projective then its measurement statistics cannot be recovered from a single operator $\hat{M}$ in the sense that (5.117) would hold. In fact, the requirement that

$$\mathbb{E}_{|\psi\rangle\langle\psi|}(M) = \mathrm{Tr}\,|\psi\rangle\langle\psi|\,\hat{M} = \left\langle \psi, \hat{M}\psi \right\rangle$$

holds for every unit vector $\psi \in \mathcal{H}$ specifies $\hat{M}$ uniquely, and if M is not projective then there always exists a pure state ψ such that (5.116) does not hold. For this reason, we will restrict our considerations to projective measurements for the rest.

Motivated by the above observations, we introduce the following notion:

Definition 5.34. We say that an operator $T \in \mathcal{B}(\mathcal{H})$ is *simple* if there exist finitely many projections $P_1, \ldots, P_r$ such that $\sum_{k=1}^r P_k = I$, and complex numbers $z_1, \ldots, z_r$, such that $T = \sum_{k=1}^r z_k P_k$.

The following exercise establishes a bijection between finite-outcome complex-valued projective measurements and simple operators.

Exercise 5.35. Let $T \in \mathcal{B}(\mathcal{H})$. Show that the following are equivalent:

(i) T is simple.

(ii) There exists a finite-outcome complex-valued PVM $M = (M_x)_{x\in\mathcal{X}} \in \mathrm{PVM}_f(\mathcal{H}, \mathbb{C})$ such that

$$T = \widehat{M} := \sum_{z\in\mathbb{C}} zM_z.$$

(iii) T is normal, it has finitely many distinct eigenvalues $z_1, \ldots, z_r$, and the projections P_k onto the eigen-subspaces corresponding to the z_k satisfy $\sum_{k=1}^r P_k = I$.

Show that the map $M \mapsto \widehat{M}$ gives a bijection between finitely supported complex PVMs and simple operators on $\mathcal{H}$, under which real PVMs correspond to self-adjoint operators.

Now we move on to the case of general projective measurements. Our main goal is to make sense of the integral (5.111), and to establish the one-to-one correspondence between real-valued PVMs and self-adjoint operators that we may anticipate based on the bijection between finitely supported real PVMs and simple self-adjoint operators.

We will start with the more general task of defining an operator $\int f\,dP$ for any PVM $P \in \mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$, and measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$. We will follow the same approach as in measure theory, first defining the integral for simple functions, and then extend it to more general measurable functions by approximating arguments.

For the rest of the section, we fix a PVM $P \in \mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$, and always work with this fixed PVM, unless otherwise stated.

For a measurable set $A \in \mathcal{A}$, let us introduce the notations

$$\int \mathbf{1}_A\,dP := P(\mathbf{1}_A) := P(A).$$

Recall that a measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ is *simple* if there exist pairwise disjoint measurable sets $A_1, \ldots, A_m \in \mathcal{A}$, $A_n \cap_{n \neq m} A_m = \emptyset$, and constants $c_1, \ldots, c_m \in \mathbb{C}$, such that $f = \sum_{k=1}^m c_k \mathbf{1}_{A_k}$. For any such function, we define

$$\int f\,dP := P(f) := \sum_{k=1}^m c_k P(A_k),$$

and call it the integral of f w.r.t. P. The following properties are easy to verify:

Exercise 5.36. Show that if f, g are simple measurable functions, and $\lambda, \eta \in \mathbb{C}$, then

- $P(\lambda f + \eta g) = \lambda P(f) + \eta P(g)$ (linearity)
- $P(fg) = P(f)P(g)$ (multiplicativity)
- $P(\overline{f}) = P(f)^*$
- $f \geq 0 \iff P(f) \geq 0$ (positivity)
- $\|P(f)\psi\|^2 = \int |f|^2\,dP_\psi = \|f\|^2_{L^2(P_\psi)}$.

Remark 5.37. The set of simple measurable functions $\mathcal{M}_{\mathrm{symp}}(\mathcal{X}, \mathcal{A}, \mathbb{C})$ is a *-algebra with the usual pointwise operations. The first three properties in Exercise 5.36 can be summarized as saying that the map $f \mapsto P(f)$ is a *-algebra morphism.

Remark 5.38. The multiplicativity property

$$\int fg\,dP = P(fg) = P(f)P(g) = \int f\,dP \int g\,dP$$

might seem surprising when compared to the usual integral theory of functions w.r.t. scalar-valued measures, where it does not hold in general. As we have already pointed out in Remark 5.31, the resolution of this apparent contradiction is that scalar-valued probability measures correspond to POVMs on a one-dimensional Hilbert space, and in this case the PVMs are exactly the Dirac measures (if all singletons are measurable), for which the integral is multiplicative:

$$\int fg\,d\delta_{x_0} = (fg)(x_0) = f(x_0)g(x_0) = \int f\,d\delta_{x_0} \int g\,d\delta_{x_0}.$$

Lemma 5.39. For a measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$, the following are equivalent:

(i) $f \in L^2(\mathcal{X}, \mathcal{A}, P_\psi)$, i.e., $\int |f|^2\,dP_\psi < +\infty$.

(ii) For any sequence $(f_n)_{n\in\mathbb{N}}$ of simple measurable functions such that $f_n \to f$ pointwise, and $|f_n| \leq |f|$, $n \in \mathbb{N}$, the sequence $(P(f_n)\psi)_{n\in\mathbb{N}}$ is convergent.

(iii) For some sequence $(f_n)_{n\in\mathbb{N}}$ of simple measurable functions such that $f_n \to f$ pointwise, and $|f_1| \leq |f_2| \leq \ldots \leq |f|$, the sequence $(P(f_n)\psi)_{n\in\mathbb{N}}$ is convergent.

Moreover, if any (and hence all) of the above holds, then

$$P(f)\psi := \lim_{n\to+\infty} P(f_n)\psi \tag{5.118}$$

is the same for any sequence $(f_n)_{n\in\mathbb{N}}$ as in (ii) or in (iii), and

$$\|P(f)\psi\|^2 = \int |f|^2\,dP_\psi = \|f\|_{L^2(P_\psi)}\,. \tag{5.119}$$

Proof. (i)$\Longrightarrow$(ii): We have

$$\|P(f_n)\psi - P(f_m)\psi\|^2 = \|P(f_n - f_m)\psi\|^2 = \int |f_n - f_m|^2\,dP_\psi \xrightarrow[n,m\to+\infty]{} 0,$$

where the first two equalities follow from the properties established in Exercise 5.36, and in the last step we used the Lebesgue dominated convergence theorem, which is applicable as $4|f|^2$ is an integrable dominating function, since $|f_n - f_m|^2 \leq (|f_n| + |f_m|)^2 \leq (2|f|)^2 = 4|f|^2$. Thus, the sequence $(P(f_n)\psi)_{n\in\mathbb{N}}$ is Cauchy, and hence it is convergent.

(ii)$\Longrightarrow$(iii) is trivial.
(iii)$\Longrightarrow$(i): Let $P(f)\psi := \lim_{n\to+\infty} P(f_n)\psi$. Then

$$\begin{array}{rcl} \lim\limits_{n\to+\infty} \|P(f_n)\psi\|^2 & = & \|P(f)\psi\|^2 \\ \| & & \\ \lim\limits_{n\to+\infty} \int |f_n|^2\, dP_\psi & = & \int |f|^2\, dP_\psi, \end{array} \tag{5.120}$$

where the last equality follows by the monotone convergence theorem. Thus, $\int |f|^2\, dP_\psi = \|P(f)\psi\|^2 < +\infty$.

The independence of the limit of the approximating sequence follows as if $(f_n)_{n\in\mathbb{N}}$ and $(\tilde{f}_n)_{n\in\mathbb{N}}$ are two approximating sequences as in (ii), then $\hat{f}_{2n-1} := f_n$, $\hat{f}_{2n} := \tilde{f}_n$, $n \in \mathbb{N}$, is again an approximating sequence as in (ii), and both $(P(f_n)\psi)_{n\in\mathbb{N}}$, and $(P(\tilde{f}_n)\psi)_{n\in\mathbb{N}}$ are subsequences of $(P(\hat{f}_n)\psi)_{n\in\mathbb{N}}$, and therefore they have to have the same limits. Finally, (5.119) is immediate from (5.120). □

Definition 5.40. For a measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$, we define

$$\mathcal{D}(P(f)) := \{\psi \in \mathcal{H} :\, f \in L^2(P_\psi)\}, \qquad \text{and} \qquad P(f) :\, \mathcal{D}(P(f)) \to \mathcal{H},$$

where $P(f)\psi$ is defined for any $\psi \in \mathcal{D}(P(f))$ as in (5.118). We call $P(f)$ the *integral of f w.r.t P*, and denote it also as

$$\int f\, dP := P(f).$$

Proposition 5.41. For any measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$, $\mathcal{D}(P(f))$ is a dense subspace, and $P(f)$ is a linear operator on it.

Proof. For every $n \in \mathbb{N}$, let $A_n := \{n-1 \le |f| < n\} \in \mathcal{A}$, and $B_N := \cup_{n=1}^N A_n = \{0 \le |f| < N\}$. Then $\cup_{n=1}^{+\infty} A_n = \mathcal{X}$, and thus, for any $\psi \in \mathcal{H}$,

$$\psi_N := P(B_N)\psi = \sum_{n=1}^{N} P(A_n)\psi \xrightarrow[n\to+\infty]{} P(\mathcal{X})\psi = \psi, \tag{5.121}$$

according to (5.104). For any $N \in \mathbb{N}$, we have

$$\int_{\mathcal{X}} |f|^2\, dP_{\psi_N} = \int_{\mathcal{X}} |f|^2\, dP_\psi\big|_{B_n} = \int_{B_N} |f|^2\, dP_\psi \le \int_{B_N} N^2\, dP_\psi = N^2 < +\infty,$$

where we used Lemma 5.30 in the first equality, and hence $\psi_N \in \mathcal{D}(P(f))$. Thus, by (5.121), any vector $\psi \in \mathcal{H}$ can be approximated in norm by vectors in $\mathcal{D}(P(f))$, i.e., $\mathcal{D}(P(f))$ is dense.

If $\psi \in \mathcal{D}(P(f))$ and $\lambda \in \mathbb{C}$ then by (5.108),

$$\int |f|^2 \, dP_{\lambda\psi} = |\lambda|^2 \int |f|^2 \, dP_\psi < +\infty,$$

and hence $\lambda\psi \in \mathcal{D}(P(f))$. If $\psi_1, \psi_2 \in \mathcal{D}(P(f))$ then

$$\int |f|^2 \, dP_{\psi_1+\psi_2} \leq 2 \int |f|^2 \, d(P_{\psi_1} + P_{\psi_2}) = 2 \int |f|^2 \, dP_{\psi_1} + 2 \int |f|^2 \, dP_{\psi_2} < +\infty,$$

where the inequality is due to (5.106), and thus $\psi_1 + \psi_2 \in \mathcal{D}(P(f))$. Therefore, $\mathcal{D}(\mathcal{P}(f))$ is a linear subspace of $\mathcal{H}$.

Finally, if $(f_n)_{n\in\mathbb{N}}$ is any sequence of functions as in Lemma 5.39, then for any $\psi_1, \psi_2 \in \mathcal{D}(P(f))$,

$$\begin{aligned} P(f)(\psi_1 + \psi_2) &= \lim_{n\to+\infty} P(f_n)(\psi_1 + \psi_2) = \lim_{n\to+\infty} P(f_n)\psi_1 + \lim_{n\to+\infty} P(f_n)\psi_2 \\ &= P(f)\psi_1 + P(f)\psi_2, \end{aligned}$$

and $P(f)(\lambda\psi) = \lambda P(f)\psi$ follows similarly for any $\psi \in \mathcal{D}(P(f))$ and $\lambda \in \mathbb{C}$. □

Definition 5.42. For a PVM $P \in \mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$, the map $f \mapsto P(f)$ on the *-algebra of complex-valued measurable functions on $(\mathcal{X}, \mathcal{A})$ is called the *functional calculus* for P.

We will explore the properties of this calculus a bit later, but before that, let us briefly return to the problem of assigning an operator to a real (or, more generally, complex) valued PVM, as outlined at the beginning of this section.

We start with the following:

Proposition 5.43. For any measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$, and any $\psi \in \mathcal{D}(P(f))$, we have $f \in L^1(P_\psi)$, and

$$\langle \psi, P(f)\psi \rangle = \int f \, dP_\psi. \tag{5.122}$$

Proof. Since P_ψ is a finite measure, $L^2(P_\psi) \subseteq L^1(P_\psi)$ due to the Cauchy-Schwarz inequality:

$$\int |f| \, dP_\psi = \int |f| \cdot 1 \, dP_\psi \leq \left(\int |f|^2 \, dP_\psi \right)^{1/2} \underbrace{\left(\int 1 \, dP_\psi \right)^{1/2}}_{=\|\psi\|} < +\infty.$$

Note that if f is simple then it is bounded, and hence in $L^2(P_\psi)$. Moreover, if f is a characteristic function then (5.122) is just the definition of P_ψ, from which (5.122) follows immediately for simple functions. For a general $f \in L^2(P_\psi)$, we may choose a sequence $(f_n)_{n\in\mathbb{N}}$ of simple measurable functions with $|f_n| \leq |f_{n+1}|$, $n \in \mathbb{N}$, and obtain (5.122) by the monotone convergence theorem. □

The above can also be generalized to expectations w.r.t. density operators, as follows:

Exercise 5.44. Let $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ be measurable, and ϱ be a density operator with a decomposition $\varrho = \sum_{j \in J} p_j \, |\psi_j\rangle\langle\psi_j|$, where J is finite or $J = \mathbb{N}$, $(p_j)_{j \in J}$ is a probability distribution, and for each $j \in J$, $\|\psi_j\| = 1$ and $\psi_j \in \mathcal{D}(P(f))$. Show that

$$\operatorname{Tr} \varrho P(f) = \int f \, dP_\varrho. \tag{5.123}$$

Definition 5.45. Let $P \in \mathrm{PVM}(\mathcal{H}, \mathbb{K})$ be a real- or complex-valued PVM. We define the *operator corresponding to* P as

$$\hat{P} := P(\mathrm{id}_\mathbb{K}) = \int \mathrm{id}_\mathbb{K} \, dP.$$

By Proposition 5.43, we have

$$\mathbb{E}_{|\psi\rangle\langle\psi|}(P) := \int_\mathbb{K} \mathrm{id}_\mathbb{K} \, dP_\psi = \left\langle \psi, \hat{P}\psi \right\rangle, \qquad \psi \in \mathcal{D}(\hat{P}), \tag{5.124}$$

as required in (5.112).

Remark 5.46. Note that (5.124) holds for every $\psi \in \mathcal{D}(\hat{P})$, and this is the maximal set of vectors for which the expression $\left\langle \psi, \hat{P}\psi \right\rangle$ makes sense. However, it may happen that $\psi \notin \mathcal{D}(P(f))$, i.e., $f \notin L^2(\mathcal{X}, \mathcal{A}, P_\psi)$, but $f \in L^1(\mathcal{X}, \mathcal{A}, P_\psi)$, and hence $\int_\mathbb{K} \mathrm{id}_\mathbb{K} \, dP_\psi$ is well-defined and finite. In this case, the expectation value $\mathbb{E}_{|\psi\rangle\langle\psi|}(P)$ exists and is finite, but it cannot be expressed in the form $\left\langle \psi, \hat{P}\psi \right\rangle$.

The next question is whether (5.124) can be extended for higher moments, i.e., if

$$\int_\mathbb{K} \mathrm{id}_\mathbb{K}^m \, dP_\psi = \left\langle \psi, \hat{P}^m \psi \right\rangle, \qquad m \in \mathbb{N}. \tag{5.125}$$

To this end, we introduce the following:

Definition 5.47. Let $P \in \mathrm{PVM}(\mathcal{H}, \mathbb{K})$ be a real- or complex-valued PVM. For any measurable function $f : \mathbb{K} \to \mathbb{C}$, we define the function f of $\hat{P}$ as

$$f(\hat{P}) := P(f) = \int f \, dP.$$

With this definition, we have

$$\Big\langle \psi, f(\hat{P})\psi \Big\rangle = \langle \psi, P(f)\psi\rangle = \int f\, dP_\psi,$$

according to Proposition 5.43. In particular,

$$\Big\langle \psi, (\mathrm{id}_{\mathbb{K}}^m)(\hat{P})\psi \Big\rangle = \langle \psi, P(\mathrm{id}_{\mathbb{K}}^m)\psi\rangle = \int \mathrm{id}_{\mathbb{K}}^m\, dP_\psi, \qquad \psi \in \mathcal{D}(P(\mathrm{id}_{\mathbb{K}}^m)). \quad (5.126)$$

To see that this indeed yields (5.125), we need to verify that $(\mathrm{id}_{\mathbb{K}}^m)(\hat{P}) = \hat{P}^m$, $m \in \mathbb{N}$, i.e., that the functional calculus introduced in Definition 5.47 is compatible with the conventional definition of taking powers of an operator. We will show in Corollary 5.60 that this is true, and hence $\hat{P}$ is indeed the operator that we set out to define at the beginning of the section. Note, however, that, similarly to the case of the expectation value discussed in Remark 5.46, (5.125) will hold only for $\psi \in \mathcal{D}(\hat{P}^m)$, which may be strictly smaller than the set of states ψ for which the m-th moment is well-defined and finite.

Before proceeding further in this direction, let us address another very natural question, namely, whether the functional calculus introduced in Definition 5.47 is compatible with the concept of a function of a POVM introduced in Definition 5.19.

We start with the following:

Lemma 5.48. Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces, $P \in \mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$, $g : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ measurable, and $f : (\mathcal{Y}, \mathcal{B}) \to \mathbb{C}$ measurable. Then

$$P(f \circ g) = \int_{\mathcal{X}} (f \circ g)\, dP = \int_{\mathcal{Y}} f\, d(P \circ g^{-1}) = \int_{\mathcal{Y}} f\, dg_* P = (g_* P)(f) = f(\widehat{g_* P}). \quad (5.127)$$

Proof. All the equalities are by definition, except for the second one. If $f = \mathbf{1}_A$ for some $A \in \mathcal{A}$ then $\mathbf{1}_A \circ g = \mathbf{1}_{g^{-1}(A)}$ yields $\int_{\mathcal{X}} (f \circ g)\, dP = P(g^{-1}(A)) = (P \circ g^{-1})(A) = \int_{\mathcal{Y}} \mathbf{1}_A\, d(P \circ g^{-1})$. From this, the assertion follows immediately when f is a simple measurable functions. Now, let $(f_n)_{n\in\mathbb{N}}$ be a sequence of simple measurable functions such that $|f_n| \leq |f|$, $n \in \mathbb{N}$. This implies that $(f_n \circ g)_{n\in\mathbb{N}}$ is a sequence of simple measurable functions such that $|f_n \circ g| \leq |f \circ g|$, $n \in \mathbb{N}$. By Lemma 5.39,

$$\begin{aligned} \psi \in \mathcal{D}(P(f \circ g)) &\iff \exists \lim_{n\to+\infty} \underbrace{P(f_n \circ g)}_{=(g_*P)(f_n)} \psi \\ &\iff \exists \lim_{n\to+\infty} (g_* P)(f_n)\psi \iff \psi \in \mathcal{D}((g_* P)(f_n)), \end{aligned}$$

and if ψ is as above then

$$P(f \circ g)\psi = \lim_{n\to+\infty} P(f_n \circ g)\psi = \lim_{n\to+\infty} (g_* P)(f_n)\psi = (g_* P)(f_n)\psi.$$

□

Remark 5.49. We may call the second equality in (5.127) a *change of variables* formula for the integral w.r.t. projection-valued measures.

Applying Lemma 5.48 with $g = \mathrm{id}_{\mathbb{K}}$ yields the following:

Corollary 5.50. For any real- or complex-valued PVM $P \in \mathrm{PVM}(\mathcal{H}, \mathbb{K})$, and any measurable function $f : \mathbb{K} \to \mathbb{C}$,

$$f(\hat{P}) = \widehat{f_* P}.$$

That is, the functional calculus introduced in Definition 5.47 is compatible with Definition 5.19 in the sense that a function of the operator corresponding to a PVM is the operator corresponding to the function of the PVM.

Remark 5.51. Recall that a measurable functional calculus can be very easily defined for multiplication operators: If $g : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}$ is measurable, and $f : \mathbb{K} \to \mathbb{C}$ is measurable, then

$$f(M_g) := M_{f \circ g}$$

has all the good algebraic properties that we may expect from a functional calculus.

Similarly, we may define a measurable functional calculus for all operators that can be written in the form $P(g)$ for some PVM $P \in \mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ and measurable function $g : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}$, by

$$f(P(g)) := P(f \circ g) \tag{5.128}$$

for any measurable $f : \mathbb{K} \to \mathbb{C}$. By lemma 5.48,

$$P(g) = \widehat{g_* P}, \qquad \text{and thus} \qquad f(P(g)) = f\left(\widehat{g_* P}\right).$$

That is, the operators that can be written in the form $P(g)$ for a PVM on a *general* measurable space $(\mathcal{X}, \mathcal{A})$ and some complex-valued measurable function g, are exactly the operators that can be written in the form $\hat{Q}$ for a real- or complex-valued PVM Q, and thus the functional calculus defined in (5.128) is the same as the one defined in Definition 5.47.

We will give some important illustrations of the concepts introduced in Definition 5.45 in Examples 5.11 and 5.68.

Let us now return to the investigation of the properties of the functional calculus $f \mapsto P(f)$. We start with noting that in general, $P(f)$ does not need to be everywhere defined or bounded. As we will see below, these hold if and only if f is bounded P-almost everywhere, a concept that we introduce next.

Similarly to ordinary measures, we may define the (P,∞)-norm of a measurable function $f:(\mathcal{X},\mathcal{A})\to\mathbb{C}$ as

$$\begin{aligned}\|f\|_{P,\infty} &:= \inf\{C>0:\ P(\{x\in\mathcal{X}:\ |f(x)|>C\})=0\}\\ &= \sup\{C\geq 0:\ P(\{x\in\mathcal{X}:\ |f(x)|\geq C\})\neq 0\}.\end{aligned}$$

We say that f is *P-bounded*, if $\|f\|_{P,\infty}<+\infty$. We denote the set of P-bounded measurable functions by

$$L^\infty_P := \{f:(\mathcal{X},\mathcal{A})\to\mathbb{C}\text{ measurable}:\ \|f\|_{P,\infty}<+\infty\}.$$

Exercise 5.52. Show that

$$\|f\|_{P,\infty} = \sup_{\psi\in\mathcal{H}}\|f\|_{P_\psi,\infty}\,. \tag{5.129}$$

Solution: Clearly, if $P(\{x\in\mathcal{X}:\ |f(x)|>C\})=0$ for some $C>0$ then

$$P_\psi(\{x\in\mathcal{X}:\ |f(x)|>C\}) = \langle\psi, P(\{x\in\mathcal{X}:\ |f(x)|>C\})\psi\rangle = 0$$

for any $\psi\in\mathcal{H}$, and thus $\sup_{\psi\in\mathcal{H}}\|f\|_{P_\psi}\leq\|f\|_{P,\infty}$. Conversely, if $P(\{x\in\mathcal{X}:\ |f(x)|\geq C\})\neq 0$ for some $C>0$ then there exists some $\psi\neq 0$ such that

$$0<\langle\psi, P(\{x\in\mathcal{X}:\ |f(x)|\geq C\})\psi\rangle = P_\psi(\{x\in\mathcal{X}:\ |f(x)|\geq C\}),$$

and taking the supremum over all such C yields $\|f\|_{P,\infty}\leq\|f\|_{P_\psi}$.

Exercise 5.53. Show that L^∞_P is a vector space.

Proposition 5.54.

$$\|P(f)\| = \|f\|_{P,\infty}\,.$$

In particular, $P(f)$ is a bounded operator if and only if the function f is P-bounded, in which case $P(f)$ is also everywhere defined.

Proof. Let $\psi\in\mathcal{H}$ be an arbitrary vector. Then

$$\int|f|^2\,dP_\psi \leq \int\|f\|^2_{P_\psi,\infty}\,dP_\psi = \|f\|^2_{P_\psi,\infty}\,P_\psi(\mathcal{X}) \leq \|f\|^2_{P,\infty}\,\|\psi\|^2 < +\infty,$$

where the second inequality is due to (5.129). Thus, $\psi\in\mathcal{D}(\mathcal{P}(f))$ for every $\psi\in\mathcal{H}$, and $\|P(f)\psi\|^2 = \int|f|^2\,dP_\psi \leq \|f\|_{P,\infty}\,\|\psi\|^2$ according to (5.119), whence $\|P(f)\|\leq$

$\|f\|_{P,\infty}$. Conversely, if $C \geq 0$ is such that $P_C := (\{x \in \mathcal{X} : |f(x)| \geq C\}) \neq 0$ then there exists a $0 \neq \psi \in \operatorname{ran} P_C$, and

$$\begin{aligned}\|P(f)\psi\|^2 = \|P(f)P_C\psi\|^2 &= \int_{\mathcal{X}} |f|^2 \, dP_{P_C\psi} = \int_{\{|f|\geq C\}} |f|^2 \, dP_\psi \\ &\geq C^2 P_\psi(\{|f| \geq C\}) = C^2 \langle \psi, P_C\psi\rangle = C^2 \|\psi\|^2,\end{aligned}$$

where the third equality is due to Lemma 5.30. Thus, $\|P(f)\| \geq C$. Taking the supremum over all such C yields $\|P(f)\| \geq \|f\|_{P,\infty}$. □

Now we are ready to state and prove the fundamental algebraic properties of the functional calculus for PVMs:

Proposition 5.55. Let $f, g : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ be measurable.

(i) $P(\lambda f) = \lambda P(f)$ for any $\lambda \in \mathbb{C} \setminus \{0\}$, and $P(0 \cdot f) = 0 \in \mathcal{B}(\mathcal{H})$.

(ii) $P(f) + P(g) \subseteq P(f + g)$,
and equality holds if at least one of the functions is P-bounded.

(iii) $P(f)P(g) \subseteq P(fg)$, and $\mathcal{D}(P(f)P(g)) = \mathcal{D}(P(g)) \cap \mathcal{D}(P(fg))$.
In particular, $P(f)P(g) = P(fg) \iff \mathcal{D}(P(g)) \cap \mathcal{D}(P(fg)) = \mathcal{D}(P(fg))$, which holds when f is P-bounded.

(iv) $P(f)^* = P(\overline{f})$.

(v) $P(f)^*P(f) = P(|f|^2) = P(f)P(f)^*$.

Proof. (i) Trivial.

(ii) We have $|f + g|^2 \leq (|f| + |g|)^2 = |f|^2 + |g|^2 + 2|f||g| \leq 2(|f|^2 + |g|^2)$, where the last inequality follows from the inequality between the geometric and the square mean. Thus, if $\psi \in \mathcal{D}(P(f) \cap P(g)) = \mathcal{D}(P(f) + P(g))$ then

$$\int |f + g|^2 \, dP_\psi \leq 2 \int |f|^2 \, dP_\psi + 2 \int |g|^2 \, dP_\psi < +\infty,$$

and hence $\psi \in \mathcal{D}(P(f + g))$. If $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are approximating sequences of simple normal functions as in Lemma 5.39 then $P(f_n + g_n) = P(f_n) + P(g_n)$ by Exercise 5.36, and $P(f_n + g_n)\psi = P(f_n)\psi + P(g_n)\psi$ follows by taking the limit in n.

Assume now that f is P-bounded, so that $\mathcal{D}(P(f)) = \mathcal{H}$, and thus $\mathcal{D}(P(f) + P(g)) = \mathcal{D}(P(g))$. Let $\psi \in \mathcal{D}(P(f+g))$. Then $|g|^2 = |g+f-f|^2 \leq (|g+f| + |f|)^2 \leq 2(|g+f|^2 + |f|^2)$ yields

$$\int |g|^2 \, dP_\psi \leq \underbrace{\int |f+g|^2 \, dP_\psi}_{\leq +\infty} + \underbrace{\int |f|^2 \, dP}_{\leq \|f\|^2_{P,\infty}\|\psi\|^2} < +\infty,$$

and hence $\psi \in \mathcal{D}(P(g))$. This implies that $D(P(f) + P(g)) = D(P(g)) \supseteq \mathcal{D}(P(f+g))$, and hence, by the above $P(f) + P(g) = P(f+g)$.

(iii) Let $(f_n)_{n\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}}$ be sequences of bounded measurable functions converging pointwise to f, g, respectively, such that $|f_n| \leq |f_{n+1}|$, $|g_n| \leq |g_{n+1}|$, $n \in \mathbb{N}$. Note that $P(f_n g_m) = P(f_n)P(g_m)$ by Exercise 5.36. Let $\psi \in \mathcal{D}(P(g))$. Then $\lim_{m\to+\infty} P(g_m)\psi = P(g)\psi$ by Lemma 5.39, and

$$P(f_n)P(g)\psi = \lim_{m\to+\infty} P(f_n)P(g_m)\psi = \lim_{m\to+\infty} P(f_n g_m)\psi = P(f_n g)\psi, \tag{5.130}$$

where the first equality follows from the fact that $P(f_n)$ is bounded, and the last equality follows from Lemma 5.39 by the fact that $\int |f_n g|^2 \, dP_\psi \leq \|f_n\|^2_\infty \int |g|^2 \, dP_\psi < +\infty$, and hence $\psi \in \mathcal{D}(P(f_n g))$. (One may also argue by noting that $f_n g_m$ converges pointwise to $f_n g$ as $m \to +\infty$, and $|f_n g_m| \leq |f_n g|$.) Therefore,

$$\int |f_n g|^2 \, dP_\psi = \|P(f_n g)\psi\|^2 = \|P(f_n)P(g)\psi\|^2 = \int |f_n|^2 \, dP_{P(g)\psi}.$$

This yields, by the monotone convergence theorem, that

$$\int |fg|^2 \, dP_\psi = \lim_{n\to+\infty} \int |f_n g|^2 \, dP_\psi = \lim_{n\to+\infty} \int |f_n|^2 \, dP_{P(g)\psi} = \int |f|^2 \, dP_{P(g)\psi}.$$

Thus, by Lemma 5.39, $\psi \in \mathcal{D}(P(fg)) \iff P(g)\psi \in \mathcal{D}(P(f))$. Finally, if either (and hence both) of these holds, then

$$P(fg)\psi = \lim_{n\to+\infty} P(f_n g)\psi = \lim_{n\to+\infty} P(f_n)P(g)\psi = P(f)P(g)\psi, \tag{5.131}$$

where the first equality is due to the fact that $\psi \in \mathcal{D}(P(fg))$, he second equality follows from (5.130), and the third equality follows from $P(g)\psi \in \mathcal{D}(P(f))$.

(iv) Let $\psi_1, \psi_2 \in \mathcal{D}(P(f)) = D(P(\overline{f}))$, and let $(f_n)_{n\in\mathbb{N}}$ is a sequence of bounded measurable functions converging to f, respectively, such that $|f_n| \leq |f_{n+1}|$, $n \in \mathbb{N}$. Then

$$\langle \psi_1, P(f)\psi_2\rangle = \lim_{n\to+\infty} \langle \psi_1, P(f_n)\psi_2\rangle = \lim_{n\to+\infty} \left\langle P(\overline{f}_n)\psi_1, \psi_2\right\rangle = \left\langle P(\overline{f})\psi_1, \psi_2\right\rangle,$$

where the second equality is due to Exercise 5.36. Thus, $P(\overline{f}) \subseteq P(f)^*$. If f is bounded then $P(\overline{f})$ is everywhere defined, according to Proposition 5.54, and in this case $P(\overline{f}) = P(f)^*$.

Now, let $\psi \in \mathcal{D}(P(f)^*)$, and $g_n := \mathbf{1}_{\{|f|\leq n\}}$, $n \in \mathbb{N}$. Then

$$\mathcal{D}(P(f)P(g_n)) = \underbrace{\mathcal{D}(P(g_n))}_{=\mathcal{H}} \cap \underbrace{\mathcal{D}(P(fg_n))}_{=\mathcal{H}} = \mathcal{H} \Longrightarrow P(f)P(g_n) = P(fg_n),$$

where we used that g_n and fg_n are bounded, and the implication follows from the previous point. Thus,

$$P(g_n)P(f)^* = P(g_n)^*P(f)^* \subseteq [P(f)P(g_n)]^* = P(fg_n)^* = P(\overline{fg_n}) = P(\overline{f}g_n),$$

where in the first equality we used that g_n is bounded, and in the third equality we used that fg_n is bounded. Hence,

$$\int |g_n|^2\, dP_{P(f)^*\psi} = \|P(g_n)^*P(f)^*\psi\|^2 = \left\|P(\overline{f}g_n)\psi\right\|^2 = \int |fg_n|^2\, dP_\psi.$$

Thus,

$$\begin{aligned} +\infty > \|P(f)^*\psi\|^2 &= \int \mathbf{1}_{\mathcal{X}}\, dP_{P(f)^*\psi} = \lim_{n\to+\infty} \int |g_n|^2\, dP_{P(f)^*\psi} \\ &= \lim_{n\to+\infty} \int |fg_n|^2\, dP_\psi = \int |f|^2\, dP_\psi, \end{aligned}$$

where the second and the last equality follow by the monotone convergence theorem. Thus, $\psi \in \mathcal{D}(P(f)) = \mathcal{D}(P(\overline{f}))$, and therefore $P(f)^* = P(\overline{f})$.

(v) Note that $\psi \in D(P(|f|^2) \Longleftrightarrow \int |f|^4\, dP_\psi < +\infty$, and in this case,

$$\int |f|^2\, dP_\psi = \int |f|^2 \cdot 1\, dP_\psi \leq \left(\int |f|^4\, dP_\psi\right)^{1/2} \left(\int 1\, dP_\psi\right)^{1/2} < +\infty,$$

due to the Cauchy-Schwarz inequality, so $\psi \in \mathcal{D}(P(f))$. Thus, $\mathcal{D}(P(f)P(\overline{f})) = \mathcal{D}(P(|f|^2) \cap \mathcal{D}(P(f)) = \mathcal{D}(P(|f|^2)$, and similarly, $\mathcal{D}(P(\overline{f})P(f)) = \mathcal{D}(P(|f|^2)$, which implies

$$P(f)^*P(f) = P(\overline{f})P(f) = P(\overline{f}f) = P(|f|^2) = P(f\overline{f}) = P(f)P(f)^*.$$

□

Proposition 5.54 and the linearity established in Proposition 5.55 yield immediately the following continuity property:

Corollary 5.56. Let f be P-bounded, and $(f_n)_{n\in\mathbb{N}}$ be a sequence of P-bounded measurable functions such that $\lim_{n\to+\infty} \|f - f_n\|_{P,\infty} = 0$. Then

$$\lim_{n\to+\infty} P(f_n) = P(f)$$

in operator norm.

Proof. By Proposition 5.54, $P(f_n) \in \mathcal{B}(\mathcal{H})$, $n \in \mathbb{N}$, and

$$\|P(f_n) - P(f)\| = \|P(f_n - f)\| = \|f_n - f\|_{P,\infty} \xrightarrow[n\to+\infty]{} 0,$$

where the first equality follows from Proposition 5.55, and the second equality from Proposition 5.54. □

Corollary 5.57. All the identities in Exercise 5.36 hold when f, g are P-bounded. That is, the functional calculus $f \mapsto P(f)$ is a *-algebra morphism from L^∞_P to $\mathcal{B}(\mathcal{H})$.

The algebraic properties established in Proposition 5.55 readily translate to properties of the functional calculus for $\hat{P}$, introduced in Definition 5.47:

Proposition 5.58. Let $P \in \mathrm{PVM}(\mathcal{H}, \mathbb{K})$ be a real- or complex-valued PVM and $\hat{P}$ the corresponding operator. For any measurable functions $f, g : \mathbb{K} \to \mathbb{C}$, and scalar $\lambda \in \mathbb{C} \setminus \{0\}$, we have

$$(\lambda f)(\hat{P}) = \lambda f(\hat{P}), \tag{5.132}$$

$$f(\hat{P}) + g(\hat{P}) \subseteq (f+g)(\hat{P}), \tag{5.133}$$

$$f(\hat{P})g(\hat{P}) \subseteq (fg)(\hat{P}), \tag{5.134}$$

$$\overline{f}(\hat{P}) = (f(\hat{P}))^*, \tag{5.135}$$

$$f(\hat{P})(f(\hat{P}))^* = |f|^2(\hat{P}). \tag{5.136}$$

For $\lambda = 0$, (5.132) holds as $(0 \cdot f)(\hat{P}) = 0 \in \mathcal{B}(\mathcal{H})$.

Moreover, equality holds in (5.133) if at least one of the functions is bounded. In (5.134), we have

$$\mathcal{D}\left(f(\hat{P})g(\hat{P})\right) = \mathcal{D}(f(\hat{g})) \cap \mathcal{D}((f+g)(\hat{P})),$$

and equality holds in (5.134) if and only if $\mathcal{D}(g(\hat{P})) \cap \mathcal{D}((fg)(\hat{P})) = \mathcal{D}((fg)(\hat{P}))$.

These allow us to settle the question raised around formulas (5.125)–(5.126). We start with the following:

Exercise 5.59. Show that for $n, m \in \mathbb{N}$, $n \leq m$,

$$\mathcal{D}\left((\mathrm{id}_{\mathbb{K}}^m)(\hat{P})\right) \subseteq \mathcal{D}\left((\mathrm{id}_{\mathbb{K}}^n)(\hat{P})\right), \tag{5.137}$$

where by definition, $\mathrm{id}_{\mathbb{K}}^0 := \mathbf{1}_{\mathbb{K}}$. Show that for any $n, m \in \mathbb{N}$,

$$(\mathrm{id}_{\mathbb{K}}^m)(\hat{P}) \cdot (\mathrm{id}_{\mathbb{K}}^n)(\hat{P}) = (\mathrm{id}_{\mathbb{K}}^{n+m})(\hat{P}).$$

(Hint: Use the Hölder inequality.)

Solution:

The assertions are trivial when $n = 0$, since then $\mathrm{id}_{\mathbb{K}}^n(\hat{P}) = I$, and hence we assume that $n > 0$. Let $p := m/n$, and $q := 1 - 1/p = 1 - n/m = (m-n)/n$ be its Hölder conjugate. For any $\psi \in \mathcal{H}$,

$$\begin{aligned}
\int_{\mathbb{K}} |\,\mathrm{id}_{\mathbb{K}}^n\,|^2\, dP_\psi &= \int_{\mathbb{K}} |\,\mathrm{id}_{\mathbb{K}}^2\,|^n \cdot 1\, dP_\psi \leq \left\| |\,\mathrm{id}_{\mathbb{K}}\,|^{2n} \right\|_p \|1\|_q \\
&= \left(\int_{\mathbb{K}} \left(|\,\mathrm{id}_{\mathbb{K}}\,|^{2n} \right)^{m/n}\, dP_\psi \right)^{n/m} \left(\int_{\mathbb{K}} 1^{m/(n-m)}\, dP_\psi \right)^{(m-n)/m} \\
&= \left(\int_{\mathbb{K}} |\,\mathrm{id}_{\mathbb{K}}^2\,|^m\, dP_\psi \right)^{n/m} \|\psi\|^{2(m-n)/m},
\end{aligned}$$

where the inequality follows from Hölder's inequality. Thus, if $\psi \in \mathcal{D}((\mathrm{id}_{\mathbb{K}}^m)(\hat{P}))$ then $\psi \in \mathcal{D}((\mathrm{id}_{\mathbb{K}}^n)(\hat{P}))$, proving (5.137).

By Proposition 5.58,

$$\mathcal{D}\left((\mathrm{id}_{\mathbb{K}}^m)(\hat{P}) \cdot (\mathrm{id}_{\mathbb{K}}^n)(\hat{P})\right) = \mathcal{D}((\mathrm{id}_{\mathbb{K}}^m)(\hat{P})) \cap \mathcal{D}((\mathrm{id}_{\mathbb{K}})^{n+m}(\hat{P})) = \mathcal{D}((\mathrm{id}_{\mathbb{K}}^{n+m})(\hat{P})),$$

where the last equality is due to (5.137).

Corollary 5.60. For any $m \in \mathbb{N}$,

$$\mathrm{id}_{\mathbb{K}}^m(\hat{P}) = \hat{P}^m,$$

and for any $\psi \in \hat{P}^m$, (5.125) holds, i.e.,

$$\int_{\mathbb{K}} \mathrm{id}_{\mathbb{K}}^m\, dP_\psi = \left\langle \psi, \hat{P}^m \psi \right\rangle.$$

Proof. By definition, $\hat{P} = \mathrm{id}_{\mathbb{K}}(\hat{P})$, and thus we have, for any $m \in \mathbb{N}$,

$$\hat{P}^m = \hat{P} \cdot \ldots \cdot \hat{P} = \mathrm{id}_{\mathbb{K}}(\hat{P}) \cdot \ldots \cdot \mathrm{id}_{\mathbb{K}}(\hat{P}) = \mathrm{id}^m_{\mathbb{K}}(\hat{P}),$$

where the last equality follows from Exercise 5.59. □

The above can be further generalized as follows:

Exercise 5.61. Let $P \in \mathrm{PVM}(\mathcal{H}, \mathbb{K})$ be a real- or complex-valued PVM and $\hat{P}$ the corresponding operator. Show that for any polynomial $p(z) = \sum_{k=0}^{n} c_k z^k$, we have

$$p(\hat{P}) = P(p) = \sum_{k=0}^{n} c_k \hat{P}^k.$$

(Hint: Show that $\mathcal{D}(p(\hat{P})) = \mathcal{D}(\mathrm{id}^n_{\mathbb{K}}(\hat{P})) = \mathcal{D}(p(\hat{P}))$ where $n = \deg p$. You may want to compute the integral $\int |p|^2 \, dP_\psi$ by splitting $\mathbb{K}$ to a large enough disk around the origin and its complement.)

Remark 5.62. Exercise 5.61 shows that the functional calculus developed for $\hat{P}$ is an extension of the polynomial function calculus.

Next, we turn to the following very important corollary of Proposition 5.55:

Corollary 5.63. For any measurable $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$, $P(f)$ is a *normal operator*, i.e., it is densely defined, closed, and $P(f)^* P(f) = P(f) P(f)^*$.

Proof. We have seen in Proposition 5.41 that $P(f)$ is densely defined. It is closed, as $P(f) = P(\overline{f})^*$ by Proposition 5.55, and any adjoint operator is closed, and $P(f)^* P(f) = P(f) P(f)^*$ also follows from Proposition 5.55. □

As it turns out, the above Corollary can also be reversed, and we have the following:

Theorem 5.64. (Spectral theorem for normal operators - PVM form)
Let T be a normal operator on a Hilbert space $\mathcal{H}$. Then there exists a PVM $P^T \in \mathrm{PVM}(\mathcal{H}, \mathbb{C})$, called the *spectral PVM of* T, such that

$$T = \widehat{P^T} = \int \mathrm{id}_{\mathbb{C}} \, dP^T.$$

Moreover, $\operatorname{supp} P^T = \operatorname{spec}(T)$.

Remark 5.65. The above spectral theorem is one of the central results in the theory of Hilbert space operators. Its proof is beyond the scope of these notes, and hence we omit it.

Similarly to usual measure theory, we say that a property of a function holds P-almost everywhere (a.e.) if the set of points A where it does not hold is measurable, and $P(A) = 0$. For a measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{K}$, we define its *$P$-essential range* as

$$\operatorname{essran}_P f := \operatorname{supp}(P \circ f)^{-1} = \{x \in \mathbb{K} : P(f^{-1}(B_\varepsilon(x)) \neq 0 \ \forall \varepsilon > 0\},$$

where $B_\varepsilon(x)$ is the ε-ball centered at x.

We leave the following properties of the functional calculus as an exercise:

Exercise 5.66. Let $P \in \mathrm{PVM}(\mathcal{H}, \mathcal{X}, \mathcal{A})$ be a PVM and $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ be a measurable function. Show the following:

(i) The spectrum of $P(f)$ is given by

$$\operatorname{spec}(P(f)) = \operatorname{supp}(P \circ f^{-1}) = \operatorname{essran}_P f.$$

In particular, if P is a real- or complex PVM then

$$\operatorname{spec}(\hat{P}) = \operatorname{supp} P.$$

(ii) Show that $P(f)$ is

a) self-adjoint $\iff \operatorname{essran}_P f \subseteq \mathbb{R}$;

b) unitary, $\iff \operatorname{essran}_P f \subseteq \{z \in \mathbb{C} : |z| = 1\}$;

c) projection, $\iff \operatorname{essran}_P f \subseteq \{0, 1\}$;

d) positive semi-definite $\iff \operatorname{essran}_P f \subseteq [0, +\infty)$.

Let us now consider some important examples and applications of the functional calculus developed above.

Example 5.67. Consider the multiplication PVM $M \in \mathrm{PVM}(L^2(\mathcal{X}, \mathcal{A}, \mu), \mathcal{X}, \mathcal{A})$ from Example 5.11, given by $M(A) := M_{\mathbf{1}_A}$, where the latter is the multiplication operator with the characteristic function of $A \in \mathcal{A}$. As we have already seen before, for any $\psi \in L^2(\mathcal{X}, \mathcal{A}, \mu)$,

$$M_\psi(A) = \langle \psi, M(A)\psi \rangle = \int_{\mathcal{X}} \overline{\psi} \mathbf{1}_A \psi \, d\mu = \int_A |\psi|^2 \, d\mu = (|\psi|^2 \mu)(A).$$

Thus, for any $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$ measurable,

$$\psi \in \mathcal{D}(M(f)) \iff +\infty > \int |f|^2 \, dM_\psi = \int |f|^2 |\psi|^2 \, d\mu \iff \psi \in \mathcal{D}(M_f).$$

Moreover, for any simple measurable function $f = \sum_{k=1}^{r} c_k \mathbf{1}_{A_k}$,

$$M(f) = \sum_{k=1}^{r} c_k M(A_k) = \sum_{k=1}^{r} c_k M_{\mathbf{1}_{A_k}} = M_{\sum_{k=1}^{r} c_k \mathbf{1}_{A_k}} = M_f,$$

and therefore for any measurable function $f : (\mathcal{X}, \mathcal{A}) \to \mathbb{C}$,

$$M_f = M(f) = \int_{\mathcal{X}} f \, dM = \int_{\mathbb{C}} \mathrm{id}_{\mathbb{C}} \, d(M \circ f^{-1}) = \widehat{f_* M}.$$

Definition **??** then allows us to define functions of the multiplication operator M_f, and we obtain

$$g(M_f) := g(M(f)) = \int_{\mathcal{X}} (g \circ f) \, dM = M_{g \circ f}, \tag{5.138}$$

for any measurable function $g : \mathbb{C} \to \mathbb{C}$. That is, a function g of the multiplication operator by f is the multiplication operator by $g \circ f$.

It is straightforward to verify that multiplication operators are normal, and from the point of view of Theorem 5.64, the above considerations show that the spectral PVM of the multiplication operator M_f is $M \circ f^{-1}$, i.e.,

$$P^{M_f} = M \circ f^{-1} = f_* M.$$

Example 5.68. Let us now specify Example 5.67 to the case where $(\mathcal{X}, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, with λ being the Lebesgue-measure, and let Q denote the multiplication PVM. We may consider it as a real-valued POVM, or as a complex-valued PVM with support in $\mathbb{R}$, and we can define the operator $\hat{Q}$ corresponding to Q according to Definition 5.45, as

$$\hat{Q} := \int_{\mathbb{K}} \mathrm{id}_{\mathbb{K}} \, dQ = Q(\mathrm{id}_{\mathbb{R}}) = M_{\mathrm{id}_{\mathbb{R}}},$$

the multiplication with the coordinate in $\mathbb{R}$, where the equalities follow as special cases of Example 5.67. This is the operator corresponding to the position observable in quantum mechanics. Functions of the position observable are then defined as

$$g(\hat{Q}) := M_{g \circ \mathrm{id}_{\mathbb{R}}} = M_g,$$

according to (5.138). The spectral PVM of the position operator $\hat{Q}$ is then Q itself, i..e.,

$$P^{\hat{Q}} = Q.$$

Exercise 5.69.

6 Composite systems

6.1 The tensor product of Hilbert spaces

Definition 6.1. Let $\mathcal{H}_1, \ldots, \mathcal{H}_n$, be Hilbert spaces over the same scalar field $\mathbb{K}$. A pair $(\mathcal{K}, \tau)$, where $\mathcal{K}$ is a Hilbert space over $\mathbb{K}$, and $\tau : \mathcal{H}_1 \times \ldots \times \mathcal{H}_n \to \mathcal{K}$ is an n-linear map, is called a *realization of the tensor product* of $\mathcal{H}_1, \ldots, \mathcal{H}_n$, if

(i) the inner product of $\mathcal{K}$ factorizes on $\operatorname{ran}\tau$ as

$$\langle \tau(\phi_1, \ldots, \phi_n), \tau(\psi_1, \ldots, \psi_n) \rangle = \langle \phi_1, \psi_1 \rangle \cdot \ldots \cdot \langle \phi_n, \psi_n \rangle , \tag{6.139}$$

where $\phi_i, \psi_i \in \mathcal{H}_i$, $i \in [n]$;

(ii) the subspace generated by $\operatorname{ran}\tau$ is dense in $\mathcal{K}$, i.e.,

$$\overline{\operatorname{span}}\{\tau(\psi_1, \ldots, \psi_n) : \psi_i \in \mathcal{H}_i,\, i \in [n]\} = \mathcal{K}. \tag{6.140}$$

Lemma 6.2. Assume that an n-linear map $\tau : \mathcal{H}_1 \times \ldots \times \mathcal{H}_n \to \mathcal{K}$ satisfies (6.139).

(i) τ is bounded with norm 1.

(ii) For any orthonormal systems $(e_{i,j})_{j \in J_i}$, $i \in [n]$,

$$\{e_{\underline{j}} := \tau(e_{1,j_i}, \ldots, e_{n,j_n}) : \underline{j} \in \times_{i=1}^n J_i\}$$

is an ONS in $\mathcal{K}$.

(iii) (6.140) holds if and only if for any/some ONBs $(e_{i,j})_{j \in J_i}$, $i \in [n]$,

$$\{e_{\underline{j}} = \tau(e_{1,j_i}, \ldots, e_{n,j_n}) : \underline{j} \in \times_{i=1}^n J_i\} \quad \text{is an ONB in } \mathcal{K}. \tag{6.141}$$

Proof. (i) According to (6.139),

$$\|\tau(\psi_1, \ldots, \psi_n)\| = \|\psi_1\| \cdot \ldots \cdot \|\psi_n\|$$

for any $\psi_i \in \mathcal{H}_i$, $i \in [n]$, proving the assertion.

(ii) Immediate from (6.139).

(iii) Assume that (6.141) holds for some ONBs. Then

$$\mathcal{K} = \overline{\operatorname{span}}\{e_{\underline{j}} : \underline{j} \in \times_{i=1}^n J_i\} \subseteq \overline{\operatorname{span}}\{\tau(\psi_1, \ldots, \psi_n) : \psi_i \in \mathcal{H}_i,\, i \in [n]\} \subseteq \mathcal{K},$$

and hence all containments are equalities, and (6.140) holds. Assume next that (6.140) holds, and let $(e_{i,j})_{j \in J_i}$, $i \in [n]$, be arbitrary ONBs in the respective spaces.

For any $\psi_i \in \mathcal{H}_i$, there exists a sequence $\psi_{i,m} \in \operatorname{span}\{e_{i,j}\}_{j\in J_i}$ such that $\|\psi_i - \psi_{i,m}\| \to 0$ as $m \to +\infty$. Then

$$\underbrace{\sum_{\underline{j}\in\times_{i=1}^n J_i} \prod_{i=1}^n \langle e_{i,j_i}, \psi_i\rangle\, \tau(e_{1,j_1},\dots,e_{n,j_n})}_{\in \operatorname{span}\{e_{\underline{j}}:\, \underline{j}\in\times_{i=1}^n J_i\}} = \tau(\psi_{1,m},\dots,\psi_{n,m}) \xrightarrow[m\to+\infty]{} \tau(\psi_1,\dots,\psi_n),$$

where the first equality is by the n-linearity of τ, the sum is finite by assumption, showing that the first expression is in $\operatorname{span}\{e_{\underline{j}} : \underline{j} \in \times_{i=1}^n J_i\}$, and the convergence is due to the continuity of τ, established in the first point (i). Thus,

$$\mathcal{K} = \overline{\operatorname{span}}\{\tau(\psi_1,\dots,\psi_n) :\, \psi_i \in \mathcal{H}_i,\, i \in [n]\} \subseteq \overline{\operatorname{span}}\{e_{\underline{j}} :\, \underline{j} \in \times_{i=1}^n J_i\} \subseteq \mathcal{K},$$

proving that $\operatorname{span}\{e_{\underline{j}} : \underline{j} \in \times_{i=1}^n J_i\}$ is dense in $\mathcal{K}$. By (ii), this means that $\{e_{\underline{j}} : \underline{j} \in \times_{i=1}^n J_i\}$ is an ONB in $\mathcal{K}$. □

Proposition 6.3. Let $\mathcal{H}_1,\dots,\mathcal{H}_n$, be Hilbert spaces.

(i) There exists a realization of their tensor product.

(ii) If $(\mathcal{K}_1,\tau_1)$ and $(\mathcal{K}_2,\tau_2)$ are two realizations of the tensor product of $\mathcal{H}_1,\dots,\mathcal{H}_n$, then there exists a unique unitary $U : \mathcal{K}_1 \to \mathcal{K}_2$ such that

$$\tau_2 = U \circ \tau_1 \tag{6.142}$$

(and hence $U^{-1} \circ \tau_2 = \tau_1$).

Proof. (i) For every $i \in [n]$, let $(e_{i,j})_{j\in J_i}$ be an ONB in $\mathcal{H}_i$. Then $\mathcal{H}_i$ is isomorphic to $l^2(J_i)$ via $U_i\psi := (\langle e_{i,j}, \psi\rangle)_{j\in J_i}$, $\psi \in \mathcal{H}_i$; see Corollary 4.111. For $\psi_i \in \mathcal{H}_i$, $i \in [n]$, let

$$\tau(\psi_1,\dots,\psi_n) := U_1\psi_1 \dot\otimes \dots \dot\otimes U_n\psi_n \in l^2\left(\times_{i=1}^n J_i\right) =: \mathcal{K},$$

where $\dot\otimes$ is the tensor product of functions introduced in Definition 2.96, i.e.,

$$\begin{aligned}(U_1\psi_1 \dot\otimes \dots \dot\otimes U_n\psi_n)(j_1,\dots,j_n) &= (U_1\psi_1)(j_1)\cdot\ldots\cdot(U_n\psi_1)(j_n)\\ &= \langle e_{1,j_1},\psi_1\rangle \cdot\ldots\cdot \langle e_{n,j_n},\psi_n\rangle\,, \qquad \underline{j} \in \times_{i=1}^n J_i.\end{aligned}$$

Clearly, τ is n-linear, and (6.139) is straightforward to verify. Since

$$\tau(e_{1,j_1},\dots,e_{n,j_n}) = \underbrace{U_1e_{1,j_1}}_{=\mathbf{1}_{\{j_1\}}} \dot\otimes \dots \dot\otimes \underbrace{U_ne_{1,j_n}}_{=\mathbf{1}_{\{j_n\}}} = \mathbf{1}_{(j_1,\dots,j_n)}, \qquad \underline{j} \in \times_{i=1}^n J_i,$$

is an ONB in $l^2(\times_{i=1}^n J_i)$, (6.140) holds due to Lemma 6.2.

(ii) For every $i \in [n]$, let $(e_{i,j})_{j\in J_i}$ be an ONB in $\mathcal{H}_i$. By (iii) of Lemma 6.2 and Lemma 4.110, there exists a unique unitary $U : \mathcal{K}_1 \to \mathcal{K}_2$ such that

$$U\tau_1(e_{1,j_1}, \ldots, e_{n,j_n}) = \tau_2(e_{1,j_1}, \ldots, e_{n,j_n}), \qquad \underline{j} \in \times_{i=1}^n J_i.$$

If $\psi_i \in \operatorname{span}\{e_{i,j}\}_{j\in J_i}$, $i \in [n]$, then

$$\begin{aligned} U\tau_1(\psi_1, \ldots, \psi_n) &= U \sum_{\underline{j}\in\times_{i=1}^n J_i} \prod_{i=1}^n \langle e_{i,j_i}, \psi_i\rangle \, \tau_1(e_{1,j_1}, \ldots, e_{n,j_n}) \\ &= \sum_{\underline{j}\in\times_{i=1}^n J_i} \prod_{i=1}^n \langle e_{i,j_i}, \psi_i\rangle \, \tau_2(e_{1,j_1}, \ldots, e_{n,j_n}) \\ &= \tau_2(\psi_1, \ldots, \psi_n). \end{aligned}$$

For general $\psi_i \in \mathcal{H}_i$, $i \in [n]$, $U\tau_1(\psi_1, \ldots, \psi_n) = \tau_2(\psi_1, \ldots, \psi_n)$ follows from the above by continuity; see (i) of Lemma 6.2. □

According to Proposition 6.3, any two realizations of the tensor product are equivalent in a canonical way. Hence, unless we want to explicitly specify what realization we are working with, we will simply write "the" tensor product of the Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_n$ for any realization, and denote it by

$$\mathcal{H}_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \mathcal{H}_n = \overline{\otimes}_{i=1}^n \mathcal{H}_i, \qquad \text{and} \qquad \tau(\psi_1, \ldots, \psi_n) \;\text{ by }\; \psi_1 \otimes \ldots \otimes \psi_n.$$

In the proof of Lemma 6.3 we constructed a specific realization. However, this is not always the most convenient to work with. For instance, when $\mathcal{H}_i = L^2(\mathcal{X}_i, \mathcal{A}_i, \mu_i)$ then

$$\overline{\otimes}_{i=1}^n \mathcal{H}_i = L^2(\times_{i=1}^n \mathcal{X}_i, \otimes_{i=1}^n \mathcal{A}_i, \otimes_{i=1}^n \mu_i),$$

with

$$f_1 \otimes \ldots \otimes f_n := f_1 \,\dot{\otimes}\, \ldots \,\dot{\otimes}\, f_n : (x_1, \ldots, x_n) \mapsto f_1(x_1) \cdot \ldots \cdot f_n(x_n)$$

is a more natural realization of the tensor product; see Section 4.3.

Definition 6.4. Let $\mathcal{H}_i$, $i \in [n]$, be Hilbert spaces, and for each $i \in [n]$, let $\mathcal{K}_i$ be a subspace of $\mathcal{H}_i$. We define the *algebraic tensor product* of the $\mathcal{K}_i$ as

$$\mathcal{K}_1 \otimes \ldots \otimes \mathcal{K}_n := \operatorname{span}\{\psi_1 \otimes \ldots \otimes \psi_n : \psi_i \in \mathcal{K}_i, \, i \in [n]\}. \tag{6.143}$$

It is straightforward from the definitions that if each $\mathcal{K}_i$ in Definition 6.4 is closed (i.e., a Hilbert space) then the closure of the subspace in (6.143) gives a realization of the tensor product of $\mathcal{K}_1, \ldots, \mathcal{K}_n$ with the natural n-linear map $\tau(\psi_1, \ldots, \psi_n) := \psi_1 \otimes \ldots \otimes \psi_n$, and hence we write

$$\overline{\mathcal{K}_1 \otimes \ldots \otimes \mathcal{K}_n} = \mathcal{K}_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \mathcal{K}_n.$$

More generally, we have the following:

Exercise 6.5. Show that in the setting of Definition 6.4,

$$\overline{\mathcal{K}_1 \otimes \ldots \otimes \mathcal{K}_n} = \overline{\mathcal{K}}_1 \,\overline{\otimes}\, \ldots \,\overline{\otimes}\, \overline{\mathcal{K}}_n,$$

with the natural n-linear map $\tau(\psi_1, \ldots, \psi_n) := \psi_1 \otimes \ldots \otimes \psi_n$.

6.2 The spin chain

The observable algebra

In this section we give a description of the observable algebra of an infinite spin system. To avoid technical difficulties of working with infinite tensor product of Hilbert spaces, we choose a C^*-algebraic description instead of working on the Hilbert space level. An additional advantage of this choice is that quantum and classical chains can be treated within the same formalism.

Let $\mathcal{C}_0$ denote the observable algebra of a single spin, located at one site of the infinite lattice $\mathbb{Z}$. For a classical spin-$\frac{1}{2}$ system $\mathcal{C}_0$ is the commutative algebra $\mathcal{F}(\{-1, 1\})$, while for a quantum spin-s system $\mathcal{C}_0 = \mathcal{B}(\mathcal{H})$ where $\mathcal{H}$ is a $2s+1$-dimensional Hilbert space. The observable algebra of the spins located at the sites of a finite subset $\Lambda \subset \mathbb{Z}$ is $\mathcal{C}_\Lambda := \otimes_{k \in \Lambda} \mathcal{C}_0$. For any pair $\Lambda \subset \Lambda'$ the observable algebra $\mathcal{C}_\Lambda$ can naturally be viewed as a subalgebra of $\mathcal{C}_{\Lambda'}$. More precisely, there exists a unity-preserving embedding $\iota_{\Lambda',\Lambda}$ of $\mathcal{C}_\Lambda$ into $\mathcal{C}_{\Lambda'}$, given by the formula

$$\iota_{\Lambda',\Lambda}(A) := A \otimes \left(\bigotimes_{l \in \Lambda' \setminus \Lambda} I \right) \qquad \text{for} \quad A \in \mathcal{C}_\Lambda.$$

The union of the C_Λ's can be equipped with a natural equivalence relation; for $A \in \mathcal{C}_{\Lambda_1}$, $B \in \mathcal{C}_{\Lambda_2}$ we say that A and B are equivalent, if there exists a bigger subset Λ such that the embedded images of A and B in $\mathcal{C}_\Lambda$ are the same, i.e.

$$\iota_{\Lambda,\Lambda_1}(A) = \iota_{\Lambda,\Lambda_2}(B).$$

Factorization with this equivalence relation yields a normed $*$-algebra, in which also the C^*-property is satisfied. This algebra is called the *algebra of local observables*,

and its norm completion is by definition the spin chain algebra $\mathcal{C}$. By identifying elements in different $\mathcal{C}_\Lambda$'s with their equivalence classes, we get an embedding of the finite algebras into the infinite one, which is compatible with the embeddings of the finite algebras into each other, i.e. we have embeddings $\iota_\Lambda : \mathcal{C}_\Lambda \to \mathcal{C}$ such that $\iota_\Lambda = \iota_{\Lambda'} \circ \iota_{\Lambda',\Lambda}$ holds. For the rest we identify $\mathcal{C}_\Lambda$ with its embedded image and in the following we use the same notation $\mathcal{C}_\Lambda$ for it. With this identification the local algebra can be written as the union of all the $\mathcal{C}_\Lambda$'s. Elements of the local algebra can be pictured as linear combinations of elements of the form

$$\ldots I \otimes I \otimes A_{i_1} \otimes \ldots \otimes A_{i_n} \otimes I \otimes I \otimes \ldots,$$

where $i_1, \ldots, i_n \in \mathbb{Z}$ and $A_{i_k} \in \mathcal{C}_0$.

Note that the local algebra can also be given as the union of all subalgebras of the form $\mathcal{C}_{\{-n,\ldots,n\}}$. The half-infinite chain $\mathcal{C}_\mathbb{N}$ is defined as the C^*-subalgebra of $\mathcal{C}$, generated by the union of all subalgebras of the form $\mathcal{C}_{\{0\ldots,n\}}$; $\quad n \geq 0$.

Example 6.6. The classical chain

In the classical case, a spin-$\frac{1}{2}$ particle is modeled by a two-level system with configuration space $\{-1,+1\}$. The confuguration space of a system of spins located at the sites corresponding to a finite subset $\Lambda \subset \mathbb{Z}$ is $\Omega_\Lambda := \{-1,+1\}^\Lambda$; the observable algebra $\mathcal{F}(\Omega_\Lambda)$ is isomorphic to $\mathcal{C}_\Lambda = \otimes_{k\in\Lambda} \mathcal{F}(\{-1,1\})$. If we endow the finite set $\{-1,+1\}$ with the discrete topology, then the configuration space of the infinite system $\Omega := \{-1,+1\}^\mathbb{Z}$ becomes a topological space with the product topology, which is homeomorphic to $C \times C$, where C is the Cantor set

$$C := \left\{ \sum_{n\in\mathbb{N}} \frac{x_n}{3^n} \;:\; x_n \in \{0,2\},\ n \in \mathbb{N} \right\} \subset [0,1];$$

an explicit homeomorphism is given by

$$h(\underline{\omega}) := \left(\sum_{n=-1}^{-\infty} \frac{\omega_n + 1}{3^{-n}}; \sum_{n=0}^{+\infty} \frac{\omega_n + 1}{3^{n+1}} \right); \qquad \underline{\omega} \in \Omega.$$

The Cantor set can be written in the form

$$C = \bigcap_{n\in\mathbb{N}} \bigcup_{k=0}^{2^n-1} I_{k,n},$$

where $I_{k,n}$ is an interval of the form $\left[\sum_{l=1}^n \frac{x_l}{3^l}, \sum_{l=1}^n \frac{x_l}{3^l} + \frac{1}{3^n}\right]$ with $\sum_{l=1}^n x_l 2^{l-1} = k$.

The algebra $\mathcal{C}_\Lambda$ can be identified with functions on Ω that depend only on the coordinates corresponding to Λ; these functions are obviously continuous on Ω. In

the case $\Lambda = [-n, n-1]$ the elements of $\mathcal{C}_\Lambda$ can also be pictured as functions on $C \times C$ that are constants on the sets $(I_{k,n} \times I_{l,n}) \cap (C \times C)$. Since the collection of these functions for all n forms a dense set in $\mathcal{F}_c(C \times C)$ then the observable algebra $\mathcal{C}$ can be identified by the commutative C^*-algebra $\mathcal{F}_c(C \times C)$. The algebra of the left (as well as the right) half-chain in this picture is naturally isomorphic to $\mathcal{F}_c(C)$. Note that since $C \times C$ is homeomorphic to C then also $\mathcal{C}$ is isomorphic to $\mathcal{F}_c(C)$.

The quantum spin chain algebra can also be given as the universal C^*-algebra[1] corresponding to a set of generators and relations. Two canonical choices are the *spin operators* $\{S_j^n : n \in \mathbb{Z};\ j = 1, 2, 3\}$, satisfying relations

$$\text{(R1)}\quad (S_j^n)^* = S_j^n \qquad \forall j\ \forall n \tag{6.144}$$

$$\text{(R2)}\quad [S_j^n, S_k^m] = \begin{cases} 0\,, & \text{if } n \neq m; \\ i\epsilon_{jkl} S_l^n\,, & \text{if } n = m, \end{cases} \tag{6.145}$$

and the *commuting matrix units* $\{e_{ij}^n : n \in \mathbb{Z};\ i, j = 1, \ldots, d\}$, satisfying relations

$$\text{(R1')}\quad (e_{ij}^n)^* = e_{ji}^n \tag{6.146}$$

$$\text{(R2')}\quad e_{ij}^n\, e_{kl}^m = \begin{cases} e_{kl}^m\, e_{ij}^n\,, & \text{if } n \neq m; \\ \delta_{jk}\, e_{il}^n\,, & \text{if } n = m \end{cases} \tag{6.147}$$

$$\text{(R3')}\quad \sum_{i=1}^{2s+1} e_{ii} = I. \tag{6.148}$$

Here $[.,.]$ denotes the commutator, δ_{ij} is the Kronecker delta, and ϵ_{jkl} is the Levi-Civita symbol. A set of matrix units can be given in $\mathcal{C}_\Lambda$ by the definition

$$e_{\boldsymbol{ij}} := \prod_{n \in \Lambda} e_{i_n\, j_n}; \qquad\qquad \boldsymbol{i}, \boldsymbol{j} \in \{1, \ldots, d\}^\Lambda\,. \tag{6.149}$$

For the quantum spin-$\frac{1}{2}$ case it is convenient to introduce the *Pauli operators*

$$\sigma_x^n := \sigma_1^n := 2\, S_1^n; \qquad \sigma_y^n := \sigma_2^n := 2\, S_2^n; \qquad \sigma_z^n := \sigma_3^n := 2\, S_3^n\,; \tag{6.150}$$

[1]For the definition of a universal C^*-algebra see Appendix A.1.

a commuting set of matrix units can then be given by

$$e^n_{11} := \frac{1}{2}\left(I + \sigma^n_z\right); \qquad e^n_{12} := \tfrac{1}{2}\left(\sigma^n_x + i\sigma^n_y\right);$$
$$e^n_{21} := \frac{1}{2}\left(\sigma^n_x - i\sigma^n_y\right); \qquad e^n_{22} := \tfrac{1}{2}\left(I - \sigma^n_z\right). \tag{6.151}$$

The *right shift* is defined on the generators as $\gamma\left(S^n_k\right) := S^{n+1}_k$, and is easily seen to extend to an automorphism of the spin chain. For the half-infinite chain the right shift is defined the same way, and its effect on local elements can be pictured as

$$\gamma : \; A_1 \otimes \ldots \otimes A_n \otimes I \otimes \ldots \mapsto I \otimes A_1 \otimes \ldots \otimes A_n \otimes I \otimes \ldots \; .$$

States on the spin chain

For a state φ on the spin chain the restrictions $\varphi_\Lambda := \varphi \restriction_{\mathcal{C}_\Lambda}$ form a *compatible family*, i.e. $\varphi_{\Lambda'}(A) = \varphi_\Lambda(A)$ holds when $\Lambda \subset \Lambda'$ and $A \in \mathcal{C}_\Lambda$. On the other hand, given a compatible family $\{\varphi_\Lambda\}$ of states, there exists a unique state φ on $\mathcal{C}$ such that $\varphi_\Lambda = \varphi \restriction_{\mathcal{C}_\Lambda} \quad \forall \Lambda$ holds. To determine a state of the spin chain it is actually sufficient to know all the restrictions to subalgebras of the form $\mathcal{C}_{\{-n,\ldots,n\}}$ (or $\mathcal{C}_{\{0,\ldots,n\}}$ in the half-infinite case). The density matrices of the states φ_Λ are given by

$$[\varphi_\Lambda]_{\boldsymbol{i},\boldsymbol{j}} = \varphi_\Lambda\left(e_{\boldsymbol{j}\boldsymbol{i}}\right); \qquad \boldsymbol{i},\boldsymbol{j} \in \{1,2\}^\Lambda.$$

A state φ on the (half-)infinite spin chain is called *translation-invariant* (or *shift-invariant*), if

$$\varphi \circ \gamma = \varphi$$

holds for the right shift γ. To specify a shift-invariant state it suffices to know all the restrictions

$$\varphi_n := \varphi \restriction_{\mathcal{C}_{\{0,\ldots,n-1\}}}$$

even for the two-sided infinite chain $\mathcal{C}$. Note that the restriction to $\mathcal{C}_{\mathbb{N}}$ of a shift-invariant state on $\mathcal{C}$ gives a shift-invariant state, while a shift-invariant state on $\mathcal{C}_{\mathbb{N}}$ can be uniquely extended to a shift-invariant state of $\mathcal{C}$, giving a one-to one correspondence between shift-invariant states of the full chain and the right half chain.

The set of translation-invariant states is convex; its extremal points are called *ergodic states*. Note that a pure translation-invariant state is extremal in the convex set of all states, and so it is also ergodic. An equivalent characterization of ergodicity is the following [**?**]:

Theorem 6.7. A state φ is ergodic if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi\left(a\gamma^k(b)\right) = \varphi(a)\varphi(b)\,, \qquad a,b \in \mathcal{C}. \tag{6.152}$$

Note that by continuity it is enough to check (6.152) for local elements a, b.

Example 6.8. Given a state ϱ on $\mathcal{C}_0$, we can define a compatible family of states by

$$\varphi_n(A_0 \otimes \ldots \otimes A_{n-1}) = \varrho(A_0) \cdot \ldots \cdot \varrho(A_{n-1}); \qquad A_0, \ldots, A_{n-1} \in \mathcal{C}_0.$$

The resulting state on $\mathcal{C}$ is denoted by $\varrho^{\otimes\infty}$. States of this form are called (shift-invariant) *product states*.

A product state is always ergodic, as easily seen from the characterization (6.152), and is pure if and only if ϱ is a pure state on $\mathcal{C}_0$.

By Example (6.6) and Example (**??**) a pure state on the classical spin-$\frac{1}{2}$ chain is a Dirac measure concentrated on an infinite configuration $\underline{\omega}$, and therefore it is shift-invariant if and only if $\omega_k = 1 \ \ \forall k$ or $\omega_k = -1 \ \ \forall k$. On the quantum spin chain, however, there is an abundance of shift-invariant pure states, showing a far richer structure of the state space of the quantum chain.

It is in general rather difficult to check properties (e.g. purity or ergodicity) of a state, resulting from a compatible family $\{\varrho_n\}$. Two important classes with well handable criteria are the quasi-free states, presented in section 7.4, and the finitely correlated states, that contain quantum Markov states as a subclass, which in turn contains classical Markov states.

Example 6.9. Finitely correlated states

Let $\mathcal{B}$ be a finite dimensional C^*-algebra with a state ϱ on it, and $\mathbb{E} : \mathcal{A}\otimes\mathcal{B} \to \mathcal{B}$ be a unital CP map, where $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ is the one-site algebra of the spin chain; hence $\mathbb{E}^* : \mathcal{B} \to \mathcal{A} \otimes \mathcal{B}$ is a stochastic map. $\mathbb{E}$ is related to ϱ such that

$$\mathrm{Tr}_{\mathcal{A}}\ \mathbb{E}^*(\varrho) = \varrho \tag{6.153}$$

holds. Let

$$\begin{aligned}
\omega_1 &:= \mathbb{E}^*(\varrho) \\
\omega_2 &:= (\mathrm{id}_{\mathcal{A}} \otimes \mathbb{E}^*) \circ \mathbb{E}^*(\varrho) \\
&\vdots \\
\omega_n &:= \left(\mathrm{id}_{\mathcal{A}}^{\otimes(n-1)} \otimes \mathbb{E}^*\right) \circ \ldots \circ (\mathrm{id}_{\mathcal{A}} \otimes \mathbb{E}^*) \circ \mathbb{E}^*(\varrho)\ ;
\end{aligned}$$

here ω_n is defined on $\mathcal{A}^{\otimes n} \otimes \mathcal{B}$. To obtain a family on the spin chain, we take

$$\varphi_n := \mathrm{Tr}_{\mathcal{B}}\ \omega_n;$$

on simple product operators it takes the value

$$\varphi_n(A_1 \otimes \ldots \otimes A_n) = \varrho\left(\mathbb{E}\left(A_1 \otimes \mathbb{E}\left(A_2 \otimes \ldots \mathbb{E}\left(A_n \otimes I_{\mathcal{B}}\right)\ldots\right)\right)\right) .$$

Compatibility of this family is guaranteed by the unitality of $\mathbb{E}$, while shift-invariance follows from (6.153). The induced state φ on the spin chain is called a finitely correlated state. By lemma 2.5, [?], $\mathcal{B}$ can always be taken to be $\mathcal{B}(\mathcal{K})$ for some finite dimensional Hilbert space $\mathcal{K}$, and ϱ can be assumed to be faithful. Note that the construction of the sequence ω_n resembles very much to that of Markov states (Example (6.10)). However, the construction of the sequence φ_n doesn't follow the scheme for Markovian states, as φ_{n+1} is derived from ω_{n+1} instead of φ_n. As a consequence, the resulting state φ is in general not Markovian.

A finitely correlated state is purely generated, if $\mathbb{E} = \mathrm{Ad}_{V^*}$ for an isometry $V : \mathcal{K} \to \mathcal{H} \otimes \mathcal{K}$. A purely generated state is pure if and only if the spectrum of $\hat{\mathbb{E}} : b \mapsto \mathbb{E}(I_{\mathcal{A}} \otimes b)$ intersects the complex unit circle only at $\{1\}$, and 1 is a simple eigenvalue.

A finitely correlated state is ergodic, if and only if 1 is a simple eigenvalue of $\hat{\mathbb{E}}$ (Proposition 3.1, [?]).

A special subclass is when the algebra $\mathcal{B}$ is a commutative one, isomorphic to $\mathcal{F}(\mathcal{X})$ for some finite set $\mathcal{X}$. $\mathbb{E}^*$ is specified by its values on the Dirac measures δ_x, and since the result is a state on $\mathcal{A} \otimes \mathcal{F}(\mathcal{X})$, it can uniquely be written as $\sum_y p_{xy}\, \varrho_{xy} \otimes \delta_y$, where $\{p_{xy}\}$ is a probability distribution for a fixed x, and ϱ_{xy} are states on $\mathcal{A}$. The state ϱ is of the form $\varrho = \sum_x p_x\, \delta_x$, and (6.153) is equivalent to $\{p_x\}$ being an invariant measure of the stochastic map with matrix $T_{x,y} := p_{xy}$. The resulting states are of the form

$$\varphi_n = \sum_{\{x_1,\ldots,x_{n+1}\}} \mu(x_1, \ldots, x_{n+1})\, \varrho_{x_1\, x_2} \otimes \ldots \otimes \varrho_{x_n\, x_{n+1}} ,$$

where μ is the classical Markov measure, generated by $\{p_x\}$ and T. In this case $\hat{\mathbb{E}}$ is the linear map on $\mathcal{F}(\mathcal{X})$ with matrix T, thus φ is ergodic if and only if 1 is a simple eigenvalue of T, which is equivalent to the ergodicity of the classical measure μ. We obtain a special form when ϱ_{xy} is independent of x; in this case

$$\varphi_n = \sum_{\{x_1,\ldots,x_n\}} \mu(x_1, \ldots, x_n)\, \varrho_{x_1} \otimes \ldots \otimes \varrho_{x_n} . \tag{6.154}$$

Example 6.10. Markov states

A shift-invariant state ϱ on the infinite spin chain is called a Markov state if there exists a stochastic map $\alpha : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ such that

$$\varrho_{n+1} = \left(\mathrm{id}_{\mathcal{A}}^{\otimes(n-1)} \otimes \alpha\right)(\varrho_n) \,. \tag{6.155}$$

Note that the criterion for Markovianity doesn't provide a method to construct Markovian states, as for a given map α and state φ_1 the sequence $\varphi_1 := \alpha(\varphi_1)$; $\varphi_2 := (\mathrm{id}_{\mathcal{A}} \otimes \alpha) \circ \alpha(\varphi_1)$; ... is in general not a compatible one.

Markov states are finitely correlated. Indeed, the choice

$$\mathcal{B} := \mathcal{A}; \qquad \mathbb{E} := \alpha^* \qquad \text{and} \qquad \varrho := \varphi_1$$

yields $\omega_n = \varphi_{n+1}$, and $\mathrm{Tr}_{\mathcal{B}}\, \omega_n = \mathrm{Tr}_{n+1}\, \varphi_{n+1} = \varphi_n$. Compatibility and shift-invariance implies

$$\mathrm{Tr}_1\, \alpha(\varphi_1) = \mathrm{Tr}_2\, \alpha(\varphi_1) = \varphi_1 \,.$$

Note that (6.155) is equivalent to ϱ_{n+2} saturating (SSA) with respect to the partition $\mathcal{A}^{\otimes n} \otimes \mathcal{A} \otimes \mathcal{A}$ (see Chapter **??** for details), therefore the algebraic characterization (Theorem **??**) can be used to construct Markov states. We can specify a Markov state by

- a splitting $\mathcal{H} = \bigoplus_{k=1}^{K} \mathcal{H}_k^{\mathrm{L}} \otimes \mathcal{H}_k^{\mathrm{R}}$;
- a set of states $\{\varrho_{kl} : k, l = 1, \ldots, K\}$, where ϱ_{kl} is a state on $\mathcal{H}_k^{\mathrm{R}} \otimes \mathcal{H}_l^{\mathrm{L}}$;
- a classical shift-invariant Markov measure μ on $\{1, \ldots, K\}^{\infty}$.

Let

$$\varrho_l^{\mathrm{L}} := \sum_k \frac{\mu(k,l)}{\mu(l)}\, \mathrm{Tr}_{\mathcal{H}_k^{\mathrm{R}}}\, \varrho_{kl} \qquad \text{and} \qquad \varrho_k^{\mathrm{R}} := \sum_l \frac{\mu(k,l)}{\mu(k)}\, \mathrm{Tr}_{\mathcal{H}_l^{\mathrm{L}}}\, \varrho_{kl}; \tag{6.156}$$

Then

$$\begin{aligned} \varrho_1 &= \bigoplus_k \mu(k)\, \varrho_k^{\mathrm{L}} \otimes \varrho_k^{\mathrm{R}} \,, \qquad \text{and} \\ \varrho_n &= \bigoplus_{\{k_1,\ldots,k_n\}} \mu(k_1 \ldots, k_n)\, \varrho_{k_1}^{\mathrm{L}} \otimes \varrho_{k_1 k_2} \otimes \ldots \otimes \varrho_{k_{n-1} k_n} \otimes \varrho_{k_n}^{\mathrm{R}} \,. \end{aligned} \tag{6.157}$$

Markovianity of μ implies $\mu(k_1, \ldots, k_n) = \mu(k_{n-1})\, \frac{\mu(k_1,\ldots,k_{n-1})}{\mu(k_{n-1})}\, \frac{\mu(k_{n-1},k_n)}{\mu(k_{n-1})}$, hence

$$\begin{aligned} \varrho_n = \bigoplus_{k_{n-1}} \mu(k_{n-1}) \left[\bigoplus_{\{k_1,\ldots,k_{n-2}\}} \frac{\mu(k_1, \ldots, k_{n-1})}{\mu(k_{n-1})}\, \varrho_{k_1}^{\mathrm{L}} \otimes \varrho_{k_1 k_2} \otimes \ldots \otimes \varrho_{k_{n-2} k_{n-1}} \right] \\ \otimes \left[\bigoplus_{k_n} \frac{\mu(k_{n-1}, k_n)}{\mu(k_{n-1})}\, \varrho_{k_{n-1} k_n} \otimes \varrho_{k_n}^{\mathrm{R}} \right] , \end{aligned}$$

therefore it satisfies the structural criterion (**??**) for all $n \geq 2$.

Symmetries

As in the general theory of dynamical systems, symmetries are described by the same mathematical tool as the dynamical evolution of the system, i.e. by (a group of) automorphisms. Symmetries can be local, that is, automorphisms of a local algebra $\mathcal{C}_\Lambda$, or global, i.e. automorphisms of $\mathcal{C}$. If a global automorphism preserves the local structure of the spin chain then it can be restricted to give a local automorphism on each local algebra. As for states, the thermodynamical limit of a family of local automorphisms can be defined. If a given family of local automorphisms leaves the local Hamiltonians invariant, then in an ideal case the same holds for the thermodynamical limit automorphism and the thermodynamical limit of the equilibrium states. Here we describe in some detail the automorphism group of rotations, as it is going to have some importance later on.

The map

$$\kappa_w : \ \sigma_x^k \mapsto \cos\vartheta\ \sigma_x^k + \sin\vartheta\ \sigma_y^k; \quad \sigma_y^k \mapsto -\sin\vartheta\ \sigma_x^k + \cos\vartheta\ \sigma_y^k; \quad \sigma_z^k \mapsto \sigma_z^k$$

for $w = e^{i\vartheta}$; $\vartheta \in [0, 2\Pi)$ describes a rotation of angle ϑ in the xy plane, and extends to an automorphism of the infinite chain for every w. Note that the definition of κ_w depends on the choice of spin operators σ_x, σ_y, σ_z, i.e. on the choice of the basis. The map $w \mapsto \kappa_w$ gives a representation of the complex unit circle $\mathbb{T}$ in the automorphism group of the spin chain, which we will call the *rotation group* for the sake of simplicity. Since $\kappa_w(\mathcal{C}_\Lambda) = \mathcal{C}_\Lambda$, it can be restricted to an automorphism $(\kappa_w)_\Lambda$ of the algebra $\mathcal{C}_\Lambda$.

The set $\{-1, 1\}$ forms an order two subgroup of the unit circle; the order two automorphism κ_{-1} we call the *parity automorphism*. It describes a rotation of angle 180° in the xy plane. The elements of its fixed point algebra are called *even*, while elements for which $\kappa_{-1}(a) = -a$ are called *odd*. Every element can uniquely be decomposed into a sum of an even and an odd element in the form

$$a = a_+ + a_-; \qquad a_+ = \frac{1}{2}\left(a + \kappa_{-1}(a)\right), \qquad a_- = \frac{1}{2}\left(a - \kappa_{-1}(a)\right).$$

Product of even elements is even again, therefore the set of even elements forms a subalgebra $\mathcal{C}_+$, while the set of odd elements is not closed under multiplication, and therefore not an algebra. A matrix unit e_{ij}^n is even, if $i = j$, and odd, if $i \neq j$; thus matrix units in $\mathcal{C}_\Lambda$ of the form (6.149) are even (odd) if and only if $\sum_{n\in\Lambda}(i_n - j_n)$ is even (odd).

We call a state φ *rotation-invariant*, if $\varphi\circ\kappa_w = \varphi \ \ \forall w \in \mathbb{T}$, and *even*, if $\varphi\circ\kappa_{-1} = \varphi$. Note that a state is even if and only if it vanishes on odd elements, therefore for the entries of the density matrices of an even state we have

$$[\varphi_\Lambda]_{\boldsymbol{i},\boldsymbol{j}} = 0, \qquad \text{if} \qquad \sum_{k\in\Lambda}(i_k - j_k) \ \text{ is odd.}$$

Physical models

Mathematical models describing effectively one-dimensional spin systems are in general specified by a sequence of local Hamiltonians $H_N \in \mathcal{C}_{[-N,N]}$. The thermal equilibrium state of a finite subsystem at inverse temperature β is then the Gibbs state $\varrho_{N,\beta}$ with density $e^{-\beta H_N} / \operatorname{Tr} e^{-\beta H_N}$. A state ϱ is a ground state, if it minimizes the expectation value of H_N. It is easily seen that ϱ is a ground state if and only if its support is contained in the lowest-energy eigensubspace of H_N. The zero-temperature limit $\varrho_{N,\infty} := \lim_{\beta\to\infty} \varrho_{N,\beta}$ of the Gibbs states gives a particular ground state. The thermodynamical limit of the states $\varrho_{N,\beta}$ is defined on local elements as

$$\varrho_\beta(A) := \lim_{N\to\infty} \varrho_{N,\beta}(A); \qquad\qquad A \in \mathcal{C}_{\mathrm{loc}}\,. \tag{6.158}$$

For a detailed analysis of conditions that guarantee the existence of the above limit, see [?].

Example 6.11. The XY model describes nearest-neighbor interaction of spins in the presence of a homogeneous z-directional magnetic field h. It is specified by a sequence of local Hamiltonians of the form

$$H_N := \sum_{k=-N}^{N-1} \left(\frac{1+\delta}{2}\, \sigma_x^k \sigma_x^{k+1} + \frac{1-\delta}{2}\, \sigma_y^k \sigma_y^{k+1} \right) + h \sum_{k=-N}^{N} \sigma_z^k \tag{6.159}$$

where δ is a real parameter. Special cases are the Ising model, where $|\delta| = 1$, and the XX-model, where $\delta = 0$.

The parity automorphism leaves the local Hamiltonians of the XY-model invariant, i.e. κ_{-1} is a symmetry of the local systems, and the local Gibbs states (as well as the ground states) are even, and these two properties carry through to the thermodynamical limit. The local Hamiltonians of the XX-model are invariant under the whole rotation group, and so the local Gibbs and the ground states are rotation-invariant, the and the same holds after taking the thermodynamical limit.

The first solution of the ground state problem of the XY model was given in [?] for the case $h = 0$. The case $h \neq 0$ was studied for the Ising model in [?], and for the general XY model in [?]. In Appendix C.1 we present an explicit computation, following the method of [?], that shows that the ground state of the XX model can be identified as a pure translation-invariant quasi-free state (see section 7.4 for quasi-free states on the spin chain).

6.3 Symmetric and antisymmetric tensors, Fock spaces

Let S_n denote the permutation group of $[n] := \{1, \ldots, n\}$. If $\mathcal{H}$ is a Hilbert space with an ONB $(e_i)_{i\in\mathcal{I}}$, then $(e_{i_1} \otimes \ldots \otimes e_{i_n})_{\underline{i}\in\mathcal{I}^n}$ is an ONB in $\mathcal{H}^{\overline{\otimes} n}$, according to Lemma

6.2. Moreover, for any $\sigma \in S_n$, $(e_{i_{\sigma^{-1}(1)}} \otimes \ldots \otimes e_{i_{\sigma^{-1}(n)}})_{\underline{i} \in \mathcal{I}^n}$ is also an ONB in $\mathcal{H}^{\overline{\otimes} n}$, since it is simply a permutation of the original basis elements. Hence, there is a unique unitary operator $U_{\sigma,\mathcal{H}}$ on $\mathcal{H}^{\overline{\otimes} n}$ such that

$$U_{\sigma,\mathcal{H}}(e_{i_1} \otimes \ldots \otimes e_{i_n}) = e_{i_{\sigma^{-1}(1)}} \otimes \ldots \otimes e_{i_{\sigma^{-1}(n)}}, \qquad \underline{i} \in \mathcal{I}^n.$$

Remark 6.12. When the Hilbert space $\mathcal{H}$ is clear from the context, or if it is irrelevant, we will omit it from the notation, and simply write U_σ instead of $U_{\sigma,\mathcal{H}}$.

The following is straightforward to verify.

Exercise 6.13. (i) Show that $U_\sigma = U_{\sigma,\mathcal{H}}$ is the unique bounded linear operator on $\mathcal{H}^{\overline{\otimes} n}$ satisfying

$$U_\sigma\,(\psi_1 \otimes \ldots \otimes \psi_n) = \psi_{\sigma^{-1}(1)} \otimes \ldots \otimes \psi_{\sigma^{-1}(n)}, \qquad \psi_k \in \mathcal{H},\ k \in [n].$$

(ii) Show that $S_n \ni \sigma \mapsto U_\sigma$ is a unitary representation of S_n on $\mathcal{H}^{\overline{\otimes} n}$, i.e.,

$$U_{\mathrm{id}} = I, \quad U_{\sigma_1} U_{\sigma_1} = U_{\sigma_1\sigma_2}, \quad \sigma_1, \sigma_2 \in S_n.$$

In particular, $U_{\sigma^{-1}} = U_\sigma^{-1} = U_\sigma^*$.

Remark 6.14. More generally, for any vector space V, and any $\sigma \in S_n$, the map

$$V^n \ni (v_1, \ldots, v_n) \mapsto v_{\sigma^{-1}(1)} \otimes \ldots \otimes v_{\sigma^{-1}(n)}$$

is n-linear, and hence it has a unique linearization $U_{\sigma,V}$ through the algebraic tensor product $V^{\otimes n}$, given by

$$U_{\sigma,V}\,(v_1 \otimes \ldots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes \ldots \otimes v_{\sigma^{-1}(n)}, \qquad v_k \in V,\ k \in [n].$$

It is easy to verify that $S_n \ni \sigma \mapsto U_{\sigma,V}$ is a representation of S_n on $V^{\otimes n}$.

Moreover, it is easy to see that if $V = \mathcal{H}$ is a Hilbert space then $U_{\sigma,\mathcal{H}}$ is isometric (w.r.t. the natural norm on $\mathcal{H}^{\otimes n}$) on the subspace spanned by product vectors. Thus, it has a unique unitary extension to $\mathcal{H}^{\overline{\otimes} n}$, and it coincides with the $U_{\sigma,\mathcal{H}}$ defined above.

Recall that the *sign* $\varepsilon(\sigma)$ of a permutation $\sigma \in S_n$ is defined as

$$\varepsilon(\sigma) := (-1)^{|\{i<j:\,\sigma(i)>\sigma(j)\}|},$$

where the number in the exponent is the number of *inversions* in σ. In particular, $\varepsilon(\sigma)$ is equal to $+1$ or -1. It is not too difficult to see that ε gives a 1-dimensional representation of S_n, as

$$\varepsilon(\sigma_1\sigma_2) = \varepsilon(\sigma_1)\varepsilon(\sigma_2), \qquad \sigma_1, \sigma_2 \in S_n.$$

Definition 6.15. A vector $\psi \in \mathcal{H}^{\overline{\otimes} n}$ is called *symmetric*, if

$$U_\sigma \psi = \psi \qquad \forall \sigma \in S_n,$$

and *antisymmetric*, if

$$U_\sigma \psi = \varepsilon(\sigma)\psi \qquad \forall \sigma \in S_n.$$

We denote the set of symmetric vectors in $\mathcal{H}^{\overline{\otimes} n}$ by $\vee^n \mathcal{H}$, and the set of antisymmetric vectors by $\wedge^n \mathcal{H}$.

Exercise 6.16. Show that both $\vee^n \mathcal{H}$ and $\wedge^n \mathcal{H}$ are closed subspaces.

Exercise 6.17. Show that the operators

$$P_s^{(n)} := \frac{1}{n!} \sum_{\sigma \in S_n} U_\sigma \quad \text{and} \quad P_a^{(n)} := \frac{1}{n!} \sum_{\sigma \in S_n} \varepsilon(\sigma) U_\sigma$$

are self-adjoint projections with ranges $\vee^n \mathcal{H}$ and $\wedge^n \mathcal{H}$, respectively. Show that $P_a^{(n)} P_s^{(n)} = 0$ and conclude that $\wedge^n \mathcal{H} \perp \vee^n \mathcal{H}$.

Solution: Hidden.

For $\psi_1, \ldots, \psi_n$, we use the notations

$$\psi_1 \vee \ldots \vee \psi_n := \sqrt{n!} P_s^{(n)} (\psi_1 \otimes \ldots \otimes \psi_n) = \frac{1}{\sqrt{n!}} \sum_{\sigma \in S_n} \psi_{\sigma(1)} \otimes \ldots \otimes \psi_{\sigma(n)}$$

$$\psi_1 \wedge \ldots \wedge \psi_n := \sqrt{n!} P_a^{(n)} (\psi_1 \otimes \ldots \otimes \psi_n) = \frac{1}{\sqrt{n!}} \sum_{\sigma \in S_n} \varepsilon(\sigma) \psi_{\sigma(1)} \otimes \ldots \otimes \psi_{\sigma(n)}.$$

Lemma 6.18.

$$\vee^n \mathcal{H} = \overline{\operatorname{span}}\{\psi_1 \vee \ldots \vee \psi_n : \psi_i \in \mathcal{H},\, i \in [n]\}, \tag{6.160}$$

$$\wedge^n \mathcal{H} = \overline{\operatorname{span}}\{\psi_1 \wedge \ldots \wedge \psi_n : \psi_i \in \mathcal{H},\, i \in [n]\}. \tag{6.161}$$

Proof. By the definition of the tensor product,

$$\mathcal{H}^{\overline{\otimes} n} = \overline{\operatorname{span}}\{\psi_1 \otimes \ldots \otimes \psi_n : \psi_i \in \mathcal{H},\, i \in [n]\}$$

(see (6.140)), whence (6.160) and (6.161) follow due to the linearity and continuity of $P_s^{(n)}$ and $P_a^{(n)}$, respectively. □

Exercise 6.19. Show that for any $\sigma \in S_n$,

$$U_\sigma P_s^{(n)} = P_s^{(n)} = P_s^{(n)} U_\sigma, \qquad U_\sigma P_a^{(n)} = \varepsilon(\sigma) P_a^{(n)} = P_a^{(n)} U_\sigma,$$

and prove that for any $\psi_1, \ldots, \psi_n \in \mathcal{H}$,

$$U_\sigma(\psi_1 \vee \ldots \vee \psi_n) = \psi_{\sigma^{-1}(1)} \vee \ldots \vee \psi_{\sigma^{-1}(n)} = \psi_1 \vee \ldots \vee \psi_n, \tag{6.162}$$

$$U_\sigma(\psi_1 \wedge \ldots \wedge \psi_n) = \psi_{\sigma^{-1}(1)} \wedge \ldots \wedge \psi_{\sigma^{-1}(n)} = \varepsilon(\sigma)\psi_1 \wedge \ldots \wedge \psi_n. \tag{6.163}$$

Conclude that

$$\exists\, i \neq j :\ \psi_i = \psi_j \quad \Longrightarrow \quad \psi_1 \wedge \ldots \wedge \psi_n = 0. \tag{6.164}$$

For two sequences of vectors $\vec{\psi} := (\psi_1, \ldots, \psi_n) \in \mathcal{H}^n$ and $\vec{\phi} := (\phi_1, \ldots, \phi_n) \in \mathcal{H}^n$, we define their *Gram matrix*

$$G\left(\vec{\psi}, \vec{\phi}\right)_{i,j} := \langle \psi_i, \phi_j \rangle, \qquad i, j \in [n].$$

Exercise 6.20. Show that

$$\langle x_1 \wedge \ldots \wedge x_n, y_1 \wedge \ldots \wedge y_n \rangle = \det\left(G\left(\vec{\psi}, \vec{\phi}\right)\right), \tag{6.165}$$

$$\langle x_1 \vee \ldots \vee x_n, y_1 \vee \ldots \vee y_n \rangle = \operatorname{per}\left(G\left(\vec{\psi}, \vec{\phi}\right)\right), \tag{6.166}$$

where $\operatorname{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^n A_{i,\sigma(i)}$ stands for the *permanent* of the matrix A.

Let $\mathcal{I}$ be an arbitrary set. For $\underline{i} \in \mathcal{I}^n$, the *type* of $\underline{i}$ is a probability distribution on $\mathcal{I}$, defined by

$$P_{\underline{i}}(j) := \frac{1}{n} |\{k :\ i_k = j\}|.$$

That is, $P_{\underline{i}}(j)$ is the *frequency* of $j \in \mathcal{I}$ in the sequence $\underline{i} \in \mathcal{I}^n$. $P_{\underline{i}}$ is also called the *empirical distribution* of $\underline{i}$.

Clearly, $P_{\underline{i}}$ is a finitely supported probability distribution, and all its weights can be written as non-negative rational numbers with n in the denominator. It is also clear that

$$P_{(i_1,\ldots,i_n)} = P_{(j_1,\ldots,j_n)} \quad \Longleftrightarrow \quad \exists\, \sigma \in S_n :\ \ j_k = i_{\sigma(k)},\ k \in [n]. \tag{6.167}$$

Proposition 6.21. Let $(e_i)_{i \in \mathcal{I}}$ be an ONS in $\mathcal{H}$, where $\mathcal{I}$ is some totally ordered set (such as $[d] = \{0, \ldots, d-1\}$, $\mathbb{N}$ or $\mathbb{Z}$). Then

$$\{e_{i_1} \wedge \ldots \wedge e_{i_n} :\ \underline{i} \in \mathcal{I}^n,\ i_1 < \ldots < i_n\} \tag{6.168}$$

is an ONS in $\wedge^n\mathcal{H}$, and

$$\left\{\frac{1}{\sqrt{\prod_{j\in\mathcal{I}}(nP_{\underline{i}}(j))!}}e_{i_1}\vee\ldots\vee e_{i_n}:\underline{i}\in\mathcal{I}^n,\,i_1\leq\ldots\leq i_n\right\}\tag{6.169}$$

is an ONS in $\vee^n\mathcal{H}$. Moreover, if $(e_i)_{i\in\mathcal{I}}$ is an ONB in $\mathcal{H}$ then (6.168) and (6.169) define ONBs in $\wedge^n\mathcal{H}$ and $\vee^n\mathcal{H}$, respectively.

Proof. It is clear from (6.165) and (6.166) that (6.168) and (6.169) define orthonormal systems.

Assume that $\{e_i\ :\ i\in I\}$ is an ONB in $\mathcal{H}$, and that $\wedge^n\mathcal{H}\ni\psi\perp e_{i_1}\wedge\ldots\wedge e_{i_n}$ for all $i_1<\ldots<i_n$. Then

$$\begin{aligned}
0&=\langle e_{i_1}\wedge\ldots\wedge e_{i_n},\psi\rangle\\
&=\left\langle U_\sigma(e_{\sigma(i_1)}\wedge\ldots\wedge e_{\sigma(i_n)}),\psi\right\rangle\\
&=\left\langle e_{\sigma(i_1)}\wedge\ldots\wedge e_{\sigma(i_n)},U_{\sigma^{-1}}\psi\right\rangle\\
&=\varepsilon(\sigma^{-1})\left\langle e_{\sigma(i_1)}\wedge\ldots\wedge e_{\sigma(i_n)},\psi\right\rangle\\
&=\varepsilon(\sigma^{-1})\sqrt{n!}\left\langle P_a^{(n)}(e_{\sigma(i_1)}\otimes\ldots\otimes e_{\sigma(i_n)}),\psi\right\rangle\\
&=\varepsilon(\sigma^{-1})\sqrt{n!}\left\langle e_{\sigma(i_1)}\otimes\ldots\otimes e_{\sigma(i_n)},P_a^{(n)}\psi\right\rangle\\
&=\varepsilon(\sigma^{-1})\sqrt{n!}\left\langle e_{\sigma(i_1)}\otimes\ldots\otimes e_{\sigma(i_n)},\psi\right\rangle.
\end{aligned}$$

Thus, $\psi\perp e_{\underline{j}}=e_{j_1}\otimes\ldots\otimes e_{j_n}$ if $j_k\neq_{k\neq l}j_l$. If there exist $k\neq l$ such that $j_k=j_l$ then $0=e_{j_1}\wedge\ldots\wedge e_{j_n}=\sqrt{n!}P_a^{(n)}(e_{j_1}\otimes\ldots\otimes e_{j_n})$, according to (6.164), whence

$$0=\left\langle P_a^{(n)}(e_{j_1}\otimes\ldots\otimes e_{j_n}),\psi\right\rangle=\left\langle e_{j_1}\otimes\ldots\otimes e_{j_n},P_a^{(n)}\psi\right\rangle=\langle e_{j_1}\otimes\ldots\otimes e_{j_n},\psi\rangle.$$

Thus, $\psi\perp e_{\underline{j}}=e_{j_1}\otimes\ldots\otimes e_{j_n}$ for any $\underline{j}\in I^n$, and therefore $\psi=0$, according to Lemma 6.2. □

Remark 6.22. Note that the vectors in (6.169) can be equivalently written as

$$\frac{1}{\sqrt{\prod_{j\in\mathcal{I}}(nP_{\underline{i}}(j))!}}e_{i_1}\vee\ldots\vee e_{i_n}=\frac{\sqrt{n!}}{\sqrt{\prod_{j\in\mathcal{I}}(nP_{\underline{i}}(j))!}}P_s^{(n)}(e_{i_1}\otimes\ldots\otimes e_{i_n}).$$

Corollary 6.23. We have

$$\dim\wedge^n\mathcal{H}=\binom{\dim\mathcal{H}}{n}:=\frac{\dim\mathcal{H}(\dim\mathcal{H}-1)\cdot\ldots\cdot(\dim\mathcal{H}-n+1)}{n!}$$

$$\dim\vee^n\mathcal{H}=\binom{\dim\mathcal{H}+n-1}{n}:=\frac{(\dim\mathcal{H}+n-1)\cdot\ldots\cdot\dim\mathcal{H}}{n!}=\prod_{k=1}^n\left(1+\frac{\dim\mathcal{H}-1}{k}\right)$$

(both of them equal to $\dim\mathcal{H}$ when $\mathcal{H}$ is infinite-dimensional).

Proof. Let $(e_i)_{i\in\mathcal{I}}$ be an arbitrary ONB in $\mathcal{H}$. Since $\mathcal{I}$ can be totally ordered, the assertions follow immediately from Proposition 6.21. □

Corollary 6.24. If $\mathcal{H}$ is finite-dimensional then

$$\dim\left(\wedge^{\dim\mathcal{H}}\mathcal{H}\right)=1,\qquad\text{and}\qquad n>\dim\mathcal{H}\Longrightarrow\wedge^n\mathcal{H}=\{0\},$$

while

$$\dim\mathcal{H}<\dim\vee^2\mathcal{H}<\dim\vee^3\mathcal{H}<\dots$$

Let us now give some equivalent formulations of the basis vectors in (6.168) and (6.169), for which we will need some further simple observations about types.

The *set of n-types* $\mathcal{P}_n(\mathcal{I})$ on $\mathcal{I}$ is defined as the collection of all probability distribution on $\mathcal{I}$ as above, i.e.,

$$\mathcal{P}_n(\mathcal{I}):=\{P_{\underline{i}}\}_{\underline{i}\in\mathcal{I}^n}.$$

We have the natural identifications

$$\mathcal{P}_n\ \equiv\ \left\{\underline{n}\in\mathbb{N}^{\mathcal{I}}:\ |\underline{n}|:=\sum_{k\in\mathcal{I}}n_k=n\right\}\ \equiv\ \mathcal{I}^n/_{\sim},\tag{6.170}$$

where in the last expression the factorization is according to the equivalence relation on the RHS of (6.167), and the correspondences are given by $P(k)=n_k/|\underline{n}|=P_{\underline{i}}(k)$, $k\in\mathcal{I}$.

Now let $(e_i)_{i\in\mathcal{I}}$ be an ONS (ONB) in $\mathcal{H}$. We may introduce

$$e^s_P:=e^s_{\underline{n}}:=\sqrt{\frac{|\underline{n}|!}{\prod\limits_{k\in\mathcal{I}}n_k!}}P^{(n)}_s\left(\otimes_{k:\,n_k>1}e_k^{\otimes n_k}\right)=\frac{1}{\sqrt{\prod_{j\in\mathcal{I}}(nP_{\underline{i}}(j))!}}e_{i_1}\vee\ldots\vee e_{i_n},\tag{6.171}$$

where the n-type P, the sequence $\underline{n}\in\mathbb{N}^{\mathcal{I}}$ and the sequence $\underline{i}=(i_1,\dots,i_n)\in\mathcal{I}^n$ correspond to each other according to (6.170). With these notations, Proposition 6.21 tells that

$$(e^s_P)_{P\in\mathcal{P}_n(\mathcal{I})}\,,\qquad\left(e^s_{\underline{n}}\right)_{\underline{n}\in\mathbb{N}^{\mathcal{I}},|\underline{n}|=n}$$

are orthonormal systems (orthonormal bases) in $\vee^n\mathcal{H}$, and they can be naturally identified according to (6.171) and (6.170). Note that with these parametrizations we actually do not need a total ordering of $\mathcal{I}$ to uniquely specify the basis vectors in $\vee^n\mathcal{I}$.

The situation is slightly different in the antisymmetric case. First, (6.164) implies that here we need to restrict to sequences $\underline{n}$ such that $n_k = 0$ or 1 for each k, or equivalently, to n-types P such that $P(k) = 0$ or $1/n$. Another slight complication stems from the fact that for each $P \in \mathcal{P}_n(\mathcal{I})$, $\mathcal{E}_P := \{e_{i_1} \wedge \ldots \wedge e_{i_n} : P_{\underline{i}} = P\}$ has exactly two elements, which are parallel to each other, according to (6.163), and we need to choose exactly one element from each $\mathcal{E}_P$, $P \in \mathcal{P}_n$, to get an ONS (ONB) in $\wedge^n\mathcal{H}$. Hence, we cannot unambiguously specify an ONS (ONB) in $\wedge^n\mathcal{H}$ from one given in $\mathcal{H}$ without some extra assumption, e.g., a total order on $\mathcal{I}$, as was done in Proposition 6.21. This, however, is not a serious issue, as any set can be totally ordered, and, moreover, the ONB in $\mathcal{H}$ is usually parametrized by $[d]$, $\mathbb{N}$, or $\mathbb{Z}$, anyway, which carry a natural total order. In these cases we may define

$$e_P^a := e_{\underline{n}}^a := \wedge_{k:n_k=1} e_k = e_{k_1} \wedge \ldots \wedge e_{k_n}, \tag{6.172}$$

where $k_1 < \ldots < k_n$ are the k values for which $n_k = 1$.

The ways of writing the basis vectors as in (6.171) and (6.172) is called the *occupational number representation*. The picture behind this terminology is that the e_i represent distinguished physical states of the system, and the sequence $\underline{n}$ tells how many of the n particles occupy each of the given states. The notation

$$|n_1, n_2, \ldots\rangle$$

is also used for either type of basis vectors above; note, however, that the dependence on the choice of ONB in $\mathcal{H}$ is suppressed in this notation.

On top of the symmetrized tensor products, we also have the following generating set for the symmetric subspace:

Proposition 6.25.

$$\vee^n\mathcal{H} = \overline{\operatorname{span}}\{\psi^{\otimes n} : \psi \in \mathcal{H}\}. \tag{6.173}$$

Proof. Let $(e_i)_{i\in\mathcal{I}}$ be an ONB in $\mathcal{H}$. By the above, it is sufficient to prove that $e_P^s \in \overline{\operatorname{span}}\{\psi^{\otimes n} : \psi \in \mathcal{H}\}$ for every $P \in \mathcal{P}_n(\mathcal{I})$. Let $r := |\operatorname{supp}\mathcal{P}(I)|$; then we can identify $\tilde{I}$ with $[r]$, and we may consider P as an element of $\mathcal{P}_n([r])$.

For any $\underline{t} \in \mathbb{R}^r$, let $\psi(\underline{t}) := \sum_{k=1}^r t_k e_k$. Then

$$\begin{aligned}
\psi(\underline{t})^{\otimes n} &= \sum_{P\in\mathcal{P}_n([r])} t_1^{nP(1)} \cdot \ldots \cdot t_r^{nP(r)} \sum_{\underline{k}\in[r]^n:\, P_{\underline{k}}=P} e_{k_1} \otimes \ldots \otimes e_{k_n} \\
&= \sum_{P\in\mathcal{P}_n([r])} t_1^{nP(1)} \cdot \ldots \cdot t_r^{nP(r)} \sqrt{n!}\, \underbrace{e_1 \vee \ldots \vee e_1}_{nP(1)\text{times}} \vee \ldots \underbrace{e_r \vee \ldots \vee e_r}_{nP(r)\text{times}} \\
&= \sum_{P\in\mathcal{P}_n([r])} t_1^{nP(1)} \cdot \ldots \cdot t_r^{nP(r)} \sqrt{n!(nP(1))! \cdot \ldots \cdot (nP(r))!}\, e_P^s.
\end{aligned}$$

Thus,

$$e_P^s = \sqrt{n!(nP(1))! \cdot \ldots \cdot (nP(r))!}^{-1} \frac{\partial}{\partial t_1^{nP(1)} \ldots \partial t_r^{nP(r)}} \psi(\underline{t})^{\otimes n} \Big|_{t_1=\ldots=t_r=0},$$

and hence $e_P^s \in \overline{\text{span}}\{\psi^{\otimes n} : \psi \in \mathcal{H}\}$, since derivatives are obtained by linear operations and taking limits. □

Remark 6.26. For $n = 2$, (6.173) can be proved in a straightforward manner, as

$$\begin{aligned}\text{span}\{\psi^{\otimes 2} : \psi \in \mathcal{H}\} \ni &(x+y) \otimes (x+y) - (x-y) \otimes (x-y) \\ &= 2(x \otimes y + y \otimes x) = 4P_s^{(2)}(x \otimes y) = \frac{4}{\sqrt{2}} x \vee y.\end{aligned}$$

There are various physical problems where the number of particles (fermions or bosons) may change during the time evolution of the system. To model such situations, we introduce the *full Fock space*, the *antisymmetric Fock space* and the *symmetric Fock space*, respectively, as

$$\mathcal{F}(\mathcal{H}) := \oplus_{n\in\mathbb{N}} \mathcal{H}^{\overline{\otimes}\, n}, \qquad \mathcal{F}_a(\mathcal{H}) := \oplus_{n\in\mathbb{N}} \wedge^n \mathcal{H}, \qquad \mathcal{F}_s(\mathcal{H}) := \oplus_{n\in\mathbb{N}} \vee^n \mathcal{H},$$

where $\mathcal{H}^{\otimes 0}$, $\wedge^0\mathcal{H}$ and $\vee^0\mathcal{H}$ are, by definition, the one-dimensional Hilbert space $\mathbb{C}$. The dimension of $\mathcal{F}_a(\mathcal{H})$ is $2^{\dim \mathcal{H}}$, which is finite when $\mathcal{H}$ is finite dimensional (it is a result of the fact that $\wedge^n\mathcal{H} = 0$ when $n > \dim \mathcal{H}$), while $\mathcal{F}(\mathcal{H})$ and $\mathcal{F}_s(\mathcal{H})$ are always infinite dimensional.

Both the antisymmetric and the symmetric Fock spaces are subspaces of the full Fock space, and, since their projectors are $\oplus_{n\in\mathbb{N}} P_a^{(n)}$ and $\oplus_{n\in\mathbb{N}} P_s^{(n)}$, respectively, and, since

$$\left(\oplus_{n\in\mathbb{N}} P_a^{(n)}\right)\left(\oplus_{n\in\mathbb{N}} P_s^{(n)}\right) = \oplus_{n\in\mathbb{N}} P_a^{(n)} P_s^{(n)} = 0,$$

the two subspaces are orthogonal to each other.

A basis vector given by the formula (6.168) or (6.169) is uniquely determined by the numbers $N_k := \#\{j : i_j = k\}$, so the set of basis vectors can be identified by sequences $\underline{N} \in (\mathbb{N}^I)_*$, where $_*$ refers to the fact that only finitely many terms of the sequence can be different from 0. Of course, N_k can only take the values 0 or 1 in the antisymmetric case. We introduce the notations

$$e_{\underline{N}}^a := \wedge_{k:N_k=1} e_k \qquad \text{and} \qquad e_{\underline{N}}^s := \frac{1}{\sqrt{\prod_{k\in I} N_k!}} \vee_{k:N_k>0} \left(\vee^{N_k} e_k\right).$$

Rewriting formulas (6.168) and (6.169) using these notations, we have that the sets

$$\{e^a_{\underline{N}} \ : \ \underline{N} \in (\mathbb{N}^I)_*, \sum_{k\in I} N_k = n\} \quad \text{and} \quad \{e^s_{\underline{N}} \ : \ \underline{N} \in (\mathbb{N}^I)_*, \sum_{k\in I} N_k = n\}$$

are orthonormal bases for $\wedge^n\mathcal{H}$ and $\vee^n\mathcal{H}$, respectively, and so the sets

$$\{e^a_{\underline{N}} \ : \ \underline{N} \in (\mathbb{N}^I)_*\} \qquad \text{and} \qquad \{e^s_{\underline{N}} \ : \ \underline{N} \in (\mathbb{N}^I)_*\}$$

are orthonormal bases for $\mathcal{F}_a(\mathcal{H})$ and $\mathcal{F}_s(\mathcal{H})$, respectively.

6.4 Operators

Respectively, an operator $A \in \mathcal{B}(\mathcal{H}^{\otimes n})$ is symmetric, if it commutes with all the operators of the representation, i.e.

$$U_\sigma A U^*_\sigma = A \qquad \forall \sigma \in S_n$$

and antisymmetric, if

$$U_\sigma A U^*_\sigma = \varepsilon(\sigma) A \qquad \forall \sigma \in S_n.$$

Again, symmetric and antisymmetric operators form subspaces $\vee^n\mathcal{B}(\mathcal{H})$ and $\wedge^n\mathcal{B}(\mathcal{H})$, that are closed with respect to the operator norm.

Exercise 6.27. Show that $U_\pi \ (A_1 \otimes \ldots \otimes A_n) \ U^*_\pi = A_{\pi(1)} \otimes \ldots \otimes A_{\pi(n)}$, and the operators $\mathcal{P}^{(n)}_s, \mathcal{P}^{(n)}_a : \mathcal{B}(\mathcal{H})^{\otimes n} \to \mathcal{B}(\mathcal{H})^{\otimes n}$, given by

$$\mathcal{P}_s(X)^{(n)} = \frac{1}{n!} \sum_{\pi\in S_n} U_\pi \, X \, U^*_\pi; \qquad \mathcal{P}_a(X)^{(n)} = \frac{1}{n!} \sum_{\pi\in S_n} \varepsilon(\sigma) \, U_\pi \, X \, U^*_\pi$$

are idempotents with corresponding ranges $\vee^n\mathcal{B}(\mathcal{H})$ and $\wedge^n\mathcal{B}(\mathcal{H})$. Restrict the above operators to $\mathcal{B}_2(\mathcal{H}^{\otimes n})$ (the Hilbert-Schmidt operators on $\mathcal{H}^{\otimes n}$), and show that in this case $\mathcal{P}^{(n)}_s$ and $\mathcal{P}^{(n)}_a$ are self-adjoint projections with orthogonal ranges.

Exercise 6.28. Show that symmetric operators leave the subspaces $\vee^n\mathcal{H}$ and $\wedge^n\mathcal{H}$ invariant, while antisymmetric operators interchange them.

Exercise 6.29. Show that both $P^{(n)}_s$ and $P^{(n)}_a$ are symmetric operators.

We will use the notations

$$A_1 \vee \ldots \vee A_n \ := \ \mathcal{P}^{(n)}_s A_1 \otimes \ldots \otimes A_n = \frac{1}{n!} \sum_{\sigma\in S_n} A_{\sigma(1)} \otimes \ldots \otimes A_{\sigma(n)}$$

$$A_1 \wedge \ldots \wedge A_n \ := \ \mathcal{P}^{(n)}_a A_1 \otimes \ldots \otimes A_n = \frac{1}{n!} \sum_{\sigma\in S_n} \varepsilon(\sigma) A_{\sigma(1)} \otimes \ldots \otimes A_{\sigma(n)}.$$

(Note that this notation is slightly incosistent, as there is a $\sqrt{n!}$ difference from the definitions of $x_1 \vee \ldots \vee x_n$ and $x_1 \wedge \ldots \wedge x_n$ given for vectors.) These operators can equivalently defined as the unique bounded linear extensions of

$$\begin{aligned} A_1 \vee \ldots \vee A_n \, x_1 \vee \ldots \vee x_n &:= (A_1 x_1) \vee \ldots \vee (A_n x_n) \\ A_1 \wedge \ldots \wedge A_n \, x_1 \wedge \ldots \wedge x_n &:= (A_1 x_1) \wedge \ldots \wedge (A_n x_n) \, . \end{aligned}$$

Exercise 6.30. Let $\mathcal{H}$ be a finite dimensional Hilbert space and $A \in \mathcal{B}(\mathcal{H})$. Let $E := \{e_1, \ldots, e_d\}$ be an orthonormal base w.r.t. which the matrix of A is upper triangular (see Exercise **??**). Show that the matrix of $A^{\otimes m}$, $\wedge^m A$ and $\vee^m A$ are also upper triangular in the bases canonically obtained from E, when the base elements are ordered lexikografically (i.e. $e_{\mathbf{i}} \succ e_{\mathbf{j}}$ iff $i_a > j_a$ for some $1 \leq a \leq m$ and $i_k = j_k$ for all $k < a$).

Solution: a) in the case of $A^{\otimes m}$ we have

$$\left\langle e_{i_1} \otimes \ldots \otimes e_{i_m}, A^{\otimes m} e_{j_1} \otimes \ldots \otimes e_{j_m} \right\rangle = \prod_{k=1}^{m} \langle e_{i_k}, A e_{j_l} \rangle$$

where $i_1, \ldots, i_m$ and $j_1, \ldots, j_m$ are arbitrary. This is obviously 0 if there exists a k such that $i_k > j_k$.

b) we have

$$\begin{aligned} \left\langle e_{i_1} \wedge \ldots \wedge e_{i_m}, A^{\otimes m} e_{j_1} \wedge \ldots \wedge e_{j_m} \right\rangle &= \det G(\mathbf{i}, \mathbf{j}) \\ \left\langle e_{i_1} \vee \ldots \vee e_{i_k}, A^{\otimes k} e_{j_1} \vee \ldots \vee e_{j_k} \right\rangle &= P_G(\mathbf{i}, \mathbf{j}) \, , \end{aligned}$$

where $G(\mathbf{i}, \mathbf{j})_{kl} := \langle e_{i_k}, A e_{j_l} \rangle$. $i_1 < \ldots < i_m$ and $j_1 < \ldots < j_m$ for $\wedge^m A$ and $i_1 \leq \ldots \leq i_m$ and $j_1 \leq \ldots \leq j_m$ for $\vee^m A$. It is easy to see that if $e_{\mathbf{i}} \succ e_{\mathbf{j}}$ then $G(\mathbf{i}, \mathbf{j})_{[a,d] \times [1,a]} = 0$ hence both the determinant and the permanent of $G(\mathbf{i}, \mathbf{j})$ is 0.

Exercise 6.31. Let $\{\lambda_1, \ldots, \lambda_d\}$ be the eigenvalues of $A \in \mathcal{B}(\mathcal{H})$, counted with multiplicity and arranged so that $|\lambda_1| \geq \ldots \geq |\lambda_d|$ (here $d = \dim \mathcal{H}$). Show that

(i)

$$\begin{aligned} \sigma(A^{\otimes k}) &= \{\lambda_{i_1} \cdot \ldots \cdot \lambda_{i_k} \, : \, 1 \leq i_1, \ldots, i_k \leq d\} \, , \\ \sigma(\vee^k A) &= \{\lambda_{i_1} \cdot \ldots \cdot \lambda_{i_k} \, : \, 1 \leq i_1 \leq \ldots \leq i_k \leq d\} \, , \\ \sigma(\wedge^k A) &= \{\lambda_{i_1} \cdot \ldots \cdot \lambda_{i_k} \, : \, 1 \leq i_1 < \ldots < i_k \leq d\} \, . \end{aligned}$$

(ii) Let $\mu_1(A) \geq \ldots \geq \mu_d(A)$ be the singular values of A. Show that

$$\begin{aligned} |\lambda_1(\wedge^k A)| &= \prod_{i=1}^{k} |\lambda_i(A)|\,, \\ \mu_1(\wedge^k A) &= \prod_{i=1}^{k} \mu_i(A)\,. \end{aligned}$$

Note that $\mu_1(A) = \|A\|$ and $\mu_1(\wedge^k A) = \left\|\wedge^k A\right\|$.

(iii) Show that

$$\prod_{i=1}^{k} |\lambda_i(A)| \leq \prod_{i=1}^{k} \mu_i(A)\,.$$

(iv) Show that if $A, B \in \mathcal{B}(\mathcal{H})$ then

$$\prod_{i=1}^{k} \mu_i(AB) \leq \prod_{i=1}^{k} \mu_i(A)\mu_i(B)\,.$$

If, moreover, AB is normal then

$$\prod_{i=1}^{k} \mu_i(AB) \leq \prod_{i=1}^{k} \mu_i(BA)\,.$$

Solution:

(i) follows from Exercise 6.30.

(ii) the first formula follows from the previous point and immediately yields the second, as $\mu_i(A) = \lambda_i(|A|)$.

(iii)

$$\prod_{i=1}^{k} |\lambda_i(A)| = |\lambda_1(\wedge^k A)| \leq \left\|\wedge^k A\right\| = \mu_1(\wedge^k A) = \prod_{i=1}^{k} \mu_i(A)\,.$$

(iv)

$$\begin{aligned} \prod_{i=1}^{k} \mu_i(AB) &= \mu_1(\wedge^k AB) = \left\|\wedge^k AB\right\| = \left\|(\wedge^k A)\,(\wedge^k B)\right\| \\ &\leq \left\|\wedge^k A\right\| \left\|\wedge^k B\right\| = \mu_1(\wedge^k A)\mu_1(\wedge^k B) = \prod_{i=1}^{k} \mu_i(A)\mu_i(B)\,. \end{aligned}$$

If AB is normal then $\wedge^k AB$ is normal as well, hence $|\lambda_1(\wedge^k AB)| = \big\|\wedge^k AB\big\| = \mu_1(\wedge^k AB)$. Since the spectral radius of a product does not depend on the order of the product,

$$\begin{aligned}\prod_{i=1}^{k} \mu_i(AB) &= \mu_1(\wedge^k AB) = |\lambda_1(\wedge^k AB)| = |\lambda_1\big((\wedge^k A)\,(\wedge^k B)\big)| \\ &= |\lambda_1\big((\wedge^k B \wedge^k A)\big)| = |\lambda_1(\wedge^k BA)| \leq \big\|\wedge^k BA\big\| \\ &= \mu_1(\wedge^k BA)\,.\end{aligned}$$

Exercise 6.32. Let $A \in \mathcal{B}(\mathcal{H})$ be compact with singular value decomposition $A = \sum_k \mu_k\, |f_k\rangle\langle e_k|$.

(i) Show that $A^{\otimes n}$ and $\wedge^n A$ are also compact, with singular value decompositions

$$\begin{aligned}A^{\otimes n} &= \sum_{k_1,\ldots,k_n} \mu_{k_1} \cdot \ldots \cdot \mu_{k_n}\, |f_{k_1} \otimes \cdot \ldots \otimes f_{k_n}\rangle\langle e_{k_1} \otimes \cdot \ldots \otimes e_{k_n}|\,, \\ \wedge^n A &= \sum_{k_1<\ldots<k_n} \mu_{k_1} \cdot \ldots \cdot \mu_{k_n}\, |f_{k_1} \wedge \cdot \ldots \wedge f_{k_n}\rangle\langle e_{k_1} \wedge \cdot \ldots \wedge e_{k_n}|\,.\end{aligned}$$

(ii) Conclude that if $A \in \mathcal{B}_p(\mathcal{H})$ then $A^{\otimes n} \in \mathcal{B}_p(\mathcal{H}^{\otimes n})$ and $\wedge^n A \in \mathcal{B}_p(\wedge^n \mathcal{H})$, and

$$\big\|A^{\otimes n}\big\|_p = \|A\|_p^n\,, \qquad \|\wedge^n A\|_p \leq \frac{1}{n!}\,\|A\|_p^n\,. \tag{6.174}$$

(iii) Show that if $\operatorname{rk} A = r$ then

$$\operatorname{rk} A^{\otimes n} = r^n, \qquad \operatorname{rk} \wedge^n A = \binom{r}{n}\,,$$

both of them being infinite when A is of infinite rank. In particular, if P is a finite-rank projection onto the subspace spanned by the orthonormal vectors $e_1, \ldots, e_r$ then

$$\wedge^r P = |e_1 \wedge \ldots \wedge e_r\rangle\langle e_1 \wedge \ldots \wedge e_r|\,.$$

Solution: The first statement is obvious for A with finite rank. One can easily see that

$$\|\wedge^n A - \wedge^n B\| \leq \big\|A^{\otimes n} - B^{\otimes n}\big\| \leq \|A - B\| \sum_{k=0}^{n-1} \|B\|^k\, \|A\|^{n-k}$$

for arbitrary bounded operators $A, B \in \mathcal{B}(\mathcal{H})$. Form this it follows that if a sequence A_m of finite-rank operators converge to A in norm then also $A_m^{\otimes n} \to A^{\otimes n}$ and $\wedge^n A_m \to \wedge^n A$, from which the first assertion follows. The statement about the p-norms is an immediate consequence.

Exercise 6.33. Let $A \in \mathcal{B}(\mathcal{H})$ and define the operator A_F by

$$\mathrm{dom}(A_F) := \left\{ x \in \mathcal{F}_a(\mathcal{H}) \,:\, \sum_n \left\| (\wedge^n A) P_a^{(n)} x \right\|^2 < \infty \right\}, \qquad A_F x := \sum_n \wedge^n A P_a^{(n)} x.$$

We also use the notation $A_F = \oplus_n \wedge^n A$.

Show that:

(i) A_F is bounded if $\|A\| \leq 1$. Moreover, if $\|A\|, \|B\| \leq 1$ then $\|A_F - B_F\| \leq \|A - B\|$, hence the map $A \mapsto A_F$ from $\mathcal{B}(\mathcal{H})$ to $\mathcal{B}(\mathcal{F}_a(\mathcal{H}))$ is continuous on the unit ball of $\mathcal{B}(\mathcal{H})$.

(ii) $A_F \in \mathcal{B}_p(\mathcal{F}_a(\mathcal{H}))$ if and only $A \in \mathcal{B}_p(\mathcal{H})$, and

$$\|A_F\|_p \leq e^{\|A\|_p}.$$

(iii) If A is of finite rank then

$$\mathrm{Tr}\, A_F = \det(I + A), \tag{6.175}$$

where $\det(I + A) := (1 + \lambda_1) \cdot \ldots \cdot (1 + \lambda_k)$, with $\lambda_1, \ldots, \lambda_k$ being the non-zero eigenvalues of A, counted with multiplicity.

Solution: The statements follow from the facts $\|\wedge^n A\| \leq \|A^{\otimes n}\|$ and (6.174).

By the above exercise, $\mathrm{Tr}\, A_F$ is finite for any trace-class operator A, and (6.175) motivates to define

$$\det(I + A) := \mathrm{Tr}\, A_F.$$

Exercise 6.34. Let $A \in \mathcal{B}(\mathcal{H})$ be a normal trace-class operator, and assume that -1 is not in its spectrum. Show that

$$\mathrm{Tr}\, A_F c(x_1)^* \ldots c(x_n)^* c(y_m) \ldots c(y_1) = \delta_{m,n} \det(I + A) \det \left\{ \left\langle y_{j_k}, \frac{A}{I + A} x_{i_l} \right\rangle \right\}_{k,l=1}^n.$$

Solution: First assume that A is of finite rank with eigen-decomposition $A = \sum_{k=1}^{r} \lambda_k |e_k\rangle\langle e_k|$. Choose a base of $\mathcal{H}$ consisting of $e_1, \ldots, e_r$ as the first r elements. First let $k_1 < \ldots < k_p,\ i_1 < \ldots < i_n,\ j_1 < \ldots < j_m$, and compute

$$\begin{aligned} T(k_1, \ldots, k_p) &:= \left\langle e_{k_1} \wedge \ldots \wedge e_{k_p}, A_F c(e_{i_1})^* \ldots c(e_{i_n})^* c(e_{j_m}) \ldots c(e_{j_1}) e_{k_1} \wedge \ldots e_{k_p} \right\rangle \\ &= \left\langle (Ae_{k_1}) \wedge \ldots \wedge (Ae_{k_p}), c(e_{i_1})^* \ldots c(e_{i_n})^* c(e_{j_m}) \ldots c(e_{j_1}) e_{k_1} \wedge \ldots e_{k_p} \right\rangle \\ &= \lambda_{k_1} \cdot \ldots \cdot \lambda_{k_p} \left\langle c(e_{i_1}) \ldots c(e_{i_n}) e_{k_1} \wedge \ldots \wedge e_{k_p}, c(e_{j_m}) \ldots c(e_{j_1}) e_{k_1} \wedge \ldots e_{k_p} \right\rangle . \end{aligned}$$

This last expression can be non-zero only if $m = n$ and $\{i_1, \ldots, i_n\} = \{j_1, \ldots, j_n\} \subset \{k_1, \ldots, k_p\}$. Thus

$$\begin{aligned} T &:= \operatorname{Tr} A_F c(e_{i_1})^* \ldots c(e_{i_n})^* c(e_{j_m}) \ldots c(e_{j_1}) \\ &= \delta_{m,n}\delta_{i_1,j_1} \cdot \ldots \cdot \delta_{i_n,j_n} \sum_{\{i_1,\ldots,i_n\} \subset \{k_1,\ldots,k_p\}} \lambda_{k_1} \cdot \ldots \cdot \lambda_{k_p} \\ &= \delta_{m,n}\delta_{i_1,j_1} \cdot \ldots \cdot \delta_{i_n,j_n} \lambda_{i_1} \cdot \ldots \cdot \lambda_{i_n} \sum_{X \subset \{1,\ldots,r\} \setminus \{i_1,\ldots,i_n\}} \lambda_X \\ &= \delta_{m,n}\delta_{i_1,j_1} \cdot \ldots \cdot \delta_{i_n,j_n} \lambda_{i_1} \cdot \ldots \cdot \lambda_{i_n} \prod_{t \in \{1,\ldots,r\} \setminus \{i_1,\ldots,i_n\}} (1 + \lambda_t) \\ &= \delta_{m,n}\delta_{i_1,j_1} \cdot \ldots \cdot \delta_{i_n,j_n} \frac{\lambda_{i_1}}{1+\lambda_{i_1}} \cdot \ldots \cdot \frac{\lambda_{i_n}}{1+\lambda_{i_n}} \det(I + A) \\ &= \delta_{m,n} \det(I + A) \det \left\{ \left\langle e_{j_k}, \frac{A}{I+A} e_{i_l} \right\rangle \right\}_{k,l=1}^{n} . \end{aligned}$$

Now for a general A with eigen-decomposition $A = \sum_k \lambda_k |e_k\rangle\langle e_k|$ we have $A_N := \sum_{k=1}^{N} \lambda_k |e_k\rangle\langle e_k| \to A$ as $N \to \infty$, and by continutity of the above terms we get the desired statement when $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are eigenvectors of A. The assertion for general vectors follows by multilinearity of both sides.

Exercise 6.35. Let $Q \in \mathcal{B}(\mathcal{H})$, and assume that 1 is not in the spectrum of Q. Show that

(i) Q is trace-class if and only if $\left(\frac{Q}{I-Q}\right)_F$ is trace-class;

(ii) $0 \le \left(\frac{Q}{I-Q}\right)_F$ if and only if $0 \le Q \le I$;

(iii) if Q is trace-class with $0 \le Q \le I$ then

$$\mathcal{D}\omega_Q := \det(I - Q) \left(\frac{Q}{I-Q}\right)_F$$

is a density operator, and

$$\operatorname{Tr} \mathcal{D}\omega_Q c(x_1)^* \ldots c(x_n)^* c(y_m) \ldots c(y_1) = \delta_{m,n} \det \{\langle y_{j_k}, Q x_{i_l} \rangle\}_{k,l=1}^{n} .$$

Solution: The first assertion can easily be seen by drawing the graphs of the functions $\frac{x}{1-x}$ and $\frac{x}{1+x}$. Obviously, $0 \leq \left(\frac{Q}{I-Q}\right)_F$ if and only if $0 \leq \frac{Q}{I-Q}$, which is equivalent to $0 \leq Q \leq I$ (see again the graphs). The third statement follows immediately from the previous exercise.

6.5 Second quantization basics

If $A \in \mathcal{B}(\mathcal{H})$ then $A^{\otimes m}$ leaves $\vee^m \mathcal{H}$ invariant, and we denote its restriction to $\vee^m \mathcal{H}$ by $\vee^m A$. The *Fock operator* A_F, corresponding to A, is

$$A_F := \bigoplus_{m=0}^{\infty} \vee^m A \quad \text{with} \quad \mathrm{dom}(A_F) := \left\{ \oplus_{m=0}^{\infty} x_m \in \mathcal{F}(\mathcal{H}) \, : \, \sum_{m=0}^{\infty} \left\| (\vee^m A)\, x_m \right\|^2 < \infty \right\}.$$

Note that the Fock operators are closed, and

$$\mathcal{F}_f(\mathcal{H}) := \left\{ \oplus_{m=0}^{M} x_m \, : \, x_m \in \vee^m \mathcal{H} \ \ \forall m, \ M \in \mathbb{N} \right\}$$

is a common core for all Fock operators, on which $A_F B_F = (AB)_F$ holds. If $A \geq 0$ then we also have $(A_F)^t = (A^t)_F$ on $\mathcal{F}_f(\mathcal{H})$ for any $t \in \mathbb{R}$, with the convention $0^t := 0$, $t \in \mathbb{R}$. Fock operators are also characterized by the property $A_F x_F = (Ax)_F$, $x \in \mathcal{H}$.

If $A \geq 0$ is a finite-rank operator and $A = \sum_{k=1}^{r} \lambda_k \, |e_k\rangle\langle e_k|$ is an eigen-decomposition of A, then

$$\vee^m A = \sum_{m_1, \ldots, m_r = m} \lambda_{\underline{m}} \, |e_{\underline{m}}\rangle\langle e_{\underline{m}}|$$

is an eigen-decomposition of $\vee^m A$, where

$$\lambda_{\underline{m}} := \lambda_1^{m_1} \cdot \ldots \cdot \lambda_r^{m_r}, \qquad e_{\underline{m}} := \frac{1}{\sqrt{m_1! \ldots m_r! m!}} \sum_{\sigma \in S_m} U_\sigma^{(m)} e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r},$$

and $U_\sigma^{(m)}$, $\sigma \in S_m$, denotes the standard unitary representation of the symmetric group S_m on $\mathcal{H}^{\otimes m}$. As a consequence,

$$\sum_{m=0}^{\infty} \mathrm{Tr} \vee^m A = \prod_{k=1}^{r} \left(\sum_{m=0}^{\infty} \lambda_k^m \right), \tag{6.176}$$

which is finite if and only if $A < I$, in which case A_F is trace-class with

$$\mathrm{Tr}\, A_F = \det (I - A)^{-1}. \tag{6.177}$$

If A is a bounded operator then $A \otimes I^{\otimes(m-1)} + I \otimes A \otimes I^{m-2} + \ldots + I^{\otimes(m-1)} \otimes A$ leaves $\vee^m \mathcal{H}$ invariant, and we denote its restriction to $\vee^m \mathcal{H}$ by $\Gamma_m(A)$. The second-quantized version of A is

$$\Gamma(A) := \oplus_{m=0}^{\infty} \Gamma_m(A).$$

Assume that A is of finite rank with an eigen-decomposition $A = \sum_{k=1}^{r} \lambda_k |e_k\rangle\langle e_k|$. Then, since $\Gamma_m(A) U_\sigma^{(m)} = U_\sigma^{(m)} \Gamma_m(A)$ for all $m \in \mathbb{N}$ and $\sigma \in S_m$,

$$\begin{aligned}
\Gamma(A) e_{\underline{m}} &= \Gamma_m(A) e_{\underline{m}} \\
&= \frac{1}{\sqrt{m_1! \ldots m_r! m!}} \sum_{\sigma \in S_m} \Gamma_m(A) U_\sigma^{(m)} e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r} \\
&= \frac{1}{\sqrt{m_1! \ldots m_r! m!}} \sum_{\sigma \in S_m} U_\sigma^{(m)} \Gamma_m(A) e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r} \\
&= \frac{1}{\sqrt{m_1! \ldots m_r! m!}} \sum_{\sigma \in S_m} U_\sigma^{(m)} \left(m_1\lambda_1 + \ldots + m_r\lambda_r\right) e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r} \\
&= \left(m_1\lambda_1 + \ldots + m_r\lambda_r\right) \frac{1}{\sqrt{m_1! \ldots m_r! m!}} \sum_{\sigma \in S_m} U_\sigma^{(m)} e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r} \\
&= \left(m_1\lambda_1 + \ldots + m_r\lambda_r\right) e_{\underline{m}},
\end{aligned}$$

and hence,

$$\Gamma(A) = \sum_{m=0}^{\infty} \sum_{m_1, \ldots, m_r = m} \left(m_1\lambda_1 + \ldots + m_r\lambda_r\right) |e_{\underline{m}}\rangle\langle e_{\underline{m}}| \ .$$

Since

$$\left(\log A_F\right) e_{\underline{m}} = \left(\log \lambda_{\underline{m}}\right) e_{\underline{m}} = \left(m_1 \log \lambda_1 + \ldots + m_r \log \lambda_r\right) e_{\underline{m}},$$

we get

$$\log A_F = \Gamma(\log A).$$

Now let A and B be of finite rank with $A = \sum_{k=1}^{r} \lambda_k |e_k\rangle\langle e_k| < I$. Then,

$$\operatorname{Tr} A_F \Gamma(B) = \sum_{m=0}^{\infty} \sum_{m_1, \ldots, m_r = m} \lambda_{\underline{m}} \langle e_{\underline{m}}, \Gamma_m(B) e_{\underline{m}} \rangle \, .$$

Since $\Gamma_m(B)$ is permutation-invariant, we get

$$\begin{aligned}
\langle e_{\underline{m}}, \Gamma_m(B) e_{\underline{m}} \rangle &= \frac{1}{m_1! \ldots m_r! m!} \sum_{\sigma, \pi \in S_m} \left\langle e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r}, U_{\sigma^{-1}}^{(m)} \Gamma_m(B) U_\pi^{(m)} e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r} \right\rangle \\
&= \frac{1}{m_1! \ldots m_r!} \sum_{\tau \in S_m} \left\langle e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r}, U_\tau^{(m)} \Gamma_m(B) e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r} \right\rangle .
\end{aligned}$$

Orthogonality of the e_k's gives

$$\left\langle e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r}, U_\tau^{(m)} \Gamma_m(B) e_1^{\otimes m_1} \otimes \ldots \otimes e_r^{\otimes m_r} \right\rangle = m_1! \cdot \ldots \cdot m_r! \sum_{k=1}^{r} m_k \left\langle e_k, B e_k \right\rangle,$$

and therefore

$$\begin{aligned}
\operatorname{Tr} A_F \Gamma(B) &= \sum_{k=1}^{r} \langle e_k, B e_k \rangle \sum_{m=0}^{\infty} \sum_{m_1,\ldots,m_r=m} \lambda_{\underline{m}} m_k \\
&= \frac{1}{\det(I-A)} \sum_{k=1}^{r} \langle e_k, B e_k \rangle (1-\lambda_k) \sum_{n=0}^{+\infty} n \lambda_k^n \\
&= \frac{1}{\det(I-A)} \sum_{k=1}^{r} \langle e_k, B e_k \rangle \frac{\lambda_k}{1-\lambda_k} \\
&= \frac{1}{\det(I-A)} \operatorname{Tr} \frac{A}{I-A} B.
\end{aligned}$$

7 Fermionic systems

7.1 The CAR algebra

The observable algebra of an indeterminate number of Fermions with one-particle Hilbert space $\mathcal{H}$ is called the CAR (**c**anonical **a**nticommutation **r**elations) algebra on $\mathcal{H}$; in notation $CAR(\mathcal{H})$. Mathematically it can be described as the universal C^*-algebra given by the set of generators $\{c(x) : x \in \mathcal{H}\}$ satisfying relations

$$\text{(R1")} \quad \text{The map } x \mapsto c(x) \text{ is complex anti-linear;} \tag{7.178}$$

$$\text{(R2")} \quad \{c(x), c(y)\} = 0; \tag{7.179}$$

$$\text{(R3")} \quad \{c^*(x), c(y)\} = \langle y, x\rangle\, I; \tag{7.180}$$

where $\{\,.\,,\,.\,\}$ denotes the anti-commutator.

The map $x \mapsto c(x)$ is easily seen to be norm-preserving, which, by (anti-) linearity implies that $CAR(\mathcal{H})$ can equally be specified as the universal C^*-algebra with generators $\{c_n : n = 1, \ldots, \dim \mathcal{H}\}$ satisfying

$$\{c_n, c_m\} = 0 \qquad \text{and} \qquad \{c_n^*, c_m\} = \delta_{nm} I\,. \tag{7.181}$$

The relation between the two pictures is that c_n equals to $c(e_n)$, where $\{e_n : n = 1, \ldots, \dim \mathcal{H}\}$ is an orthonormal base in $\mathcal{H}$.

A possible realization of relations (7.181) can be obtained by defining $\hat{c}_n := \left(\prod_{j=0}^{n-1} \sigma_z^j\right) \sigma_+^n$ in the spin-$\frac{1}{2}$ quantum chain, with $\sigma_+^n := \frac{1}{2}\left(\sigma_x^n + i\sigma_y^n\right)$. The set $\{\hat{c}_n : n \in \mathbb{N}\}$ is easily seen to generate the half-infinite spin chain, thus $\mathcal{C}_{\mathbb{N}}$ is a homomorphic image of $CAR(\mathcal{H})$. On the other hand, the elements

$$\begin{aligned} E_{11}^n &:= c_n c_n^* & \qquad E_{22}^n &:= c_n^* c_n \\ E_{12}^n &:= \left(\prod_{j=0}^{n-1} \left(1 - 2c_j^* c_j\right)\right) c_k & \qquad E_{21}^n &:= \left(\prod_{j=0}^{n-1} \left(1 - 2c_j^* c_j\right)\right) c_k^* \end{aligned}$$

in $CAR(\mathcal{H})$ satisfy relations (R1'), (R2') and (R3') of commuting matrix units, moreover, they generate $CAR(\mathcal{H})$, therefore $CAR(\mathcal{H})$ is isomorphic to $\mathcal{C}_{\mathbb{N}}$. An explicit isomorphism is obtained by extending the map

$$\mathbb{J} : E_{ij}^n \mapsto e_{ij}^n,$$

where $\{e_{ij}^n\}$ is an arbitrary set of commuting matrix units in $\mathcal{C}_{\mathbb{N}}$. This isomorphism is called the *Jordan-Wigner isomorphism*.

A particular representation of the CAR algebra is the *Fock representation*. The Hilbert space carrying it is the antisymmetric Fock space

$$\mathcal{F}(\mathcal{H}) := \bigoplus_{n\in\mathbb{N}} \wedge^n \mathcal{H}\,, \tag{7.182}$$

where $\wedge^n\mathcal{H}$ is the n-th antisymmetric tensor power of $\mathcal{H}$, spanned by vectors of the form

$$x_1 \wedge \ldots \wedge x_n := \frac{1}{\sqrt{n!}} \sum_{\sigma\in S_n} \varepsilon(\sigma) x_{\sigma(1)} \otimes \ldots \otimes x_{\sigma(n)}\,.$$

Here $\varepsilon(\sigma)$ is the sign of the permutation σ. Representants of $c(y)^*$ are the creation operators, defined by

$$a^*(y)\,(x_1 \wedge \ldots \wedge x_n) := y \wedge x_1 \wedge \ldots \wedge x_n\,.$$

The adjoints $a(y)$ are called the annihilation operators. The number operator N is the unbounded operator that equals to $n\,I_{\wedge^n\mathcal{H}}$ when restricted to $\wedge^n\mathcal{H}$. It can also be expressed in the form $N = \sum_k a_k^* a_k$, where $a_k = a(e_k)$ for some orthonormal base $\{e_k\}$ in $\mathcal{H}$.

7.2 Quasi-free morphisms

An isometry $V : \mathcal{H} \to \mathcal{K}$ between the Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ defines a map

$$\gamma_V\,(c(x)) := c\,(Vx)$$

from $CAR(\mathcal{H})$ to $CAR(\mathcal{K})$. It preserves the CAR relations, i.e.

$$\{\gamma_V\,(c^*(x))\,, \gamma_V\,(c(y))\} = \langle y, x\rangle\, I \tag{7.183}$$

$$\{\gamma_V\,(c(x))\,, \gamma_V\,(c(y))\} = 0 \tag{7.184}$$

holds. As a consequence, γ_V extends to a homomorphism from $CAR(\mathcal{H})$ to $CAR(\mathcal{K})$, with range $CAR\,(V\,(\mathcal{H}))$. Such a homomorphism is called a *quasi-free homomorphism*. In particular, a unitary U defines a quasi-free automorphism of $CAR(\mathcal{H})$.

Example 7.1. The *parity automorphism* α is the quasi-free automorphism generated by the unitary $-I$, i.e.

$$\alpha(c(x)) = c(-x).$$

The fixed point algebra of α is

$$CAR(\mathcal{H})_+ := \{b \in CAR(\mathcal{H}) \ : \ \alpha(b) = b\};$$

its elements are called *even*, while elements with the property $\alpha(b) = -b$ are called *odd*. $CAR(\mathcal{H})_+$ is generated by elements of the form

$$c^*_{i_1} \dots c^*_{i_n} c_{j_m} \dots c_{j_1}; \qquad n+m \text{ is even },$$

while the subset of the odd elements is the closure of the linear span of elements of the form

$$c^*_{i_1} \dots c^*_{i_n} c_{j_m} \dots c_{j_1}; \qquad n+m \text{ is odd }.$$

Any element $b \in CAR(\mathcal{H})$ can uniquely be decomposed into a sum of an even and an odd element; the decomposition is of the form

$$b = b_+ + b_- ; \qquad b_+ = \frac{1}{2}\left(b + \alpha(b)\right), \qquad b_- = \frac{1}{2}\left(b - \alpha(b)\right).$$

A functional φ on $CAR(\mathcal{H})$ is called *even*, if $\varphi \circ \alpha = \varphi$, or equivalently, if φ vanishes on odd elements.

Example 7.2. The *gauge group* of $CAR(\mathcal{H})$ is the group of automorphisms $(\kappa_w)_{w\in\mathbb{T}}$, where κ_w is the quasi-free automorphism given by the unitary wI on the one-particle space, and $\mathbb{T}$ is the complex unit circle. The subalgebra

$$\{b \ : \ \kappa_w(b) = b \ \ \forall w \in \mathbb{T}\}$$

is called the gauge-invariant part of $CAR(\mathcal{H})$, and is generated by the monomials

$$c^*_{i_1} \dots\dots c^*_{i_n} c_{j_m} \dots c_{j_1}; \qquad n = m.$$

A state φ is called gauge-invariant, if $\varphi \circ \kappa_w = \varphi$ holds for all $w \in \mathbb{T}$.

Note that $\kappa_{-1} = \alpha$ is the parity automorphism.

Example 7.3. The shift operator

$$S\,\delta_k := \delta_{k+1}$$

on $\mathcal{H} := l^2(\mathbb{Z})$, where $\{\delta_k \ : \ k \in \mathbb{Z}\}$ is the standard base in $\mathcal{H}$, is a unitary operator; the quasi-free automorphism generated by it is called the *shift automorphism*, and is denoted by γ. The shift automorphism is characterized by the property

$$\gamma(c_k) = c_{k+1}$$

with $c_k = c(\delta_k)$.

A state ω on $CAR\left(l^2\left(\mathbb{Z}\right)\right)$ is called *shift-invariant*, if

$$\omega = \omega \circ \gamma$$

holds.

Example 7.4. In the Fock representation every quasi-free automorphism β with one-particle unitary U is of the form $\beta = \mathrm{Ad}_{U_\mathrm{F}}$, where $U_\mathrm{F} := \bigoplus_n \wedge^n U$ is a unitary on the Fock space. The unitary $\wedge^n U$ on the n-particle space is defined by $(\wedge^n U)\ (x_1 \wedge \ldots \wedge x_n) := Ux_1 \wedge \ldots \wedge Ux_n$.

Example 7.5. The unitary evolution of an interaction-free Fermionic system is the typical physical example of a quasi-free evolution. The second-quantized Hamiltonian $\hat{H}$ is the direct sum of the restrictions of $H_n := H \otimes I \otimes \ldots \otimes I + I \otimes H \otimes I \otimes \ldots \otimes I + I \otimes \ldots \otimes I \otimes H$ onto the antisymmetrized n-particle spaces, where H is the one-particle Hamiltonian. The corresponding unitary group is $U(t) := e^{-it\hat{H}}$, and the evolution of the annihilation operators in the Fock representation is

$$c(x)(t) = e^{it\hat{H}}\, c(x)\, e^{-it\hat{H}} = c\left(e^{itH}x\right) \tag{7.185}$$

7.3 Quasi-free states

Suppose an interaction-free fermionic system is described by a Hamilton operator H with discrete spectrum, such that $e^{-\beta\hat{H}}$ is a trace-class operator for all positive β, where $\hat{H}$ is the second-quantized Hamiltonian. Then the Gibbs state ϱ_β of the system in the Fock representation has density operator $\frac{e^{-\beta\hat{H}}}{\mathrm{Tr}\ e^{-\beta\hat{H}}}$, and its value on monomials can be expressed as

$$\varrho_\beta\ (c(x_1)^* \ldots c(x_n)^* c(y_m) \ldots c(y_1)) = \delta_{m,n} \det\{\langle y_i, Q\ x_j\rangle\,, \tag{7.186}$$

where Q has the same eigenvectors as H with corresponding eigenvalues $\frac{e^{-\beta\, h_i}}{1+e^{-\beta\, h_i}}$, where the h_i's are the eigenvalues of H.

In general, given an operator Q on the one-particle Hilbert space, a functional defined on monomials by the right-hand side of (7.186) extends to a state of the CAR algebra if and only if $0 \le Q \le I$; the resulting state is called a *quasi-free state* with *symbol* Q. A quasi-free state is pure if and only if its symbol Q is a projection. The two extremes are the *Fock state* and the *anti-Fock state*, with symbol 0 and I; these two states are completely determined by

$$\omega_\mathrm{F}(c(x)^* c(x)) = 0 \qquad \text{and} \qquad \omega_\mathrm{aF}(c(x)c(x)^*) = 0\,. \tag{7.187}$$

The Fock representation is the GNS representation of the Fock state; the representing vector is the vacuum vector $\Omega \in \wedge^0\mathcal{H}$.

For a subspace $\mathcal{K} \subset \mathcal{H}$ let $\mathcal{A}_\mathcal{K}$ be the subalgebra of $CAR(\mathcal{H})$, generated by $\{c(x)\ :\ x \in \mathcal{K}\}$. The restriction of a quasi-free state to $\mathcal{A}_\mathcal{K}$ is easily seen to be a quasi-free state itself, with symbol PQP, where P is the projection onto $\mathcal{K}$.

For a finite dimensional Hilbert space $\mathcal{H}$ the symbol Q yields a base of the space, consisting of its eigenvectors. The corresponding Jordan-Wigner isomorphism sends the quasi-free state into a product state on the spin chain, with density

$$D = \bigotimes_{k=1}^{\dim \mathcal{H}} \begin{bmatrix} q_k & 0 \\ 0 & 1 - q_k \end{bmatrix} \tag{7.188}$$

where the q_k are the eigenvalues of Q.

A quasi-free state ω_Q on $CAR\left(l^2\left(\mathbb{Z}\right)\right)$ is translation-invariant if and only if its symbol is, i.e. it commutes with the shift operator of $l^2\left(\mathbb{Z}\right)$. Shift-invariant operators are also called Toeplitz operators. A shift-invariant symbol is uniquely determined by a sequence $q : \mathbb{Z} \to \mathbb{C}$, such that the matrix of Q in the standard base of $l^2\left(\mathbb{Z}\right)$ is of the form $Q_{k,l} = q(k-l)$. Shift-invariance implies that the Fourier transformation maps the symbol of a shift-invariant quasi-free state to a multiplication operator $M_{\hat{q}}$ on $\mathcal{L}^2\left(\mathbb{T}\right)$, where $\mathbb{T}$ is the one dimensional torus and $\hat{q}$ is a real-valued measurable function on $\mathbb{T}$. By identifying $\mathbb{T}$ with the interval $[0, 2\pi)$, the function $\hat{q}$ can be viewed as a function on $[0, 2\pi)$, satisfying $0 \leq \hat{q}(\vartheta) \leq 1$ for almost all $\vartheta \in [0, 2\pi)$. The defining sequence q is then the sequence of Fourier coefficients of $\hat{q}$:

$$q(k) = \frac{1}{2\pi} \int_0^{2\pi} e^{-ik\vartheta}\, \hat{q}(\vartheta)\, d\vartheta.$$

The state is pure if and only if $\hat{q}$ is the characteristic function χ_K of a subset $K \subset [0, 2\pi]$.

7.4 Quasi-free states on the spin chain

Let $\mathcal{A} := CAR\left(l^2(\mathbb{Z})\right)$ and $\mathcal{C}$ be the two-sided quantum spin-$\frac{1}{2}$ chain with corresponding shift automorphisms $\gamma_{\mathcal{A}}$ and $\gamma_{\mathcal{C}}$. Our aim in this section is to give translation-invariant states on the two-sided spin chain, by carrying shift-invariant states on $\mathcal{A}$ to $\mathcal{C}$. Since these algebras are isomorphic to each other, we can carry any state φ on $\mathcal{A}$ to a state $\varphi \circ \tau^{-1}$ on $\mathcal{C}$, where τ is any isomorphism from $\mathcal{A}$ onto $\mathcal{C}$. However, the image of a shift-invariant state need not be shift-invariant on $\mathcal{C}$, unless τ is compatible with the shifts of the two algebras, i.e.

$$\gamma_{\mathcal{C}} \circ \tau = \tau \circ \gamma_{\mathcal{A}}$$

holds. Note that the Jordan-Wigner isomorphism gives an isomorphism between $\mathcal{A}$ and the one-sided chain $\mathcal{C}_{\mathbb{N}}$, but the procedure described there cannot be modified to obtain an isomorphism onto $\mathcal{C}$. Moreover, the Jordan-Wigner isomorphism is not compatible with the shift of $\mathcal{A}$ and the shift of the one-sided chain.

Not having an isomorphism at hand that is compatible with the shifts, we use a more involved method, developed by Araki in [**?**], to obtain shift-invariant states on $\mathcal{C}$ from shift-invariant states on $\mathcal{A}$. The basic idea is to embed the algebra $\mathcal{A}$ into a bigger algebra $\tilde{A}$, extend the shift $\gamma_{\mathcal{A}}$ to an automorphism $\tilde{\gamma}$ of $\tilde{A}$ and the shift-invariant state φ to a $\tilde{\gamma}$-invariant state $\tilde{\varphi}$ on $\tilde{A}$, and then to find a $\tilde{\gamma}$-invariant subalgebra $\hat{\mathcal{A}}$ of $\tilde{\mathcal{A}}$, restrict $\tilde{\gamma}$ and $\tilde{\varphi}$ to $\hat{\mathcal{A}}$, and, as a last step, to find an explicit isomorphism between $\hat{\mathcal{A}}$ and $\mathcal{C}$ that is compatible with the restriction $\hat{\gamma}$ and $\gamma_{\mathcal{C}}$.

The algebra

We work in the Fock representation of $CAR\left(l^2\left(\mathbb{Z}\right)\right)$, and identify $\mathcal{A}$ with the algebra generated by the Fock space annihilation operators $\{a(x) \ : \ x \in l^2\left(\mathbb{Z}\right)\}$. We introduce the unbounded self-adjoint operator

$$N_- := \sum_{k<0} a_k^* a_k,$$

and, by means of functional calculus, the bounded operator

$$T := (-1)^{N_-},$$

which is a self-adjoint unitary, i.e.

$$T^* = T; \qquad T^2 = I. \tag{7.189}$$

Evaluating the product Ta_kT on vectors of the form $x_1 \wedge \ldots \wedge x_n$, and taking into account that the linear span of vectors of this form is dense in the Fock space, we get

$$Ta_kT = \begin{cases} a_k & \text{if } k \geq 0; \\ -a_k & \text{if } k < 0. \end{cases} \tag{7.190}$$

In other words, the map $b \mapsto TbT$ is the quasi-free automorphism generated by the unitary

$$U = \sum_{k\geq 0} |\delta_k\rangle\langle\delta_k| - \sum_{k<0} |\delta_k\rangle\langle\delta_k| \,. \tag{7.191}$$

Now we define $\tilde{\mathcal{A}}$ to be the C^*-algebra generated by $\mathcal{A}$ and T. It is shown in Appendix C.2 that T is not an element of $\mathcal{A}$, therefore $\tilde{\mathcal{A}}$ is strictly bigger than $\mathcal{A}$. Clearly, every element of $\tilde{\mathcal{A}}$ can uniquely be written in the form $a+Tb$ with $a, b \in \mathcal{A}$, that is,

$$\tilde{\mathcal{A}} = \mathcal{A} + T\mathcal{A}.$$

A straightforward computation verifies that the following maps are automorphisms of $\tilde{\mathcal{A}}$:

$$\begin{aligned}\beta\,(a+Tb) &:= a-Tb;\\ \tilde{\gamma}\,(a+Tb) &:= \gamma(a)+T\hat{\sigma}_z^0\gamma(b),\\ \tilde{\kappa}_w\,(a+Tb) &:= \kappa_w(a)+T\kappa_w(b),\end{aligned}$$

with $\hat{\sigma}_z^0 := I-2a_0^*a_0$, and that $\tilde{\gamma}$ is an extension of the shift automorphism $\gamma_{\mathcal{A}}$, while $\tilde{\kappa}_w$ extends the gauge group to $\tilde{\mathcal{A}}$. We use the notation $\tilde{\alpha}$ for the extension of the parity automorphism $\alpha=\kappa_{-1}$.

The fixed point set of the automorphism $\beta^{-1}\tilde{\alpha}$ forms a C^*-algebra, which we denote by $\hat{\mathcal{A}}$, i.e.

$$\begin{aligned}\hat{\mathcal{A}} &= \{a+Tb\in\tilde{\mathcal{A}}\,:\,\tilde{\alpha}(a+Tb)=\beta(a+Tb)\}\\ &= \{a+Tb\in\tilde{\mathcal{A}}\,:\,\alpha(a)=a,\,\alpha(b)=-b\}\\ &= \mathcal{A}_+ + T\mathcal{A}_-,\end{aligned}$$

where $\mathcal{A}_+$ and $\mathcal{A}_-$ are the even and the odd parts of $\mathcal{A}$.

For an odd element $b\in\mathcal{A}$ the element $\sigma_z^0 b$ is odd again, and since $\gamma\,(\mathcal{A}_+)=\mathcal{A}_+$ and $\gamma\,(\mathcal{A}_-)=\mathcal{A}_-$, we have

$$\tilde{\gamma}\left(\hat{\mathcal{A}}\right)=\hat{\mathcal{A}},$$

and so $\tilde{\gamma}$ can be restricted to an automorphism of $\hat{\mathcal{A}}$, which we denote by $\hat{\gamma}$, and call the shift automorphism of the algebra $\hat{\mathcal{A}}$. A similar argument shows that $\tilde{\kappa}_w$ can also be restricted to $\hat{\mathcal{A}}$; the restriction is again denoted by $\hat{\kappa}_w$.

Using the notation $\hat{\sigma}_z^k := I-2a_k^*a_k$ we define the elements

$$A_k := \begin{cases}\prod_{\ell=0}^{k-1}\hat{\sigma}_z^l\,, & \text{if } k>0;\\ I\,, & \text{if } k=0;\\ \prod_{\ell=k}^{-1}\hat{\sigma}_z^l\,, & \text{if } k<0.\end{cases}$$

and

$$\hat{E}_{11}^k := a_ka_k^*,\quad \hat{E}_{22}^k := a_k^*a_k,\quad \hat{E}_{12}^k := TA_ka_k,\quad \hat{E}_{21}^k := TA_ka_k^*. \tag{7.192}$$

The elements $\hat{E}_{ij}^k$ are easily seen to be in the subalgebra $\hat{\mathcal{A}}$, and a direct computation shows that they satisfy the relations (R1'), (R2') and (R3') of commuting matrix units, moreover,

$$\hat{\gamma}\left(\hat{E}_{ij}^k\right)=\hat{E}_{ij}^{k+1}\,.$$

This implies that the map

$$\Pi : \hat{E}^k_{ij} \mapsto e^k_{ij}$$

extends to an isomorphism between the algebras $\hat{\mathcal{A}}$ and $\mathcal{C}$, and that

$$\gamma_{\mathcal{C}} \circ \Pi = \Pi \circ \hat{\gamma}, \tag{7.193}$$

i.e. Π intertwines the shifts of the two algebras.

With these notations we have

$$\Pi^{-1}\left(\sigma^k_z\right) = \hat{\sigma}^k_z,$$

and

$$\hat{\sigma}^k_x := \Pi^{-1}\left(\sigma^k_x\right) = \hat{E}^k_{12} + \hat{E}^k_{21}; \qquad \hat{\sigma}^k_y := \Pi^{-1}\left(\sigma^k_y\right) = -i\hat{E}^k_{12} + i\hat{E}^k_{21}\,. \tag{7.194}$$

Since

$$\hat{\kappa}_w\left(\hat{E}^k_{11}\right) = \hat{E}^k_{11}; \quad \hat{\kappa}_w\left(\hat{E}^k_{22}\right) = \hat{E}^k_{22}; \quad \hat{\kappa}_w\left(\hat{E}^k_{12}\right) = \overline{w}\hat{E}^k_{12}; \quad \hat{\kappa}_w\left(\hat{E}^k_{21}\right) = w\hat{E}^k_{21};$$

then we have

$$\hat{\kappa}_w\left(\hat{\sigma}^k_x\right) = \cos\vartheta\ \hat{\sigma}^k_x + \sin\vartheta\ \hat{\sigma}^k_y; \quad \hat{\kappa}_w\left(\hat{\sigma}^k_y\right) = -\sin\vartheta\ \hat{\sigma}^k_x + \cos\vartheta\ \hat{\sigma}^k_y; \quad \hat{\kappa}_w\left(\hat{\sigma}^k_z\right) = \hat{\sigma}^k_z;$$

with $w = e^{i\vartheta}$, that is,

$$\Pi \circ \hat{\kappa}_w = \kappa^{\mathcal{C}}_w \circ \Pi, \tag{7.195}$$

where now $\kappa^{\mathcal{C}}_w$ denotes the rotation group of the spin chain. In particular, the even part of the spin chain is identified with the subalgebra

$$\hat{\mathcal{A}}_+ := \{a + Tb \in \hat{\mathcal{A}}\ :\ \hat{\alpha}\,(a + Tb) = a + Tb\} = \mathcal{A}_+,$$

while the odd part is mapped by Π^{-1} to

$$\hat{\mathcal{A}}_- := \{a + Tb \in \hat{\mathcal{A}}\ :\ \hat{\alpha}\,(a + Tb) = -\,(a + Tb)\} = T\mathcal{A}_-.$$

Since the decomposition of an element into the sum of an even and an odd element is unique both in the CAR algebra and the spin chain, the map $W : \mathcal{A} \to \hat{\mathcal{A}}$;

$$Wb := b_+ + Tb_-$$

is well-defined, and gives a linear isomorphism between the two algebras, with the locality property

$$W\left(\mathcal{A}_\Lambda\right) = \hat{\mathcal{A}}_\Lambda \qquad \text{for an interval}\ \ \Lambda = [a, b];\quad a \le 0 \le b,$$

where $\hat{\mathcal{A}}_\Lambda$ is the algebra generated by $\{\hat{E}^k_{ij} \ : \ i,j \in \{1,2\};\ k \in \Lambda\}$. Note, however, that $W(\mathcal{A}_\Lambda) = \hat{\mathcal{A}}_\Lambda$ doesn't hold for a general Λ. W is also compatible with the translations, and the gauge (rotation) groups of the algebras:

$$\hat{\gamma} \circ W = W \circ \gamma; \qquad \hat{\kappa}_w \circ W = W \circ \kappa_w.$$

The restriction onto the intersection $\mathcal{A} \cap \hat{\mathcal{A}} = \mathcal{A}_+$ is simply the identity of $\mathcal{A}_+$.

The algebra $\mathcal{A}_{[0,N-1]}$ is also generated by a set of matrix units $\left\{E^{(N)}_{\boldsymbol{ij}} := \prod_{k=0}^{N-1} E^k_{ij} \ : \ \boldsymbol{i},\boldsymbol{j} \in \{1,2\}^N\right\}$, where

$$E^k_{11} := a_k a_k^*, \quad E^k_{22} := a_k^* a_k, \quad E^k_{12} := A_k a_k, \quad E^k_{21} := A_k a_k^*.$$

Such a matrix unit is even (odd) if and only if $\sum_{k=0}^{n-1}(i_k - j_k)$ is even (odd). We can define matrix units the same way in $\hat{\mathcal{A}}_{[0,N-1]}$, i.e. $\hat{E}^{(N)}_{\boldsymbol{ij}} := \prod_{k=0}^{N-1} \hat{E}^k_{ij}$. Relations (7.189) and (7.190) imply that

$$\hat{E}^{(N)}_{\boldsymbol{ij}} = \begin{cases} E^{(N)}_{\boldsymbol{ij}}, & \text{when } \sum_{k=0}^{n-1}(i_k - j_k) \text{ is even;} \\ TE^{(N)}_{\boldsymbol{ij}}, & \text{when } \sum_{k=0}^{n-1}(i_k - j_k) \text{ is odd;} \end{cases}$$

therefore we have

$$W\left(E^{(N)}_{\boldsymbol{ij}}\right) = \hat{E}^{(N)}_{\boldsymbol{ij}}.$$

States

If φ_0 is a state on $\mathcal{A}_+$, then

$$\varphi(b) := \varphi_0(b_+)$$

defines a linear functional on $\mathcal{A}$, which is the trivial extension of φ_0. Since

$$|\varphi(b)| = \left|\varphi_0\left(\frac{1}{2}\left(b + \alpha(b)\right)\right)\right| \leq \left\|\frac{1}{2}\left(b + \alpha(b)\right)\right\| \leq \frac{1}{2}\left(\|b\| + \|\alpha(b)\|\right) = \|b\|$$

then $\|\varphi\| = 1 = \varphi(I)$, and so, by lemma (**??**), the extension is a state, which is obviously even. On the other hand, any even state on $\mathcal{A}$ can be recovered by this method from its restriction to $\mathcal{A}_+$, so we have a one-to-one correspondence between states on $\mathcal{A}_+$ and even states of $\mathcal{A}$. Repeating the same argument for $\hat{\mathcal{A}}$ we get a one-to one correspondence between states of $\mathcal{A}_+$ and even states of $\hat{\mathcal{A}}$. We denote

the even extension of φ_0 to $\hat{\mathcal{A}}$ by $\hat{\varphi}$. Note that $\hat{\varphi}$ is the restriction to $\hat{\mathcal{A}}$ of the state $\tilde{\varphi}$ on $\tilde{\mathcal{A}}$, defined by

$$\tilde{\varphi}\left(a+Tb\right) := \varphi(a).$$

The positivity of $\tilde{\varphi}$ follows from

$$\tilde{\varphi}\left(\left(a+Tb\right)^*\left(a+Tb\right)\right) = \varphi(a^*a+b^*b) \geq 0.$$

We can also express the one-to-one correspondence between even states φ of $\mathcal{A}$ and even states of $\hat{\mathcal{A}}$ by the help of the operator W from the previous section:

$$\varphi = \hat{\varphi}\circ W.$$

The states $\varphi, \hat{\varphi}$ and φ_0 not only determine each other uniquely, but they are essentially the same in the sense that both φ and $\hat{\varphi}$ are trivial extensions of φ_0. Note that the same statements hold when we consider the local algebras $\mathcal{A}_N$, $\hat{\mathcal{A}}_N$ and $(\mathcal{A}_N)_+$.

Since W intertwines the shifts of $\mathcal{A}$ and $\hat{\mathcal{A}}$, and $\gamma(\mathcal{A}_+) = \mathcal{A}_+$, then the shift-invariance of any of the above three states implies the shift-invariance of the other two.

The density matrices of the restrictions $\varphi_N := \varphi \restriction_{\mathcal{A}_{[0,N-1]}}$ and $\hat{\varphi}_N := \hat{\varphi} \restriction_{\hat{\mathcal{A}}_{[0,N-1]}}$ are the same:

$$[\varphi_N]_{\boldsymbol{i},\boldsymbol{j}} = \varphi\left(E^{(N)}_{\boldsymbol{ji}}\right) = \hat{\varphi}\circ W\left(E^{(N)}_{\boldsymbol{ji}}\right) = \hat{\varphi}\left(\hat{E}^{(N)}_{\boldsymbol{ji}}\right) = [\hat{\varphi}_N]_{\boldsymbol{i},\boldsymbol{j}}\,.$$

In the shift-invariant case these restrictions determine the state completely. As a consequence, the purity of φ implies the purity of $\hat{\varphi}$ for even shift-invariant states.

Example 7.6. Let φ be the Fock state on $\mathcal{A}$, with corresponding symbol $Q = 0$. The restriction φ_N to $\mathcal{A}_{[0,N-1]}$ is again a quasi-free state with symbol 0_N, where 0_N is the $N\times N$ zero operator on the subspace spanned by $\{\delta_0,\ldots,\delta_{N-1}\}$. Since the symbol of φ_N is a projection then φ_N is pure, implying that $\hat{\varphi}_N$ is also pure for every N, which immediately yields that $\hat{\varphi}$ is a pure product state on $\mathcal{C}$. To identify it, it is enough to know the density of $\hat{\varphi}_1$. Since the state is even, $[\hat{\varphi}_1]_{12} = \hat{\varphi}_1(\hat{E}^1_{21}) = 0 = \hat{\varphi}_1(\hat{E}^1_{12}) = [\hat{\varphi}_1]_{21}$. For the diagonal matrix elements we have $[\hat{\varphi}_1]_{11} = \hat{\varphi}_1(\hat{E}^1_{11}) = \varphi(a_1a_1^*) = 1$ and $[\hat{\varphi}_1]_{22} = \hat{\varphi}_1(\hat{E}^1_{22}) = \varphi(a_1^*a_1) = 0$. The density is then

$$[\hat{\varphi}_1] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = |{\uparrow}\rangle\langle{\uparrow}|,$$

with $|{\uparrow}\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and so $\hat{\varphi}$ is the pure product state $|{\uparrow}\rangle\langle{\uparrow}|^{\otimes\infty}$.

A completely similar argument shows that if φ is the anti-Fock state, then $\hat{\varphi} = |{\downarrow}\rangle\langle{\downarrow}|^{\otimes\infty}$.

A Bosonic systems

Let (H,σ) be a real symplectic space and $\kappa \in \mathbb{R}\backslash\{0\}$. We say that a map $W : H \to \mathcal{A}$ to a C^*-algebra $\mathcal{A}$ is a realization of the (κ,σ)-canonical commutation relations if $\mathcal{A}$ is generated by $\{W(x) : x \in H\}$, and the following relations hold:

$$W(x)^* = W(-x) \tag{A.196}$$
$$W(x)W(y) = e^{-i\kappa\sigma(x,y)}W(x+y)\,. \tag{A.197}$$

Obviously, $\kappa\sigma$ is again a symplectic form, hence the introduction of κ may seem superfluous in the definition. However, we follow this terminology in order to be as compatible as possible with the various conventions appearing in the literature. For instance, one can find $\kappa = \frac{1}{2}$ in [?], $\kappa = -\frac{1}{2}$ in [?], $\kappa = -1$ in [?] and $\kappa = \frac{1}{2\hbar}$ in [?].

Note that if such a realization exists then $W(x)$ is a unitary with $W(x)^{-1} = W(-x)$ for all $x \in H$ and $W(0) = I$, independent of the concrete realization of the CCR relations. Indeed,

$$W(x)W(0) = e^{-i\kappa\sigma(x,0)}W(x+0) = W(x), \qquad W(0)W(x) = e^{i\kappa\sigma(x,0)}W(0+x) = W(x),$$

and

$$W(x)^*W(x) = W(-x)W(x) = e^{-i\kappa\sigma(-x,x)}W(0) = I,$$
$$W(x)W(x)^* = W(x)W(-x) = e^{-i\kappa\sigma(x,-x)}W(0) = I.$$

Two realizations $W_1 : H \to \mathcal{A}_1$ and $W_2 : H \to \mathcal{A}_2$ are said to be *isomorphic* if there exists a C^*-algebra isomorphism $\alpha : \mathcal{A}_1 \to \mathcal{A}_2$ such that $\alpha \circ W_1 = W_2$. If $\mathcal{A}_1 \subset \mathcal{B}(\mathcal{H}_1)$ and $\mathcal{A}_2 \subset \mathcal{B}(\mathcal{H}_2)$ with some Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ and α is implemented by a untary $U : \mathcal{H}_1 \to \mathcal{H}_2$ then the two realizations are said to be *unitarily equivalent* to each other.

Two questions arises naturally: whether a realization exists, and if so, whether it is unique up to isomorphism. It is easy to see that the answer to the first question is affirmative. Indeed, define

$$\mathcal{H} := l^2(H) := \left\{ f : H \to \mathbb{C},\ \sum_{x\in H} |f(x)|^2 < +\infty \right\},$$

which is a Hilbert space with the inner product

$$\langle f, g\rangle := \sum_{x\in H} \overline{f(x)}g(x).$$

Let M_x be the multiplication operator by the function $e^{i\kappa\sigma(x,.)}$ and S_x be the translation by x, $(S_x f)(y) := f(y-x)$. Define

$$W(x) := M_x S_x, \quad \text{i.e.,} \quad (W(x)f)(y) := e^{i\kappa\sigma(x,y)}f(y-x)\,.$$

Then,

$$\begin{aligned}\langle W(x)^* g, f\rangle &= \langle g, W(x) f\rangle = \sum_{y\in H} \overline{g(y)} e^{i\kappa\sigma(x,y)} f(y-x) \\ &= \sum_{u\in H} \overline{g(u+x)} e^{i\kappa\sigma(x,u+x)} f(u) = \sum_{u\in H} \overline{g(u+x) e^{-i\kappa\sigma(x,u)}} f(u) \\ &= \sum_{u\in H} \overline{(W(-x)g)(u)} f(u) = \langle W(-x)g, f\rangle \ ,\end{aligned}$$

and

$$\begin{aligned}(W(x_1)W(x_2)f)(y) &= e^{i\kappa\sigma(x_2,y)} (W(x_1)f)(y-x_2) \\ &= e^{i\kappa\sigma(x_2,y)} e^{i\kappa\sigma(x_1,y-x_2)} f(y-x_2-x_1) \\ &= e^{-i\kappa\sigma(x_1,x_2)} e^{i\kappa\sigma(x_1+x_2,y)} f(y-(x_2+x_1)) \\ &= e^{-i\kappa\sigma(x_1,x_2)} \left(W(x_1+x_2)f\right)(y) \,,\end{aligned}$$

hence the CCR relations (A.196) hold.

Note that, in the above construction the non-degeneracy of σ didn't play a role, hence we can construct a realization of the (κ,σ)-CCR relations for any anti-symmetric bilinear form σ.

Theorem A.1. (Slawny) Any two realizations of the (κ,σ)-CCR relations are isomorphic to each other if κ is non-degenerate.

The proof of the above theorem is somewhat involved; we refer the interested reader to [**?**, Theorem 5.2.8] or [**?**, Theorem 2.1]. We will later give a proof in the case when H is finite-dimensional, following [**?**]. As we would like to benefit from the above uniqueness theorem, we will assume for the rest that σ is non-degenerate.

By the above, we can talk about *the* C^*-algebra realizing the (κ,σ)-CCR relations, which we will denote by $\mathrm{CCR}(H,\kappa,\sigma)$. In sections A.1 and A.2, we will describe two convenient and widely used special representations for finite-dimensional H, the *Fock representation* and the *Schrödinger representation*. We will refer to the above given representation on $l^2(H)$ as the *standard representation*.

Corollary A.2. $\mathrm{CCR}(H,\kappa,\sigma)$ is simple.

Proof. Assume that $\mathcal{I}$ is a two-sided ideal in $\mathrm{CCR}(H,\kappa,\sigma)$, and let $\pi : CCR(H,\kappa,\sigma) \to CCR(H,\kappa,\sigma)/\mathcal{I}$ be the factorization map, with kernel $\mathcal{I}$. Then, $\{\pi(W(x)) : x\in H\}$ is again a representation of the (κ,σ)-CCR relations, and hence π is an isomorphism, by Theorem A.1. □

Corollary A.3. The set $\{W(x) : x\in H\}$ is linearly independent.

Proof. The statement is independent of which realization we use, hence we can work in the standard representation. If $\sum_{k=1}^n \lambda_k W(x_k) = 0$ then, for any $l = 1, \ldots, n$,

$$0 = \left(\sum_{k=1}^n \lambda_k W(x_k) \right) \mathbf{1}_{\{0\}}(x_l) = \sum_{k=1}^n \lambda_k e^{i\sigma(x_k, x_l)} \mathbf{1}_{\{x_k\}}(x_l) = \lambda_l \,.$$

□

Lemma A.4. There exists a faithful tracial state τ on $CCR(H, \kappa, \sigma)$ such that $\tau(W(x)) = \delta_{x,0}$.

Proof. Define

$$\tau_0 \left(\sum_{x \in H} \lambda_x W(x) \right) := \lambda_0$$

on $\mathcal{A}_0 := \mathrm{span}\{W(x) : x \in H\}$. By the above Corollary, τ_0 is a well-defined linear functional on $\mathcal{A}_0$, and positivity and tracial properties are easy to see. By positivity, for any $a \in \mathcal{A}_0$,

$$|\tau_0(a)|^2 \leq \tau_0(a^* a) \tau_0(I) \leq \|a\|^2 \, \tau_0(I)^2 \,,$$

hence τ_0 is bounded, and therefore have a unique extension τ onto $\mathcal{A} := \mathrm{CCR}\,(H, \kappa, \sigma)$, with $\|\tau\| = \|\tau_0\| \leq \tau(I) = 1$. On the other hand, $\tau(I) = 1$ implies $\|\tau\| = 1 = \tau(I)$, from which positivity of τ follows. The tracial property is obviously inherited by τ, and hence $\{a \in \mathcal{A} : \tau(x^* x) = 0\}$ is a two-sided ideal. Faithfulness of τ then follows from the simplicity of $\mathcal{A}$. □

A.1 Fock representation

Let $\mathcal{H}$ be a complex vector space, $\sigma(x, y) := \mathfrak{Im}\, \langle x, y \rangle$ be its standard symplectic form and $\kappa > 0$. Denote the nth symmetric tensor power of $\mathcal{H}$ by $\vee^n \mathcal{H}$ for all $n \geq 1$ and let $\vee^0 \mathcal{H} := \mathbb{C}$. Let

$$\mathcal{F}(\mathcal{H}) := \bigoplus_{n=0}^{+\infty} \vee^n \mathcal{H}$$

be the *symmetric Fock space*. For every $x \in \mathcal{H}$ we define the corresponding *Fock vector* by

$$x_F := \bigoplus_{n=0}^{+\infty} \frac{1}{\sqrt{n!}} x^{\otimes n} \,.$$

We have

$$\langle x_F, y_F\rangle = e^{\langle x,y\rangle}, \qquad \|x_F\|^2 = e^{\|x\|^2}, \qquad x, y \in \mathcal{H}.$$

The vector

$$0_F = 1 \oplus 0 \oplus 0 \oplus \dots$$

is called the *vacuum vector*.

Lemma A.5. The Fock vectors are linearly independent and their linear span is dense in $\mathcal{F}(\mathcal{H})$.

Define the linear operators $C(x)$, $x \in \mathcal{H}$ on the linear span of the Fock vectors by

$$C(x)y_F := e^{-\frac{1}{2}\|x\|^2 - \langle x,y\rangle}\,(y+x)_F\ .$$

Since the Fock vectors are linearly independent, the above operators are well-defined, and an easy computation shows that

$$\langle C(x)y_F, C(x)z_F\rangle = e^{\langle y,z\rangle} = \langle y_F, z_F\rangle\ .$$

As a consequence, $C(x)$ is norm-preserving and therefore has a unique unitary extension to $\mathcal{F}(\mathcal{H})$, which we denote also by $C(x)$. Moreover, we have the following:

$$\begin{aligned}\langle C(x)^* y_F, z_F\rangle &= \langle y_F, C(x)z_F\rangle = e^{-\frac{1}{2}\|x\|^2 - \langle x,z\rangle} e^{\langle y, z+x\rangle} = e^{-\frac{1}{2}\|x\|^2 + \langle y,x\rangle + \langle y-x,z\rangle}\\ &= \langle C(-x)y_F, z_F\rangle\,,\end{aligned}$$

and

$$\begin{aligned}C(x_1)C(x_2)y_F &= e^{-\frac{1}{2}\|x_2\|^2 - \langle x_2,y\rangle} C(x_1)\,(y+x_2)_F\\ &= e^{-\frac{1}{2}\|x_2\|^2 - \langle x_2,y\rangle} e^{-\frac{1}{2}\|x_1\|^2 - \langle x_1, y+x_2\rangle}\,(y+x_2+x_1)_F\\ &= e^{-\,\mathfrak{Im}\langle x_1,x_2\rangle} e^{-\frac{1}{2}\|x_1+x_2\|^2 - \langle x_1+x_2,y\rangle}\,(y+x_2+x_1)_F\\ &= e^{-\,\mathfrak{Im}\langle x_1,x_2\rangle} C(x_1+x_2)y_F,\end{aligned}$$

that is,

$$C(x)^* = C(-x), \qquad C(x_1)C(x_2) = e^{-i\,\mathfrak{Im}\langle x_1,x_2\rangle} C(x_1+x_2), \qquad x, x_1, x_1 \in \mathcal{H}.$$

Consequently, for

$$W_\kappa(x) := C(\sqrt{\kappa}x), \qquad x \in \mathcal{H}$$

we have

$$W_\kappa(x)^* = W_\kappa(-x), \qquad W_\kappa(x_1)W_\kappa(x_2) = e^{-i\kappa\,\mathfrak{Im}\langle x_1,x_2\rangle}W_\kappa(x_1+x_2), \qquad x, x_1, x_2 \in \mathcal{H},$$

i.e., the operators $W_\kappa(x)$, $x \in \mathcal{H}$ yield a realization of the (κ,σ)-CCR relations for the standard symplectic form σ.

For $\kappa < 0$, one can modify the above construction by defining

$$W_\kappa(x) := C\left(\sqrt{|\kappa|}Jx\right),$$

with some anti-unitary J. Then again, $W_\kappa(x)$, $x \in \mathcal{H}$ yield a realization of the (κ,σ)-CCR relations for the standard symplectic form σ.

By Lemma B.17, any finite-dimensional symplectic space is isomorphic to $(\mathbb{C}^d, \sigma_{\mathbb{C}^d})$ with $d = \dim H/2$. Choosing $\mathcal{H} := \mathbb{C}^d$ with its standard inner product to be the Hilbert space in the above construction, we get a representation of CCR(H,κ,σ). This gives an alternative proof for the existence of a realization of the (κ,σ)-CCR relations on H when H is finite-dimensional.

Note that for $\kappa > 0$,

$$W_\kappa(x)y_F = e^{-\frac{1}{2}\kappa\|x\|^2}e^{-\sqrt{\kappa}\langle x,y\rangle}\left(y+\sqrt{\kappa}x\right)_F.$$

In particular,

$$W_\kappa(x)0_F = e^{-\frac{1}{2}\kappa\|x\|^2}\left(\sqrt{\kappa}x\right)_F.$$

As a consequence, the normalized Fock vector (also referred to as *coherent state*) is obtained from the vacuum by

$$e^{-\frac{1}{2}\|x\|^2}x_F = W_\kappa\left(\frac{1}{\sqrt{\kappa}}x\right)0_F.$$

A.2 Schrödinger representation

For every $x \in \mathbb{R}^d$, consider the corresponding multiplication and translation operators

$$(M_x f)(y) := e^{ixy}f(y), \qquad (S_x f)(y) := f(y-x), \qquad y \in \mathbb{R}^d$$

on $L^2(\mathbb{R}^d)$. A straightforward computation yields

$$S_x^* = S_{-x}, \quad S_xS_y = S_{x+y}, \qquad M_x^* = M_{-x}, \quad M_xM_y = M_{x+y}, \qquad x, y \in \mathbb{R}^d,$$

by which S_x and M_x are unitaries for all $x \in \mathbb{R}^d$, and both $x \mapsto S_x$ and $x \mapsto M_x$ give unitary representations of $\mathbb{R}^d$ on $L^2(\mathbb{R}^d)$. Moreover,

$$\begin{aligned} S_{x_1} M_{x_2} f(y) &= (M_{x_2} f)(y - x_1) = e^{ix_2(y-x_1)} f(y - x_1) \\ &= e^{-ix_1x_2} e^{ix_2y} f(y - x_1) \\ &= e^{-ix_1x_2} M_{x_2} S_{x_1} f(y). \end{aligned} \tag{A.198}$$

For every $(x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$, define

$$W_S(x_1, x_2) := e^{-\frac{i}{2}x_1x_2} M_{x_2} S_{x_1},$$

i.e.,

$$W_S(x_1, x_2) f(y) = e^{-\frac{i}{2}x_1x_2} e^{ix_2y} f(y - x_1).$$

By the above,

$$W_S(x_1, x_2)^* = e^{\frac{i}{2}x_1x_2} S_{-x_1} M_{-x_2} = e^{\frac{i}{2}x_1x_2} e^{-\frac{i}{2}x_1x_2} M_{-x_2} S_{-x_1} = W_S(-x_1, -x_2),$$

$$\begin{aligned} W_S(x_1, x_2) W_S(y_1, y_2) &= e^{-\frac{i}{2}x_1x_2} M_{x_2} S_{x_1} e^{-\frac{i}{2}y_1y_2} M_{y_2} S_{y_1} \\ &= e^{-\frac{i}{2}x_1x_2} e^{-\frac{i}{2}y_1y_2} M_{x_2} e^{-ix_1y_2} M_{y_2} S_{x_1} S_{y_1} \\ &= e^{-\frac{i}{2}(x_1y_2 - x_2y_1)} e^{-\frac{i}{2}(x_1+y_1)(x_2+y_2)} M_{x_1+y_1} S_{x_2+y_2} \\ &= e^{-\frac{i}{2}(x_1y_2 - x_2y_1)} W_S(x_1 + y_1, x_2 + y_2). \end{aligned}$$

Hence, $\{W_S(x_1, x_2) : (x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d\}$ gives a realization of the $\left(\frac{1}{2}, \sigma_{\mathbb{R}^{2d}}\right)$-commutation relations on $\mathbb{R}^d \times \mathbb{R}^d$. In order to get a realization of the $(\kappa, \sigma_{\mathbb{R}^{2d}})$-commutation relations, one can define

$$W_{S,\kappa}(x_1, x_2) := \begin{cases} W_S\left(\sqrt{2\kappa}\, x_1, \sqrt{2\kappa}\, x_2\right), & \kappa > 0, \\ W_S\left(\sqrt{2|\kappa|}\, x_2, \sqrt{2|\kappa|}\, x_1\right), & \kappa < 0. \end{cases}$$

For $g \in L^2(\mathbb{R}^d)$, the Fourier transform of g can be written as

$$\hat{g}(x) := (\mathcal{F}_d g)(x) = \lim_{M \to \infty} \frac{1}{\sqrt{2\pi}^d} \int_{[-M,M]^d} e^{-ixy} g(y)\, d\lambda(y), \qquad x \in \mathbb{R}^d$$

and one has the inversion formula

$$g(y) = \lim_{M \to \infty} \frac{1}{\sqrt{2\pi}^d} \int_{[-M,M]^d} e^{-ixy} \hat{g}(x)\, d\lambda(x), \qquad y \in \mathbb{R}^d.$$

Obviously, if $f, g \in L^2\left(\mathbb{R}^d\right)$ then

$$e^{-i\pi_1 \otimes \pi_2} \overline{f} \otimes \hat{g} : (y, z) \mapsto e^{-iyz} \overline{f(y)} \hat{g}(z), \qquad y, z \in \mathbb{R}^d$$

is an element of $L^2\left(\mathbb{R}^d \times \mathbb{R}^d\right)$, and we have the following:

Lemma A.6. Let $\{W_S(x) : x \in \mathbb{R}^d \times \mathbb{R}^d\}$ be the Schrödinger representation of the $(\frac{1}{2}, \sigma_{\mathbb{R}^{2d}})$-CCR relations. Then,

$$\langle f, W_S(x)\, g\rangle = e^{-\frac{i}{2}x^{(1)}x^{(2)}}\sqrt{2\pi}^d \mathcal{F}_{2d}\left(e^{-i\pi_1\otimes\pi_2}\overline{f}\otimes\hat{g}\right)\left(-x^{(2)}, x^{(1)}\right).$$

Proof.

$$\begin{aligned}
&\langle f, W(x)g\rangle = \\
&= \int \overline{f}(y)\,(W(x)g)\,(y)\,d\lambda(y) \\
&= \int \overline{f}(y) e^{-\frac{i}{2}x^{(1)}x^{(2)}} e^{ix^{(2)}y} g(y-x^{(1)})\,d\lambda(y) \\
&= e^{-\frac{i}{2}x^{(1)}x^{(2)}} \int \overline{f}(y) e^{ix^{(2)}y} g(y-x^{(1)})\,d\lambda(y) \\
&= e^{-\frac{i}{2}x^{(1)}x^{(2)}} \lim_{K\to\infty} \int_{[-K,K]^d} \overline{f}(y) e^{ix^{(2)}y} g(y-x^{(1)})\,d\lambda(y) \\
&= e^{-\frac{i}{2}x^{(1)}x^{(2)}} \lim_{K,M\to\infty} \frac{1}{\sqrt{2\pi}^d} \int_{[-K,K]^d}\int_{[-M,M]^{2d}} \overline{f}(y) e^{ix^{(2)}y} e^{iz\left(y-x^{(1)}\right)} \hat{g}(z)\,d\lambda(z)d\lambda(y) \\
&= e^{-\frac{i}{2}x^{(1)}x^{(2)}} \lim_{K,M\to\infty} \frac{1}{\sqrt{2\pi}^d} \int_{[-K,K]^d}\int_{[-M,M]^{2d}} e^{-i\left(-yx^{(2)}+zx^{(1)}\right)} e^{izy}\overline{f}(y)\hat{g}(z)\,d\lambda(z)d\lambda(y),
\end{aligned}$$

which is just $\sqrt{2\pi}^d e^{-\frac{i}{2}x^{(1)}x^{(2)}}$ times the Fourier transform of the $2d$-variable function

$$(y,z) \mapsto e^{izy}\overline{f}(y)\hat{g}(z)$$

at $(-x^{(2)}, x^{(1)})$. □

Corollary A.7.

$$\langle f, W_{S,\kappa}(x)\, g\rangle = \begin{cases} e^{-i\kappa x^{(1)}x^{(2)}}\sqrt{2\pi}^d \mathcal{F}_{2d}\left(e^{-i\pi_1\otimes\pi_2}\overline{f}\otimes\hat{g}\right)\left(-\sqrt{2\kappa}x^{(2)}, \sqrt{2\kappa}x^{(1)}\right), & \kappa > 0, \\ e^{i\kappa x^{(1)}x^{(2)}}\sqrt{2\pi}^d \mathcal{F}_{2d}\left(e^{-i\pi_1\otimes\pi_2}\overline{f}\otimes\hat{g}\right)\left(-\sqrt{2|\kappa|}x^{(1)}, \sqrt{2\kappa}x^{(2)}\right), & \kappa < 0. \end{cases}$$

Corollary A.8. Let $\{f_n : n \in \mathbb{N}\}$ and $\{g_n : n \in \mathbb{N}\}$ be orthonormal bases in $L^2(\mathbb{R}^d)$. Then,

$$\varphi_{n,m}(x) := \sqrt{\frac{|\kappa|}{\pi}}^d \langle f_n, W_{S,\kappa}(x) g_m\rangle, \qquad x \in \mathbb{R}^{2d}, \quad n,m \in \mathbb{N}$$

is an orthonormal basis in $L^2\left(\mathbb{R}^{2d}\right)$.

Proof. Since the Fourier transformation is a unitary operator, $\hat{g}_m$, $m \in \mathbb{N}$ is also an orthonormal basis in $L^2\left(\mathbb{R}^d\right)$, and hence

$$\tilde{\varphi}_{n,m}(y,z) := e^{izy}\left(\overline{f}_n \otimes \hat{g}_m\right)(y,z) = e^{izy}\overline{f}_n(y)\hat{g}_m(z), \qquad y,z \in \mathbb{R}^d$$

is an orthonormal basis in $L^2\left(\mathbb{R}^d\right) \otimes L^2\left(\mathbb{R}^d\right) = L^2\left(\mathbb{R}^{2d}\right)$. Thus, $\mathcal{F}_{2d}\varphi_{n,m},\ n,m \in \mathbb{N}$ is again an orthonormal basis in $L^2\left(\mathbb{R}^{2d}\right)$. Consequently,

$$x \mapsto \begin{cases} \sqrt{2\kappa}^d \mathcal{F}_{2d}\varphi_{n,m}\left(-\sqrt{2\kappa}x^{(2)}, \sqrt{2\kappa}x^{(1)}\right), & \kappa > 0, \\ \sqrt{2|\kappa|}^d \mathcal{F}_{2d}\varphi_{n,m}\left(-\sqrt{2|\kappa|}x^{(1)}, \sqrt{2|\kappa|}x^{(2)}\right), & \kappa < 0, \end{cases}$$

are again orthonormal bases. The statement then follows from Lemma A.7. □

Lemma A.9.

$$\int_{\mathbb{R}^{2d}} \overline{\langle f_1, W_{S,\kappa}(x)g_1\rangle}\, \langle f_2, W_{S,\kappa}(x)g_2\rangle\, d\lambda(x) = \left(\frac{\pi}{|\kappa|}\right)^d \langle f_2, f_1\rangle\, \langle g_1, g_2\rangle, \qquad f_1, f_2, g_1, g_2 \in L^2(\mathbb{R}^d).$$

Proof. By a simple integral transformation,

$$\int_{\mathbb{R}^{2d}} \overline{\langle f_1, W_{S,\kappa}(x)g_1\rangle}\, \langle f_2, W_{S,\kappa}(x)g_2\rangle\, d\lambda(x) = \left(\frac{1}{\sqrt{2|\kappa|}}\right)^{2d} \int_{\mathbb{R}^{2d}} \overline{\langle f_1, W_S(x)g_1\rangle}\, \langle f_2, W_S(x)g_2\rangle\, d^{2d}x.$$

Now, by Lemma A.6 and Parseval's identity,

$$\begin{aligned} &\int_{\mathbb{R}^{2d}} \overline{\langle f_1, W_S(x)g_1\rangle}\, \langle f_2, W_S(x)g_2\rangle\, d^{2d}x \\ &= (2\pi)^d \int_{\mathbb{R}^{2d}} e^{izy} f_1(y)\, \overline{\hat{g}_1(z)} e^{-izy} \overline{f_2(y)} \hat{g}_2(z)\, dzdy \\ &= (2\pi)^d \int_{\mathbb{R}^d} \overline{f_2(y)} f_1(y)\, dy \int_{\mathbb{R}^d} \overline{\hat{g}_1(z)} \hat{g}_2(z)\, dz \\ &= (2\pi)^d \int_{\mathbb{R}^d} \overline{f_2(y)} f_1(y)\, dy \int_{\mathbb{R}^d} \overline{g_1(z)} g_2(z)\, dz \\ &= (2\pi)^d \langle f_2, f_1\rangle\, \langle g_1, g_2\rangle. \end{aligned}$$

□

For a general finite-dimensional symplectic space (H,σ), one obtains a realization of the (κ,σ)-CCR relations by first fixing an isomorphism T with $\left(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}}\right)$ (or equivalently, by fixing a symplectic basis, see Lemma B.17) and then apply the above construction. That is, we define

$$W_\kappa(x) := W_{S,\kappa}(Tx), \qquad x \in H.$$

Note that the so obtained representation depends on the isomorphism (the symplectic basis) chosen, but any two such representations are unitarily equivalent to each other. Hence, we call any such representation *the Schrödinger representation* of the (κ, σ)-CCR relations.

Lemma A.10. Let $\{W(x) : x \in H\}$ be a continuous irreducible representation of $\mathrm{CCR}(H, \sigma, \kappa)$ for a finite-dimensional H on a Hilbert space $\mathcal{H}$. Then,

$$\int_H \overline{\langle f_1, W(x) g_1 \rangle} \langle f_2, W(x) g_2 \rangle \, d\lambda(x) = \left(\frac{\pi}{|\kappa|} \right)^d \langle f_2, f_1 \rangle \langle g_1, g_2 \rangle , \qquad f_1, f_2, g_1, g_2 \in \mathcal{H}.$$

Proof. It follows from the previous lemma, by noting that any two continuous irreducible representations of the CCR are unitarily equivalent to each other. □

Now, let $\{W_\kappa(x) : x \in H\}$ be a continuous irreducible representation of the (κ, σ)-CCR relations of a finite-dimensional H on a Hilbert space $\mathcal{H}$. The *characteristic function* of a trace-class operator $T \in \mathcal{B}_1(\mathcal{H})$ is

$$\hat{W}_\kappa[T](x) := \operatorname{Tr} T W(-x), \qquad x \in H.$$

Lemma A.11. Let $T_1, T_2 \in \mathcal{B}_2(\mathcal{H})$ be Hilbert-Schmidt operators on $\mathcal{H}$. Then, $\hat{W}[T_1], \hat{W}[T_2] \in L^2(H)$, and

$$\int_H \overline{\hat{W}_\kappa[T_1](x)} \hat{W}_\kappa[T_2](x) \, d\lambda(x) = \left(\frac{\pi}{|\kappa|} \right)^d \operatorname{Tr} T_1^* T_2$$

Proof. Let $T \in \mathcal{B}_2(\mathcal{H})$ be self-adjoint with an eigen-decomposition $T = \sum_k t_k \, |e_k\rangle\langle e_k|$. Then,

$$\hat{W}_\kappa[T](x) = \operatorname{Tr} T W(-x) = \sum_k t_k \langle e_k, W(-x) e_k \rangle .$$

By the previous lemmas,

$$f_k(x) := \langle e_k, W(-x) e_k \rangle \in L^2(H)$$

and

$$\langle f_k, f_m \rangle = \int_H \overline{f_k(x)} f_m(x) = \left(\frac{\pi}{|\kappa|} \right)^d \langle e_m, e_k \rangle \langle e_k, e_m \rangle = \delta_{m,l} \left(\frac{\pi}{\kappa} \right)^d .$$

Hence,

$$\left\| \hat{W}_{|\kappa|}[T] \right\|_2^2 = \sum_k |t_k|^2 \, \|f_k\|_2^2 = \left(\frac{\pi}{\kappa} \right)^d \sum_k |t_k|^2 = \left(\frac{\pi}{|\kappa|} \right)^d \operatorname{Tr} T^2.$$

Now, the assertion follows by polarization. □

Proposition A.12. The Schrödinger representation is irreducible.

A.3 Gaussian states

Let (H, σ) be a symplectic space and α be a positive definite symmetric form such that

$$\sigma(x,y)^2 \leq \alpha(x,x)\alpha(y,y), \qquad x,y \in H, \tag{A.199}$$

or equivalently,

$$\alpha + i\sigma \geq 0. \tag{A.200}$$

Then, there exists a unique state ϱ_α on $CCR(H,\kappa,\sigma)$ such that

$$\varrho_\alpha\left(W(x)\right) = e^{-\frac{|\kappa|}{2}\alpha(x,x)}, \qquad x \in H.$$

States defined in the above way are called *quasi-free states*.

If ϱ_α is a quasi-free state then its unitary rotation by some $W(y)$,

$$\varrho_{\alpha,y}(a) := \varrho_\alpha\left(W(y)^* a W(y)\right), \qquad a \in \mathrm{CCR}\left(H,\kappa,\sigma\right)$$

is again a state, with characteristic function

$$\varrho_{\alpha,y}\left(W(-x)\right) = \varrho_\alpha\left(W(y)^* W(-x) W(y)\right) = e^{-2i\kappa\sigma(y,x)}\varrho_\alpha\left(W(-x)\right) = e^{-2i\kappa\sigma(y,x)}e^{-\frac{|\kappa|}{2}\alpha(x,x)}.$$

Definition A.13. A state $\varrho_{\alpha,y}$ is called a *Gaussian state* with *displacement vector* y and *covariance* α.

Note that with $\alpha' := |\kappa|\alpha$, condition (A.199) becomes

$$\kappa^2\sigma(x,y)^2 \leq \alpha'(x,x)\alpha'(y,y), \qquad x,y \in H,$$

or equivalently,

$$\alpha' + i\kappa\sigma \geq 0,$$

and

$$\varrho_{\alpha,y}\left(W(-x)\right) = e^{-2i\kappa\sigma(y,x)}e^{-\frac{1}{2}\alpha'(x,x)}, \qquad x \in H.$$

In the literature (e.g., in [?]) sometimes α' is referred to as the covariance of ϱ_α. We will, however, use the previous convention.

For the rest we assume that H is finite-dimensional.

Lemma A.14. Let ϱ_α be a Gaussian state with zero displacement and let H_α be the α-canonical complexification of H. Then, ϱ_α is normal with respect to the Fock representation on $\mathcal{F}(H_\alpha)$ and

$$\mathcal{D}\varrho_\alpha = \frac{2^{\dim H_\alpha}}{\det(I+Q)}\left(\frac{Q-I}{Q+I}\right)_F = \frac{1}{\det(I+\tilde{Q})}\left(\frac{\tilde{Q}}{\tilde{Q}+I}\right)_F$$

where Q is the symbol of α and $2\tilde{Q} := Q - I$.

Proof. □

Corollary A.15. Let $\mathcal{H}$ be an arbitrary complexification of H. Every Gaussian state is normal with respect to the Fock representation on $\mathcal{F}(\mathcal{H})$.

Proof. Let $\varrho_{\alpha,y}$ be a Gaussian state and H_α be the α-canonical complexification of H. Since $\mathcal{F}(H_\alpha)$ and $\mathcal{F}(\mathcal{H})$ are both irreducible representations, there exists a unitary $U : \mathcal{F}(H_\alpha) \to \mathcal{F}(\mathcal{H})$ such that $UW_{H_\alpha}(x)U^* = W_{\mathcal{H}}(x)$, $x \in H$, where W_{H_α} and $W_{\mathcal{H}}$ are the representations of $\mathrm{CCR}(\kappa, \sigma, H)$ on $\mathcal{F}(H_\alpha)$ and $\mathcal{F}(\mathcal{H})$, respectively. Consequently, ϱ_α is given be the density

$$W_{\mathcal{H}}(y)U\mathcal{D}\varrho_\alpha U^* W_{\mathcal{H}}(y),$$

where $\mathcal{D}\varrho_\alpha$ is the density of ϱ_α on $\mathcal{F}(H_\alpha)$. □

Lemma A.16. Let ϱ_{α_1,y_1} and ϱ_{α_2,y_2} be Gaussian states. Then,

$$\mathrm{Tr}\,\mathcal{D}\varrho_{\alpha_1,y_1}\mathcal{D}\varrho_{\alpha_2,y_2} = \frac{2^d}{\sqrt{\det(\alpha_1+\alpha_2)}} e^{-2|\kappa|(\alpha_1+\alpha_2)^{-1}(y,y)},$$

where $y := y_1 - y_2$.

Proof. By [?, Chapter V, Theorem 3.2],

$$\begin{aligned}
\mathrm{Tr}\,\mathcal{D}\varrho_{\alpha_1,y_1}\mathcal{D}\varrho_{\alpha_2,y_2} &= \left(\frac{|\kappa|}{\pi}\right)^d \int_H e^{2i\kappa\sigma(y_1,x)-\frac{|\kappa|}{2}\alpha_1(x,x)} e^{-2i\kappa\sigma(y_2,x)-\frac{|\kappa|}{2}\alpha_2(x,x)}\, d\lambda(x) \\
&= \left(\frac{|\kappa|}{\pi}\right)^d \int_H e^{2i\kappa\sigma(y_1-y_2,x)-\frac{|\kappa|}{2}(\alpha_1+\alpha_2)(x,x)}\, d\lambda(x) \\
&= \left(\frac{|\kappa|}{\pi}\right)^d \int_H e^{2i\kappa\sigma(y,x)-\frac{|\kappa|}{2}\alpha(x,x)}\, d\lambda(x),
\end{aligned}$$

with $y := y_1 - y_2$ and $\alpha := \alpha_1 + \alpha_2$.

Let us fix a $|\kappa|\alpha$-canonical basis $e_1, \dots, e_{2d}$, and denote the symplectic eigenvalues of $|\kappa|\alpha$ with $a_1, \dots, a_k$. Then,

$$2i\kappa\sigma(y,x) - \frac{|\kappa|}{2}\alpha(x,x) = \sum_{k=1}^{d} 2i\kappa\left(y_{2k-1}x_{2k} - y_{2k}x_{2k-1}\right) - \frac{1}{2}a_k\left(x_{2k-1}^2 + x_{2k}^2\right),$$

where $x = \sum_{m=1}^{2d} x_m e_m$, $y = \sum_{m=1}^{2d} y_m e_m$. Hence,

$$\operatorname{Tr}\mathcal{D}\varrho_{\alpha_1,y_1}\mathcal{D}\varrho_{\alpha_2,y_2} = \prod_{k=1}^{d}\frac{|\kappa|}{\pi}\int \exp\left(2i\kappa\left(y_{2k-1}x_{2k} - y_{2k}x_{2k-1}\right) - \frac{1}{2}a_k\left(x_{2k-1}^2 + x_{2k}^2\right)\right) dx_{2k-1}dx_{2k}.$$

Let us compute the integral for a fixed k. The exponent is

$$\begin{aligned}
&-\frac{1}{2}a_k\left[x_{2k-1}^2 + \frac{4i\kappa}{a_k}y_{2k}x_{2k-1}\right] - \frac{1}{2}a_k\left[x_{2k}^2 - \frac{4i\kappa}{a_k}y_{2k-1}x_{2k}\right]\\
&= -\frac{1}{2}a_k\left[x_{2k-1}^2 - 2w_{2k-1}x_{2k-1}\right] - \frac{1}{2}a_k\left[x_{2k}^2 - 2w_{2k}x_{2k}\right]\\
&= -\frac{1}{2}a_k\left(x_{2k-1} - w_{2k-1}\right)^2 - \frac{1}{2}a_k\left(x_{2k} - w_{2k}\right)^2 + \frac{1}{2}a_k\left(w_{2k-1}^2 + w_{2k}^2\right),
\end{aligned}$$

with

$$w_{2k-1} := -\frac{2i\kappa}{a_k}y_{2k}, \qquad w_{2k} := \frac{2i\kappa}{a_k}y_{2k-1}.$$

Since

$$\int \exp\left(-\frac{1}{2}a_k\left(x_{2k-1} - w_{2k-1}\right)^2\right) dx_{2k-1} = \int \exp\left(-\frac{1}{2}a_k\left(x_{2k} - w_{2k}\right)^2\right) dx_{2k} = \sqrt{\frac{2\pi}{a_k}}$$

and

$$\frac{1}{2}a_k\left(w_{2k-1}^2 + w_{2k}^2\right) = -\frac{2\kappa^2}{a_k}\left(y_{2k-1}^2 + y_{2k}^2\right),$$

we get

$$\operatorname{Tr}\mathcal{D}\varrho_{\alpha_1,y_1}\mathcal{D}\varrho_{\alpha_2,y_2} = \prod_{k=1}^{d}\frac{2|\kappa|}{a_k}e^{-\frac{2\kappa^2}{a_k}\left(y_{2k-1}^2+y_{2k}^2\right)} = \frac{2^d|\kappa|^d}{\sqrt{\det|\kappa|\alpha}}e^{-2\kappa^2(|\kappa|\alpha)^{-1}(y,y)},$$

which yields the desired formula. □

Lemma A.17. Let $\mathcal{D}\varrho_{(\alpha,y)}$ be the density of the Gaussian state $\varrho_{(\alpha,y)}$. Then, for any $0 < t \leq 1$,

$$\left(\mathcal{D}\varrho_{(\alpha,y)}\right)^t = N_{\alpha,t}\, \mathcal{D}\varrho_{(f_t(\alpha),y)}$$

where

$$N_{\alpha,t} := 2^{td} \det\left[(\alpha+1)^t - (\alpha-1)^t\right]^{-1/2} = \prod_{k=1}^{d} \frac{2^t}{(a_k+1)^t - (a_k-1)^t} \tag{A.201}$$

with $a_1, \ldots, a_d$ the symplectic eigenvalues of α, and

$$f_t(x) := \frac{(x+1)^t + (x-1)^t}{(x+1)^t - (x-1)^t}\,.$$

Proof. Based on the computation in the next section, which should be slightly updated.

Corollary A.18. For Gaussian states ϱ_{α_1,y_1} and ϱ_{α_2,y_2} and $0 < t < 1$,

$$\begin{aligned}&\operatorname{Tr}\left(\mathcal{D}\varrho_{\alpha_1,y_1}\right)^t \left(\mathcal{D}\varrho_{\alpha_2,y_2}\right)^{1-t} \\ &\quad = N_{\alpha_1,t} N_{\alpha_2,1-t}\, \frac{2^d}{\sqrt{\det\left(f_t\left(\alpha_1\right) + f_{1-t}\left(\alpha_2\right)\right)}} e^{-2|\kappa|\left(f_t(\alpha_1)+f_{1-t}(\alpha_2)\right)^{-1}(y,y)}\,,\end{aligned}$$

where $y := y_1 - y_2$.

Proof. It follows immediately from the previous two lemmas. □

A.4 Gaussian states on complex Hilbert spaces

Let $\mathcal{H}$ be a complex Hilbert space. The second-quantized version $z \mapsto (zI)_F$ of its gauge-group $z \mapsto zI$ gives a unitary representation of the complex unit circle $\mathbb{T}$ on $\mathcal{F}(\mathcal{H})$, and

$$\gamma_z : W(x) \mapsto (zI)_F^*\, W(x)\, (zI)_F = W(zx)$$

is a (quasi-free) group of automorphisms on $\mathrm{CCR}\left(\mathcal{H}, \kappa\right)$, which is called the *gauge group* of $\mathrm{CCR}\left(\mathcal{H}, \kappa\right)$. A state φ is *gauge-invariant* if $\varphi \circ \gamma_z = \varphi$ for all z on the unit circle.

Lemma A.19. A Gaussian state $\varrho_{\alpha,y}$ is gauge-invariant if and only if its covariance α is gauge-invariant.

Proof. Obviously, $\varrho_{\alpha,y}$ is gauge-invariant if and only if its quasi-free part ϱ_α is gauge-invariant. The assertion then follows by computing expectations of two-term products of field operators; see e.g. [**?**, Formula 3.8]. □

Corollary A.20. A Gaussian state $\varrho_{\alpha,y}$ on a finite-dimensional Hilbert space $\mathcal{H}$ is gauge-invariant if and only if there exists a complex linear operator $Q \geq I$ such that $\alpha(x,y) = \mathfrak{Re}\,\langle Qx, y\rangle\,,\ x, y \in \mathcal{H}$.

Definition A.21. We say that a Gaussian state $\varrho_{\alpha,y}$ has a symbol if there exists a complex linear operator $Q \geq I$ such that $\alpha(x,y) = \mathfrak{Re}\,\langle Qx, y\rangle\,,\ x, y \in \mathcal{H}$. By the above, if $\mathcal{H}$ is finite-dimensional then a Gaussian state has a symbol if and only if it is gauge-invariant. Note that if $\mathcal{H}$ is infinite-dimensional then having a symbol is a possibly stronger condition then being gauge-invariant. For a Gaussian state with symbol Q we will also use the notation $\varrho_{\tilde{Q},y}$, where $2\tilde{Q} := Q - I$.

Example A.22. A state of $\mathrm{CCR}\,(()\,\mathcal{H})$ is *coherent* if it has density $e^{-\|y\|^2}\,|y_F\rangle\langle y_F|$ on $\mathcal{F}(\mathcal{H})$ for some $y \in \mathcal{H}$. Consider the case $\kappa > 0$. The characteristic function of a coherent state is

$$\begin{aligned}\hat{W}_\kappa\left[e^{-\|y\|^2}\,|y_F\rangle\langle y_F|\right] &= e^{-\|y\|^2}\,\langle y_F, W_\kappa(x) y_F\rangle \\ &= e^{-\|y\|^2} e^{-\frac{\kappa}{2}\|x\|^2 - \sqrt{\kappa}\langle x,y\rangle} e^{\left\langle y, y+\sqrt{\kappa}x\right\rangle} \\ &= e^{-\frac{\kappa}{2}\|x\|^2 + 2i\sqrt{\kappa}\,\mathfrak{Im}\langle y,x\rangle}.\end{aligned}$$

That is, $e^{-\|y\|^2}\,|y_F\rangle\langle y_F|$ is the density of a Gaussian state with covariance $\alpha(x,y) = \mathfrak{Re}\,\langle x, y\rangle$ and displacement $\frac{1}{\sqrt{\kappa}}y$. Obviously, a coherent state is gauge-invariant, and it has a symbol $Q = I$, or equivalently, $\tilde{Q} = 0$. Hence,

$$e^{-\|y\|^2}\,|y_F\rangle\langle y_F| = \mathcal{D}\varrho_{0,y/\sqrt{\kappa}}\,,$$

or equivalently,

$$\mathcal{D}\varrho_{0,y} = e^{-\kappa\|y\|^2}\left|\sqrt{\kappa}y_F\right\rangle\left\langle\sqrt{\kappa}y_F\right|\,.$$

Vice versa, if a Gaussian state has a symbol which is I then it is a coherent state. That is, coherent states are exactly those gauge-invariant Gaussian states that have a symbol which is the identity.

Note that

$$e^{-\|y\|^2}\,|y_F\rangle\langle y_F| = W_\kappa\left(y/\sqrt{\kappa}\right)|0_F\rangle\langle 0_F|\,W_\kappa\left(y/\sqrt{\kappa}\right)^*,$$

i.e., every coherent state is obtained from the vacuum state by a unitary rotation. The above equation can also be rewritten as

$$\mathcal{D}\varrho_{0,y} = W_\kappa(y)\ \mathcal{D}\varrho_{0,0}\,W_\kappa(y)^* = W_\kappa(y)\ |0_F\rangle\langle 0_F|\ W_\kappa(y)^*\,.$$

Example A.23. Let $\mathcal{H}$ be finite-dimensional and $\kappa > 0$. For any Borel probability measure μ on $\mathcal{H}$,

$$\begin{aligned} |\mu\rangle\langle\mu| &:= \int_{\mathcal{H}} \left|e^{-\|y\|^2/2} y_F\right\rangle\left\langle e^{-\|y\|^2/2} y_F\right| \, d\mu(y) \\ &= \int_{\mathcal{H}} W_\kappa\left(y/\sqrt{\kappa}\right) |0_F\rangle\langle 0_F| \, W_\kappa\left(y/\sqrt{\kappa}\right)^* \, d\mu(y) \end{aligned}$$

gives a density operator on $\mathcal{F}(\mathcal{H})$ with characteristic function

$$\begin{aligned} \hat{W}_\kappa\left[|\mu\rangle\langle\mu|\right](x) &= \int_{\mathcal{H}} \operatorname{Tr} W_\kappa(x) \left|e^{-\|y\|^2/2} y_F\right\rangle\left\langle e^{-\|y\|^2/2} y_F\right| \, d\mu(y) \\ &= \int_{\mathcal{H}} e^{-\frac{\kappa}{2}\|x\|^2 + 2i\sqrt{\kappa}\,\mathfrak{Im}\langle y,x\rangle} \, d\mu(y) \\ &= e^{-\frac{\kappa}{2}\|x\|^2} \int_{\mathcal{H}} e^{2i\sqrt{\kappa}\,\mathfrak{Im}\langle y,x\rangle} \, d\mu(y). \end{aligned}$$

States of this form are called *classical*.

Obviously, classical states are gauge-invariant. Note that coherent states are classical, and

$$e^{-\|y\|^2} |y_F\rangle\langle y_F| = |\delta_y\rangle\langle\delta_y| \, .$$

Example A.24. Let $\mathcal{H}$ be finite-dimensional, $\kappa > 0$ and $\varrho_{\alpha,y}$ be a gauge-invariant Gaussian state with symbol Q. Define $R := 2\,(Q - I)^{-1} = \tilde{Q}^{-1}$, where the inverse is taken on $\mathcal{H}_1 := \operatorname{supp}(Q - I) = \operatorname{supp} \tilde{Q}$. Let $r := \operatorname{rk} R$, define $\det R$ as the product of the non-zero eigenvalues of R, and let $\gamma(x, y) := \mathfrak{Re}\,\langle Rx, y\rangle$. Let λ be the Lebesgue measure of the symplectic subspace $\mathcal{H}_1$, and define the probability measure μ as

$$\mu(B) := \int_{\mathcal{H}} \mathbf{1}_B \frac{\det R}{\pi^r} e^{-\gamma(x,x)} \, d\lambda(x).$$

Note that μ is a Gaussian measure, supported on $\mathcal{H}_1$. By the above,

$$\begin{aligned} |\mu\rangle\langle\mu| &= \frac{\det R}{\pi^r} \int_{\mathcal{H}_1} e^{-\gamma(y,y)} W_\kappa\left(y/\sqrt{\kappa}\right) |0_F\rangle\langle 0_F| \, W_\kappa\left(y/\sqrt{\kappa}\right)^* \, d\lambda(y) \\ &= \det R \left(\frac{\kappa}{\pi}\right)^r \int_{\mathcal{H}_1} e^{-\kappa\gamma(u,u)} W_\kappa\left(u\right) |0_F\rangle\langle 0_F| \, W_\kappa\left(u\right)^* \, d\lambda(u), \end{aligned}$$

and the characteristic function is

$$\begin{aligned} \hat{W}_\kappa\left[|\mu\rangle\langle\mu|\right](x) &= e^{-\frac{\kappa}{2}\|x\|^2} \frac{\det R}{\pi^r} \int_{\mathcal{H}_1} e^{-\gamma(y,y) + 2i\sqrt{\kappa}\,\mathfrak{Im}\langle y,x\rangle} \, d\lambda(y) \\ &= e^{-\frac{\kappa}{2}\|x\|^2} \det R \left(\frac{\kappa}{\pi}\right)^r \int_{\mathcal{H}_1} e^{-\kappa\gamma(u,u) + 2i\kappa\sigma(u,x)} \, d\lambda(u). \end{aligned}$$

Let r_k denote the symplectic eigenvalues of γ (i.e., the eigenvalues of R), and let $e_1, \ldots, e_r$ be an eigenbasis of R. Then, e_k, ie_k, $k = 1, \ldots, r$ is a γ-canonical basis, and writing out the components of u and x in this basis, we get

$$\begin{aligned}
\gamma(u,u) - 2i\sigma(u,x) &= \sum_{k=1}^{r} r_k \left(u_{2k-1}^2 + u_{2k}^2\right) - 2i\left(u_{2k-1}x_{2k} - u_{2k}x_{2k-1}\right) \\
&= \sum_{k=1}^{r} r_k \left(u_{2k-1}^2 - 2iu_{2k-1}\frac{x_{2k}}{r_k}\right) + r_k\left(u_{2k}^2 + 2iu_{2k}\frac{x_{2k-1}}{r_k}\right) \\
&= \sum_{k=1}^{r} r_k \left(u_{2k-1} - ix_{2k}/r_k\right)^2 + \frac{x_{2k}^2}{r_k} + r_k\left(u_{2k} + ix_{2k-1}/r_k\right)^2 + \frac{x_{2k-1}^2}{r_k}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\hat{W}_\kappa\left[|\mu\rangle\langle\mu|\right](x) &= e^{-\frac{\kappa}{2}\|x\|^2} \det R \left(\frac{\kappa}{\pi}\right)^r \prod_{k=1}^{r} e^{-\frac{\kappa}{r_k}\left(x_{2k-1}^2 + x_{2k}^2\right)} \cdot \\
&\qquad \cdot \int_{\mathbb{R}^2} e^{-\kappa r_k (u_{2k-1} - ix_{2k}/r_k)^2 - \kappa r_k (u_{2k} + ix_{2k-1}/r_k)^2}\, du_{2k-1}\, du_{2k} \\
&= e^{-\frac{\kappa}{2}\|x\|^2} \det R \left(\frac{\kappa}{\pi}\right)^r \prod_{k=1}^{r} e^{-\frac{\kappa}{r_k}\left(x_{2k-1}^2 + x_{2k}^2\right)} \frac{\pi}{\kappa r_k} \\
&= e^{-\frac{\kappa}{2}\|x\|^2 - \kappa\left\langle R^{-1}x, x\right\rangle} = e^{-\frac{\kappa}{2}\|x\|^2 - \frac{\kappa}{2}\langle (Q-I)x, x\rangle} \\
&= e^{-\frac{\kappa}{2}\langle Qx, x\rangle}.
\end{aligned}$$

That is,

$$|\mu\rangle\langle\mu| = \mathcal{D}\varrho_\alpha.$$

Compare this example with Example A.25.

A.5 Powers of quasi-free states

Let ϱ be the quasi-free state corresponding to α, determined by

$$\varrho(W(x)) = e^{-\frac{1}{4}\alpha(x,x)}\,.$$

Then, $H = \oplus_{k=1}^{d} H_k$, $H_k := \mathrm{span}\{e_{2k-1}, e_{2k}\}$ and $H_{\mathbb{C}} = \oplus_{k=1}^{d} (H_k)_{\mathbb{C}}$ with $(H_k)_{\mathbb{C}} \cong \mathbb{C}$, and

$$\mathcal{F}_{\mathrm{sym}}(H_{\mathbb{C}}) \cong \otimes_{k=1}^{d} \mathcal{F}_{\mathrm{sym}}(\mathbb{C})\,, \qquad CCR(H,\sigma) \cong \otimes_{k=1}^{d} CCR(H_k, \sigma_k)\,, \qquad \varrho \cong \otimes_{k=1}^{d} \varrho_k\,,$$

with

$$\varrho_k(W(x)) = e^{-\frac{1}{4}a_k|x|^2}\,, \qquad x \in \mathbb{C}\,.$$

Let $\{|n\rangle : n \in \mathbb{N}\}$ denote the standard basis of $\mathcal{F}_{\rm sym}(\mathbb{C}) = l^2(\mathbb{N})$, and let $y_F := \oplus_{n=0}^{\infty} \frac{1}{\sqrt{n!}} y^n |n\rangle$, $y \in \mathbb{C}$. From this,

$$\langle z_F, |n\rangle\langle n| \, y_F\rangle = \begin{cases} \frac{1}{n!}(\overline{z}y)^n, & n \geq 1 \\ 1, & n = 0\,. \end{cases} \tag{A.202}$$

In this representation,

$$W(x)y_f = e^{-\frac{1}{4}|x|^2 - \frac{1}{\sqrt{2}}\overline{x}y} \left(y + \frac{1}{\sqrt{2}}x\right)_F, \qquad x, y \in \mathbb{C}\,.$$

Now, by Theorem in [?, Chapter V, Corollary 3.2], the density $\mathcal{D}\varrho_k$ of ϱ_k is given by

$$\begin{aligned} \langle z_F, \mathcal{D}\varrho_k y_F\rangle &= \frac{1}{2\pi}\int_{\mathbb{C}} \langle z_F, W(x)y_F\rangle \, e^{-\frac{1}{4}a_k|x|^2}\, dx \\ &= \frac{1}{2\pi}\int_{\mathbb{C}} e^{-\frac{1}{4}|x|^2 - \frac{1}{\sqrt{2}}\overline{x}y - \frac{1}{4}a_k|x|^2 + \overline{z}(y + \frac{1}{\sqrt{2}}x)}\, dx. \end{aligned}$$

Now, with $x = x_1 + ix_2$, $x_1, x_2 \in \mathbb{R}$, in the exponent we have

$$\begin{aligned} &-\frac{1+a_k}{4}(x_1^2 + x_2^2) - \frac{1}{\sqrt{2}}(x_1 - ix_2)y + \overline{z}y + \frac{1}{\sqrt{2}}(x_1 + ix_2)\overline{z} \\ &= -\frac{1+a_k}{4}x_1^2 - \frac{1}{\sqrt{2}}x_1[y - \overline{z}] - \frac{1+a_k}{4}x_2^2 + \frac{i}{\sqrt{2}}x_2[y + \overline{z}] + \overline{z}y \\ &= -\frac{1+a_k}{4}\left[x_1^2 + \frac{4}{\sqrt{2}(1+a_k)}x_1[y - \overline{z}]\right] - \frac{1+a_k}{4}\left[x_2^2 - \frac{4i}{\sqrt{2}(1+a_k)}x_2[y + \overline{z}]\right] + \overline{z}y \\ &= -\frac{1+a_k}{4}\left[x_1^2 + 2x_1w_1\right] - \frac{1+a_k}{4}\left[x_2^2 - 2x_2w_2\right] + \overline{z}y \\ &= -\frac{1+a_k}{4}(x_1 + w_1)^2 - \frac{1+a_k}{4}(x_2 - w_2)^2 + \frac{1+a_k}{4}(w_1^2 + w_2^2) + \overline{z}y\,, \end{aligned}$$

with

$$w_1 := \frac{\sqrt{2}}{1+a_k}[y - \overline{z}]\,, \qquad w_2 := \frac{\sqrt{2}i}{1+a_k}[y + \overline{z}]\,,$$

which gives

$$w_1^2 + w_2^2 = -\frac{8}{(1+a_k)^2}y\overline{z}\,, \quad \text{and hence} \quad \frac{1+a_k}{4}(w_1^2 + w_2^2) + \overline{z}y = \frac{a_k - 1}{a_k + 1}y\overline{z}\,.$$

Thus,

$$\begin{aligned} \langle z_F, \mathcal{D}\varrho_k y_F\rangle &= \frac{1}{2\pi}\int_{\mathbb{C}} \exp\left(-\frac{1+a_k}{4}(x_1^2 + w_1)^2 - \frac{1+a_k}{4}(x_2^2 - w_2)^2 + \frac{a_k - 1}{a_k + 1}y\overline{z}\right) dx_1\, dx_2 \\ &= \exp\left(\frac{a_k - 1}{a_k + 1}y\overline{z}\right)\frac{2}{1+a_k}\,. \end{aligned}$$

By (A.202), if $zy \neq 0$ then

$$\frac{1}{n!} = \left\langle z_F, \frac{1}{(\overline{z}y)^n} \left|n\right\rangle\!\left\langle n\right| y_F \right\rangle ,$$

and thus for $y, z \neq 0$

$$\begin{aligned} \langle z_F, \mathcal{D}\varrho_k y_F \rangle &= \frac{2}{1+a_k} \sum_{n=0}^{\infty} \left(\frac{a_k - 1}{a_k + 1} y\overline{z} \right)^n \left\langle z_F, \frac{1}{(\overline{z}y)^n} \left|n\right\rangle\!\left\langle n\right| y_F \right\rangle \\ &= \left\langle z_F, \frac{2}{1+a_k} \sum_{n=0}^{\infty} \left(\frac{a_k - 1}{a_k + 1} \right)^n \left|n\right\rangle\!\left\langle n\right| y_F \right\rangle , \end{aligned}$$

and this formula holds also when $yz = 0$. Hence,

$$\mathcal{D}\varrho_k = \frac{2}{1+a_k} \sum_{n=0}^{\infty} \left(\frac{a_k - 1}{a_k + 1} \right)^n \left|n\right\rangle\!\left\langle n\right| .$$

Now, let $f(a) := \frac{a-1}{a+1}$, $a \geq 1$, with inverse $g(x) = \frac{1+x}{1-x}$, $0 \leq x \leq 1$, and introduce the notation

$$\mathcal{D}\varrho(a) := \frac{2}{1+a} \sum_{n=0}^{\infty} \left(\frac{a-1}{a+1} \right)^n \left|n\right\rangle\!\left\langle n\right| = (1 - f(a)) \sum_{n=0}^{\infty} f(a)^n \left|n\right\rangle\!\left\langle n\right| .$$

For $a \geq 1$ and $t \in \mathbb{R}$, let $a(t) := g\left(f(a)^t\right) = \frac{1+x^t}{1-x^t}$, where $x = \frac{a-1}{a+1}$. That is, $f(a(t)) = f(a)^t$, or $\frac{a(t)-1}{a(t)+1} = \left(\frac{a-1}{a+1}\right)^t$. Then,

$$\begin{aligned} \mathcal{D}\varrho(a)^t &= (1 - f(a))^t \sum_{n=0}^{\infty} \left(f(a)^t\right)^n \left|n\right\rangle\!\left\langle n\right| \\ &= \left(\frac{2}{1+a} \right)^t \frac{1+a(t)}{2} \frac{2}{1+a(t)} \sum_{n=0}^{\infty} f(a(t))^n \left|n\right\rangle\!\left\langle n\right| \\ &= \frac{2^t}{(a+1)^t - (a-1)^t} \mathcal{D}\varrho(a(t)) . \end{aligned}$$

□

A.6 Gaussian channels

Example A.25. Let (H, σ) be a finite-dimensional symplectic space and γ be an inner product on it. Choose an arbitrary irreducible representation of CCR $(() H, \kappa, \sigma)$

on some Hilbert space $\mathcal{H}$ and define the following operation on density operators of $\mathcal{H}$:

$$\Phi : \mathcal{D}\varrho \mapsto \int \frac{\sqrt{\det \gamma}}{\pi^d} e^{-\gamma(y,y)} W_\kappa(y)\, \mathcal{D}\varrho\, W_\kappa(y)^*\, d\lambda(x).$$

That is, Φ represents a random unitary rotation, distributed according to a normal distribution on the Weyl unitaries. Obviously, Φ is positive and trace-preserving. Moreover, since unitary rotations are completely positive, Φ is actually a quantum stochastic map or a channel.

Now let us see how the characteristic function of a density $\mathcal{D}\varrho$ changes under the map Φ:

$$\begin{aligned}
\hat{W}_\kappa[\Phi(\mathcal{D}\varrho)](x) &= \operatorname{Tr} \Phi(\mathcal{D}\varrho) W(x) \\
&= \int \frac{\sqrt{\det \gamma}}{\pi^d} e^{-\gamma(y,y)} \operatorname{Tr} W_\kappa(y)\, \mathcal{D}\varrho\, W_\kappa(y)^* W(x)\, d\lambda(y) \\
&= \int \frac{\sqrt{\det \gamma}}{\pi^d} e^{-\gamma(y,y)} \operatorname{Tr} \mathcal{D}\varrho W_\kappa(y)^* W_\kappa(x) W_\kappa(y)\, d\lambda(y) \\
&= \int \frac{\sqrt{\det \gamma}}{\pi^d} e^{-\gamma(y,y)} e^{2i\kappa\sigma(y,x)} \operatorname{Tr} \mathcal{D}\varrho W_\kappa(x)\, d\lambda(y) \\
&= \hat{W}_\kappa[\mathcal{D}\varrho] \int \frac{\sqrt{\det \gamma}}{\pi^d} e^{-\gamma(y,y)} e^{2i\kappa\sigma(y,x)}\, d\lambda(y).
\end{aligned}$$

Writing out every vector in a γ-canonical basis, we get

$$\begin{aligned}
\gamma(y,y) - 2i\kappa\sigma(y,x) &= \sum_{k=1}^{d} a_k \left(y_{2k-1}^2 + y_{2k}^2\right) - 2i\kappa \left(y_{2k-1} x_{2k} - y_{2k} x_{2k-1}\right) \\
&= \sum_{k=1}^{d} a_k \left(y_{2k-1}^2 - \frac{2i\kappa x_{2k}}{a_k} y_{2k-1}\right) + a_k \left(y_{2k}^2 + \frac{2i\kappa x_{2k-1}}{a_k} y_{2k}\right) \\
&= \sum_{k=1}^{d} a_k \left(y_{2k-1} - \frac{i\kappa x_{2k}}{a_k}\right)^2 + \frac{\kappa^2 x_{2k}^2}{a_k} + a_k \left(y_{2k} + \frac{i\kappa x_{2k-1}}{a_k}\right)^2 + \frac{\kappa^2 x_{2k-1}^2}{a_k}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\hat{W}_\kappa[\Phi(\mathcal{D}\varrho)](x) &= W_\kappa[\mathcal{D}\varrho]\frac{\sqrt{\det\gamma}}{\pi^d}\prod_{k=1}^{d} e^{-\frac{\kappa^2}{a_k}\left(x_{2k-1}^2+x_{2k}^2\right)}\,. \\
&\quad \cdot\int \exp\left(-a_k\left(y_{2k-1}^2+\frac{i\kappa x_{2k}}{a_k}\right)^2 - a_k\left(y_{2k}^2-\frac{ix_{2k-1}}{a_k}\right)^2\right)\,d\lambda(y) \\
&= W_\kappa[\mathcal{D}\varrho]\frac{\sqrt{\det\gamma}}{\pi^d}e^{-\kappa^2\gamma^{-1}(x,x)}\prod_{k=1}^{d}\frac{\pi}{a_k} \\
&= W_\kappa[\mathcal{D}\varrho](x)e^{-\kappa^2\gamma^{-1}(x,x)}.
\end{aligned}$$

Now, if $\varrho=\varrho_{\alpha,y}$ is a Gaussian state, then

$$\begin{aligned}
\hat{W}_\kappa[\Phi(\mathcal{D}\varrho)](x) &= e^{-\frac{|\kappa|}{2}\alpha(x,x)+2i\kappa\sigma(y,x)}e^{-\kappa^2\gamma^{-1}(x,x)} \\
&= e^{-\frac{|\kappa|}{2}\left(\alpha(x,x)+2|\kappa|\gamma^{-1}(x,x)\right)+2i\kappa\sigma(y,x)},
\end{aligned}$$

and hence $\Phi\left(\mathcal{D}\varrho_{\alpha,y}\right)$ is the density of a Gaussian state with

$$\Phi\left(\mathcal{D}\varrho_{\alpha,y}\right)=\mathcal{D}\varrho_{\alpha+2|\kappa|\gamma^{-1},y}.$$

Note that since $\alpha+i\sigma\geq 0$ and $\gamma\geq 0$, we also have $\alpha+2|\kappa|\gamma^{-1}+i\sigma\geq 0$.

The following has been shown in [**?**]:

Theorem A.26. The output p-norm of the above channel is multiplicative for integer p's.

A The Jordan measure

A.1 The Jordan measure on $\mathbb{R}^d$

The set of d-dimensional boxes is defined as

$$\mathrm{Box}(\mathbb{R}^d) := \left\{ \times_{i=1}^d [a_i, b_i) : a_i, b_i \in \mathbb{R} \right\},$$

on which we may introduce the volume function

$$\mathrm{Vol}\left(\times_{i=1}^d [a_i, b_i) \right) := \prod_{i=1}^d (b_i - a_i).$$

To define the volume of more general subsets of $\mathbb{R}^d$, we may try to approximate them by boxes, as follows.

Definition A.1. The *Jordan outer measure* $\mathrm{Vol}_J^* : \mathcal{P}(\mathbb{R}^d) \to [0, +\infty]$ is defined as

$$\mathrm{Vol}_J^*(A) := \inf \left\{ \sum_{i=1}^r \mathrm{Vol}(T_i) : T_i \in \mathrm{Box}(\mathbb{R}^d),\, i \in [r],\, A \subseteq \cup_{i=1}^r T_i,\, r \in \mathbb{N} \right\}, \tag{A.203}$$

for any $A \subseteq \mathbb{R}^d$.

Remark A.2. Note that if $A \subseteq \mathbb{R}^d$ is unbounded then it is not possible to cover it with finitely many boxes, and hence in the definition (A.203) of its outer measure, we are taking the infimum over the empty set, which is, by definition, $+\infty$. That is, the Jordan outer measure of any unbounded set is $+\infty$.

Definition A.3. The *Jordan inner measure* $\mathrm{Vol}_{*,J} : \mathcal{P}(\mathbb{R}^d) \to [0, +\infty]$ is defined as

$$\mathrm{Vol}_{*,J}(A) := \sup \left\{ \sum_{i=1}^r \mathrm{Vol}(T_i) : T_i \in \mathrm{Box}(\mathbb{R}^d),\, i \in [r],\, \cup_{i=1}^r T_i \subseteq A,\, r \in \mathbb{N} \right\}, \tag{A.204}$$

for any $A \subseteq \mathbb{R}^d$.

Definition A.4. A set $A \subseteq \mathbb{R}^d$ is *Jordan measurable* if $\mathrm{Vol}_J^*(A) = \mathrm{Vol}_{*,J}(A)$, and in this case this common value is called its *Jordan measure*, denoted by $\mathrm{Vol}_J(A)$.

Example A.5. Let $A = \{x\}$ consist of one single point $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$; then clearly $\mathrm{Vol}_{*,J}(\{\mathbf{x}\}) = 0$. On the other hand, for every $\varepsilon > 0$, we can choose $T_\varepsilon := \times_{i=1}^d [x_i - \varepsilon, x_i + \varepsilon) \in \mathrm{Box}(\mathbb{R}^d)$ so that $\{\mathbf{x}\} \subseteq T_\varepsilon$ for every $\varepsilon > 0$, and thus

$$\mathrm{Vol}_J^*(\{\mathbf{x}\}) \leq \inf_{\varepsilon > 0} \mathrm{Vol}(T_\varepsilon) = 0$$

Hence, $\{\mathbf{x}\}$ is Jordan measurable, and its Jordan measure is $\mathrm{Vol}_J(\{\mathbf{x}\}) = 0$.

Exercise A.6. Use the definitions to show that any subset of $\mathbb{R}^d$ consisting of finitely many points is Jordan measurable, and its Jordan measure is 0.

The following properties of the Jordan measure(s) are easy to verify. We will consider these properties in a more general setting in Section A.2.

Proposition A.7.

(i) The outer Jordan measure of a set is at least as large as its inner Jordan measure, i.e.,

$$\mathrm{Vol}_{*,J}(A) \le \mathrm{Vol}^*_J(A), \qquad A \subseteq \mathbb{R}^d. \tag{A.205}$$

(ii) The Jordan measure is an extension of the volume function, i.e., every box $T \in \mathrm{Box}(\mathbb{R}^d)$ is Jordan measurable, and $\mathrm{Vol}_J(T) = \mathrm{Vol}(T)$.

(iii) (**Monotonicity**) Both the inner and the outer Jordan measures are monotonic, i.e., for $A, B \subseteq \mathbb{R}^d$,

$$A \subseteq B \implies \mathrm{Vol}_{*,J}(A) \le \mathrm{Vol}_{*,J}(B), \quad \mathrm{Vol}^*_J(A) \le \mathrm{Vol}^*_J(B),$$

and if both A and B are measurable then also $\mathrm{Vol}_J(A) \le \mathrm{Vol}_J(B)$, i.e., the Jordan measure is also monotonic.

(iv) Both the empty set and $\mathbb{R}^d$ are Jordan measurable, and

$$\mathrm{Vol}_J(\emptyset) = 0, \qquad \mathrm{Vol}_J(\mathbb{R}^d) = +\infty.$$

(v) (**Finite additivity**) The Jordan measure is finitely additive in the following sense: If $A_1, \ldots, A_r$ are disjoint Jordan measurable subsets of $\mathbb{R}^d$ then $A = \cup_{i=1}^r A_i$ is also Jordan measurable, and

$$\mathrm{Vol}_J(A) = \sum_{i=1}^r \mathrm{Vol}_J(A_i).$$

Although the Jordan measure has a very appealing geometric picture behind its construction, its properties are not good enough for the purposes of analysis and probability theory. Below we outline some of these problems.

Probably the biggest problem with the Jordan measure is that it is not countably additive in the sense that a countable union of measurable sets is not necessarily measurable. Even worse, especially for the purposes of probability theory, is that a

countable union of disjoint sets all with zero Jordan measure may not be measurable. Indeed, let $A_n := \{n\}$ for every $n \in \mathbb{N}$. Then

$$\mathrm{Vol}_J(\{n\}) = 0 \quad \forall n \in \mathbb{N}, \quad \text{but} \quad \mathrm{Vol}_{*,J}(\mathbb{N}) = 0, \quad \text{while} \quad \mathrm{Vol}^*_J(\mathbb{N}) = +\infty, \tag{A.206}$$

and hence $\mathbb{N} = \cup_{n\in\mathbb{N}}\{n\}$ is not measurable. One might think that the problem here is the unboundedness of the union, but this is not the case. Indeed, $A := \mathbb{Q} \cap [0,1]$ is a countable union of single-point sets, all with measure 0, but again it is easy to see that

$$\mathrm{Vol}_{*,J}(\mathbb{Q} \cap [0,1]) = 0 \quad \text{and} \quad \mathrm{Vol}^*_J(\mathbb{Q} \cap [0,1]) = 1, \tag{A.207}$$

and hence $\mathbb{Q} \cap [0,1]$ is not Jordan measurable.

A closely related problem is that the Jordan measure is not well connected to the topology of $\mathbb{R}^d$ in the sense that even sets with the simplest topological structure (e.g., open sets, compact sets) may not be Jordan measurable. To see this, consider the fat Cantor sets $C_{a,q}$, constructed the following way. Let $a > 0$ and $0 < q < 1$ be such that $1 > \sum_{m\in\mathbb{N}} 2^m a q^m = \frac{2aq}{1-2q}$. (There are continuum many such pairs; e.g., $a = 1$ and $q < 1/4$.) Remove from the unit interval the open interval of length aq centered in the middle of $[0,1]$, leaving $C^{(1)}_{a,q}$, which is the union of two disjoint closed intervals of equal length: $C^{(1)}_{a,q} = [0, 1/2 - aq/2] \cup [1/2 + aq/2, 1]$. Next, remove from each of these two intervals an open interval of length aq^2 centered in their middle, leaving four disjoint closed intervals of equal length, which we denote by $C^{(2)}_{a,q}$. Continuing this procedure, we get for every $n \in \mathbb{N}$ a set $C^{(n)}_{a,q}$, consisting of 2^n disjoint compact intervals of equal length, so that $C^{(1)}_{a,q} \supseteq C^{(2)}_{a,q} \supseteq \ldots$. Let us define

$$C_{a,q} := \bigcap_{n\in\mathbb{N}} C^{(n)}_{a,q}.$$

Now, it is clear that whatever non-trivial interval $[a,b)$ we take in $[0,1]$, there is an n_0 after which $C^{(n)}_{a,q}$ is the union of disjoint intervals all of length strictly smaller than $(b-a)$. Hence, $C_{a,q}$ does not contain any non-trivial interval, and therefore $\mathrm{Vol}_{*,J}(C_{a,q}) = 0$. It is also obvious that $\mathrm{Vol}_{*,J}([0,1] \setminus C_{a,q}) = \frac{2aq}{1-2q} < 1$.

Assume that $\mathrm{Vol}^*_J([0,1] \setminus C_{a,q}) < 1$, i.e., that $[0,1] \setminus C_{a,q}$ can be covered with finitely many intervals $[a_i, b_i)$, $i \in [r]$, such that $\sum_{i=1}^r (b_i - a_i) < 1$. It is easy to see that we can assume without loss of generality that all these intervals are disjoint, and we can order them so that $0 \le a_1 < b_1 \le a_2 < b_2 \le \ldots \le a_n < b_n \le 1$. Then the intervals $T_0 := [0, a_1)$, $T_i := [b_i, a_{i+1})$, $i = 1, \ldots, r-1$, and $T_r := [b_r, 1)$ are all contained in $C_{a,q}$, and hence $\mathrm{Vol}_{*,J}(C_{a,q}) \ge 1 - \sum_{i=1}^r (b_i - a_i) > 0$, a contradiction. Hence, $\mathrm{Vol}^*_J([0,1] \setminus C_{a,q}) = 1 > 0 = \mathrm{Vol}_{*,J}([0,1] \setminus C_{a,q})$, and therefore $[0,1] \setminus C_{a,q}$, an open set, is not Jordan measurable.

An exactly analogous argument shows that $\mathrm{Vol}_J^*(C_{a,q}) > 0 = \mathrm{Vol}_{*,J}(C_{a,q})$, and hence $C_{a,q}$, a compact set, is not Jordan measurable, either.

One thing that one can notice in the examples above is that restricting the number of boxes to be finite in the definition of the outer Jordan measure is a serious limitation. Hence, it is natural to consider the following generalization:

Definition A.8. The *Lebesgue outer measure* $\mathrm{Vol}^* : \mathcal{P}(\mathbb{R}^d) \to [0, +\infty]$ is defined as

$$\mathrm{Vol}^*(A) := \inf \left\{ \sum_{n\in\mathbb{N}} \mathrm{Vol}(T_n) : T_n \in \mathrm{Box}(\mathbb{R}^d),\, n \in \mathbb{N},\, A \subseteq \cup_{n\in\mathbb{N}} T_n, \right\}, \qquad A \subseteq \mathbb{R}^d. \tag{A.208}$$

One could also modify the definition of the inner measure to allow infinitely many disjoint boxes. However, it is easy to see that this would make no change:

Exercise A.9. Show that for any $A \subseteq \mathbb{R}^d$,

$$\mathrm{Vol}_{*,J}(A) = \sup \left\{ \sum_{n\in\mathbb{N}} \mathrm{Vol}(T_n) : T_n \in \mathrm{Box}(\mathbb{R}^d),\, n \in \mathbb{N},\, \cup_{n\in\mathbb{N}} T_n \subseteq A \right\}.$$

Following the same logic as for the construction of the Jordan measure, the next logical step would be to call a set $A \subseteq \mathbb{R}^d$ measurable if $\mathrm{Vol}_{*,J}(A) = \mathrm{Vol}^*(A)$, and define its measure to be this common value. This would fix the problem in some of the above examples. Indeed, it is easy to see that with this modified definition every set consisting of countably many points is measurable and has zero measure; compare with the examples in (A.206) and (A.207). It is also easy to see that with this definition, $[0,1]\backslash C_{a,q}$, the complement of the fat Cantor set, becomes measurable, and its measure is the intuitively expected value $\frac{2aq}{1-2q}$.

On the other hand, the inner measure of $C_{a,q}$ is still zero, and hence, even if $C_{a,q}$ was measurable, its measure would be zero, which is not what we expect intuitively (we would expect $1-\frac{2aq}{1-2q}$). Phrasing it more mathematically, if $C_{a,q}$ was measurable, its measure would be zero, and hence the measure of $C_{a,q}$ and the measure of $[0,1] \setminus C_{a,q}$ would not sum up to 1, i.e., the measure would not be additive. (As it turns out later, the outer Lebesgue measure of $C_{a,q}$ is indeed $1 - \frac{2aq}{1-2q}$, as expected, and hence $C_{a,q}$ is not measurable according to the above definition.)

Thus, while the above modification of the Jordan measure offers some improvement, it does not remove all the drawbacks of the Jordan measure. Interestingly, the way out turns out to be to forget about the inner measure, and define the measurability of sets, and measure their volume, solely in terms of the outer Lebesgue measure Vol^*.

A.2 Generalized Jordan measures

The above described procedure to extend the volume function from boxes to more general sets can be carried out more generally for any additive set function α on a semi-ring $\mathcal{S}$, which we outline below.

Definition A.10. Let α be an additive set function on a semi-ring $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$. Its *outer Jordan α-measure* α_J^* and *inner Jordan α-measure* $\alpha_{*,J}$ are defined as

$$\alpha_J^*(A) := \inf\left\{\sum_{i=1}^{r} \mathrm{Vol}(S_i) : S_i \in \mathcal{S},\, i \in [r],\, A \subseteq \cup_{i=1}^{r} S_i,\, r \in \mathbb{N}\right\}, \tag{A.209}$$

$$\alpha_{*,J}(A) := \sup\left\{\sum_{i=1}^{r} \mathrm{Vol}(S_i) : S_i \in \mathcal{S},\, i \in [r],\, \cup_{i=1}^{r} S_i \subseteq A,\, r \in \mathbb{N}\right\}, \tag{A.210}$$

for any $A \subseteq \mathcal{X}$. We say that a set $A \subseteq \mathcal{X}$ is *Jordan α-measurable* if $\alpha_J^*(A) = \alpha_{*,J}(A)$, and in this case this common value is called its *Jordan α-measure*, denoted by $\alpha_J(A)$.

Proposition A.11. Let α be an additive set function on a semi-ring $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$.

(i) The outer Jordan α-measure of a set is at least as large as its inner Jordan α-measure, i.e.,

$$\alpha_{*,J}(A) \le \alpha_J^*(A), \qquad A \subseteq \mathcal{X}. \tag{A.211}$$

(ii) The Jordan α-measure is an extension of α, i.e., every $S \in \mathcal{S}$ is Jordan α-measurable, and $\alpha_J(S) = \alpha(S)$.

(iii) (**Monotonicity**) Both the inner and the outer Jordan measures are monotonic, i.e., for $A, B \subseteq \mathcal{X}$,

$$A \subseteq B \implies \alpha_{*,J}(A) \le \alpha_{*,J}(B), \quad \alpha_J^*(A) \le \alpha_J^*(B),$$

and if both A and B are measurable then also $\alpha_J(A) \le \alpha_J(B)$, i.e., the Jordan α-measure is also monotonic.

(iv) (**Finite additivity**) The Jordan α-measure is finitely additive in the following sense: If $A_1, \ldots, A_r$ are disjoint Jordan α-measurable subsets of X then $A = \uplus_{i=1}^{r} A_i$ is also Jordan α-measurable, and

$$\alpha_J(A) = \sum_{i=1}^{r} \alpha_J(A_i).$$

Exercise A.12. Prove properties (i)–(iii) in Proposition A.11.

Solution: Hidden.

Exercise A.13. Prove that the Jordan α-measure is an extension of α, (i.e., (ii) of Proposition A.11), using Corollary 3.50.

Solution: Hidden.

Exercise A.14. Prove the finite additivity of the Jordan α-measure ((iv) of Proposition A.11).

Solution: Hidden.

Exercise A.15. The definitions of the inner and the outer Jordan measures seem asymmetric, as the boxes are required to be disjoint in the definition of the inner Jordan measure, but not in the definition of the outer Jordan measure.

(i) Show that the Jordan outer measure would not change if we required disjointness of the boxes in its definition.

(ii) What would happen if we did not require the disjointness of the boxes in the definition of the inner Jordan measure? What would be the inner Jordan measure of a non-trivial bounded interval in the real line?

B Symplectic spaces

B.1 Bilinear forms

Let H be a real vector space with dual space H^*. For a base $\{e_i : i \in I\}$, let $\{e_1^* : i \in I\}$ denote the dual system, i.e., $e_i^* \in H^*$, $i \in I$ and $e_i^*(e_j) = \delta_{i,j}$. The dual system forms a basis of H^* if and only if H is finite-dimensional (otherwise $\dim H^* > \dim H$). The set of bilinear forms $\mathrm{Bilin}(H, \mathbb{R})$ on H forms a real vector space, into which $H^* \otimes H^*$ is naturally embedded, through

$$\varphi_1 \otimes \varphi_2(x, y) := \varphi_1(x)\varphi_2(y)\,.$$

If H is finite-dimensional then the above embedding is also an isomorphism, otherwise $H^* \otimes H^*$ is a proper subspace of $\mathrm{Bilin}(H, \mathbb{R})$.

If $e_1, \ldots, e_d$ is a base in H then $e_i^* \otimes e_j^*$, $i, j = 1, \ldots, d$ is a base in $H^* \otimes H^*$, and hence any ω bilinear form can be expanded in this base. One can easily see that the expansion coefficients are given by $\omega(e_i, e_j)$, i.e.,

$$\omega = \sum_{i,j=1}^{d} \omega(e_i, e_j) e_i^* \otimes e_j^*\,.$$

The *matrix of* ω is $[\omega]_e$, with entries $([\omega]_e)_{ij} := \omega(e_i, e_j)$. If $f_1, \ldots, f_d$ is another basis then

$$[\omega]_f = [T]_e^{\mathrm{T}} [\omega]_e [T]_e\,, \tag{B.212}$$

where T is the linear transformation $Te_k := f_k$, and $[T]_e$ is its matrix in the basis $e_1, \ldots, e_d$, i.e., $([T]_e)_{ij} = e_i^*(Te_j) = e_i^*(f_j)$.

Any bilinear form ω induces a homomorphism $\omega^\flat$ from H to H^*, by

$$\omega^\flat : x \mapsto \omega(x, .)\,, \qquad x \in H\,, \qquad \text{i.e.,} \qquad \omega^\flat(x)y = \omega(x, y)\,.$$

One can easily see that the matrix of $\omega^\flat$ in the pair of bases $(e_1, \ldots, e_d)$, $(e_1^*, \ldots, e_d^*)$ coincides with $[\omega]_e^{\mathrm{T}}$. Indeed,

$$\begin{aligned} \omega^\flat(e_i)x &= \omega^\flat(e_i)(x) = \omega^\flat(e_i)\left(\sum_{j=1}^{d} e_j^*(x) e_j\right) = \sum_{j=1}^{d} e_j^*(x)\omega^\flat(e_i)(e_j) \\ &= \sum_{j=1}^{d} \omega(e_i, e_j) e_j^*(x) = \left(\sum_{j=1}^{d} \omega(e_i, e_j) e_j^*\right) x\,. \end{aligned}$$

The *rank* of ω is the rank of its matrix. By (B.212), the rank does not depend on the basis in which the matrix of ω is given. By the above, the rank of ω is equal to the rank of the linear map $\omega^\flat$, i.e., $\mathrm{rk}\,\omega = \dim \mathrm{ran}\,\omega^\flat$.

A bilinear form ω is called *non-degenerate*, if

$$\omega(x,y)=0 \ \ \forall y\in H \Longrightarrow x=0.$$

By the above, one can easily see that in a finite-dimensional H, the following are equivalent:

(i) ω is non-degenerate;

(ii) $\operatorname{rk}\omega=\dim H$;

(iii) $\omega^\flat$ is an isomorphism between H and H^*.

By (iii), if ω is non-degenerate then for every $\varphi\in H^*$ there exists a unique $x_\varphi\in H$ such that $\varphi(y)=\omega(x_\varphi,y)$ (namely, $x_\varphi=\left(\omega^\flat\right)^{-1}\varphi$). This is a finite-dimensional version (and generalization) of the Riesz representation theorem in Hilbert spaces.

If ω is non-degenerate then by the above, $\omega^\flat$ is invertible, and hence one can define

$$A_\alpha^{(\omega)}:=\left(\omega^\flat\right)^{-1}\circ\alpha^\flat$$

for any bilinear form α. By definition, $A_\alpha^{(\omega)}$ is a linear operator on H with the property

$$\alpha(x,y)=\omega\left(A_\alpha^{(\omega)}x,y\right), \quad x,y\in H.$$

Note that $A_\alpha^{(\omega)}$ is unique in the sense that if A is any other operator for which

$$\alpha(x,y)=\omega\left(Ax,y\right), \quad x,y\in H$$

holds then $A=A_\alpha^{(\omega)}$. Indeed,

$$0=\omega\left(A_\alpha^{(\omega)}x,y\right)-\omega\left(Ax,y\right)=\omega\left(A_\alpha^{(\omega)}x-Ax,y\right), \quad x,y\in H$$

implies $A_\alpha^{(\omega)}x-Ax=0$, $x\in H$, by the non-degeneracy of ω. If α is also non-degenerate then $A_\alpha^{(\omega)}$ is invertible, and

$$\left(A_\alpha^{(\omega)}\right)^{-1}=\left(\left(\omega^\flat\right)^{-1}\circ\alpha^\flat\right)^{-1}=\left(\alpha^\flat\right)^{-1}\circ\omega^\flat=A_\omega^{(\alpha)}, \qquad \text{i.e.,}$$
$$\alpha\left(\left(A_\alpha^{(\omega)}\right)^{-1}x,y\right)=\omega\left(x,y\right), \quad x,y\in H.$$

A bilinear form ω is *symmetric*, if

$$\omega(x,y)=\omega(y,x), \qquad x,y\in H$$

and *antisymmetric*, if

$$\omega(x,y) = -\omega(y,x)\,, \qquad x,y \in H\,.$$

A symmetric bilinear form α is an *inner product*, if it is strictly positive definite, i.e., $\alpha(x,x) > 0$ for all $x \neq 0$. Obviously, an inner product is non-degenerate. A linear transformation T is *(α-)orthogonal* if it preserves the inner product α, i.e., $\alpha(Tx,Ty) = \alpha(x,y)$, $x,y \in H$.

We use the term *symplectic form* for non-degenerate anti-symmetric forms. A linear transformation T is *(σ-)symplectic* if it preserves the symplectic form σ, i.e., $\sigma(Tx,Ty) = \sigma(x,y)$, $x,y \in H$.

If α is an inner product then for a linear operator S the bilinear form $\sigma(x,y) := \alpha(Sx,y)$ is antisymmetric if and only if

$$\alpha(Sx,y) = \alpha(x,(-S)y), \qquad x,y \in H,$$

i.e., $S^{\mathrm{T}} = -S$, with the transpose taken with respect to the inner product α, and σ is non-degenerate if and only if S is invertible.

If σ is symplectic then for a linear operator A the bilinear form $\alpha(x,y) := \sigma(Ax,y)$ is an inner product if and only if

$$\sigma(Ax,y) = -\sigma(x,Ay), \qquad 0 < \sigma(Ax,x), \qquad x,y \in H, x \neq 0\,.$$

Conversely, any such operator defines a positive definite symmetric bilinear form, and the correspondence between such operators and positive definite symmetric bilinear forms is a linear isomorphism.

Definition B.1. A pair (H,σ), where H is a real vector space and σ a symplectic form on it, is called a *symplectic space*.

Example B.2. Let $\mathcal{H}$ be a complex inner product space. Then,

$$\sigma_{\mathcal{H}}(x,y) := \mathfrak{Im}\,\langle x,y\rangle\,, \qquad x,y \in \mathcal{H}$$

defines a symplectic form, which we call the *standard symplectic form* of $\mathcal{H}$.

Example B.3. Let $H := \mathbb{R}^d \times \mathbb{R}^d$, and for $x,y \in \mathbb{R}^d$ let xy denote their standard inner product. Then,

$$\sigma_{\mathbb{R}^{2d}}\left((x^{(1)},x^{(2)}),(y^{(1)},y^{(2)})\right) := x^{(1)}y^{(2)} - x^{(2)}y^{(1)}, \qquad x^{(1)},x^{(2)},y^{(1)},y^{(2)} \in \mathbb{R}^d$$

defines a symplectic form, which we call the *standard symplectic form* of $\mathbb{R}^d \times \mathbb{R}^d$.

Definition B.4. Let (H_1, σ_1) and (H_2, σ_2) be symplectic spaces. A linear map $T : H_1 \to H_2$ is *symplectic* if

$$\sigma_2(Tx, Ty) = \sigma_1(x, y), \qquad x, y \in H_1.$$

Two symplectic spaces (H_1, σ_1) and (H_2, σ_2) are *isomorphic* to each other if there exists a symplectic isomorphism between them.

Note that the inverse of a symplectic isomorphism is again a symplectic isomorphism.

Example B.5. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be complex inner product spaces. Any isomorphism $V : \mathcal{H}_1 \to \mathcal{H}_2$ is also a symplectic map, as

$$\sigma_{\mathcal{H}_2}(Vx, Vy) = \mathfrak{Im}\,\langle Vx, Vy\rangle = \mathfrak{Im}\,\langle x, y\rangle = \sigma_{\mathcal{H}_2}(x, y).$$

Moreover, an isometry V is a symplectic isomorphism if and only if it is unitary.

Example B.6. Consider the symplectic spaces $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$ and $(\mathbb{C}^d, \sigma_{\mathbb{C}^d})$, and define the map

$$T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}^d, \qquad T : \left(x^{(1)}, x^{(2)}\right) \mapsto x^{(1)} + ix^{(2)}, \qquad x^{(1)}, x^{(2)} \in \mathbb{R}^d.$$

Then, T is a real linear isomorphism between $\mathbb{R}^d \times \mathbb{R}^d$ and $\mathbb{C}^d$ (with $\mathbb{C}^d$ considered with its real vector space structure), and a straightforward computation verifies that it is symplectic, too. Hence, the symplectic spaces $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$ and $(\mathbb{C}^d, \sigma_{\mathbb{C}^d})$ are isomorphic to each other. The inverse of T is given by

$$T^{-1} : \mathbb{C}^d \to \mathbb{R}^d \times \mathbb{R}^d, \qquad T^{-1} : x \mapsto \left(\mathfrak{Re}\, x, \mathfrak{Im}\, x\right), \qquad x \in \mathcal{H}.$$

Example B.7. Let $\mathcal{H}$ be a d-dimensional complex Hilbert space and $\sigma_{\mathcal{H}}$ be its standard symplectic form. We show that any orthonormal basis defines a symplectic isomorphism of $(\mathcal{H}, \sigma_{\mathcal{H}})$ with $(\mathbb{C}^d, \sigma_{\mathbb{C}^d})$, and therefore, by the above example, also with $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$. Indeed, if $e_1, \ldots, e_d$ is an arbitrary orthonormal basis then the coordinate map

$$T : \mathcal{H} \to \mathbb{C}^d, \qquad T : x \mapsto \left(\langle e_1, x\rangle, \ldots, \langle e_d, x\rangle\right), \qquad x \in \mathcal{H}$$

is easily seen to be a symplectic isomorphism. The corresponding symplectic isomorphism with $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$ is given by

$$S : x \mapsto \left(\left(\mathfrak{Re}\,\langle e_1, x\rangle, \ldots, \mathfrak{Re}\,\langle e_d, x\rangle\right), \left(\mathfrak{Im}\,\langle e_1, x\rangle, \ldots, \mathfrak{Im}\,\langle e_d, x\rangle\right)\right), \qquad x \in \mathcal{H}.$$

As the above examples already suggest, a symplectic space is uniquely determined by its dimension, up to symplectic isomorphisms. The analogous statement for inner product spaces can be proved using orthonormal bases, and in the next section we will prove the conjectured isomorphism of equal dimensional symplectic spaces using the concecpt of symplectic bases.

In general, one cannot define the determinant of a bilinear form if only a vector space structure is available. A heuristic reason for this is that the matrix of a bilinear form transforms as $[\alpha]_f = [S]_e^{\mathrm{T}}[\alpha]_e[S]_e$, where $Se_k := f_k$ is the basis change transformation, and hence the determinant of the matrix of α can be different in different bases. If, however, an extra structure, like a symplectic form is given then the situation is different.

Definition B.8. Let H be a finite-dimensional real vector space and ω be a non-degenerate bilinear form on it. The *ω-determinant* of a bilinear form α is defined as

$$\det_\omega \alpha := \det A_\alpha^{(\omega)} .$$

In particular, the *determinant* of a bilinear form α in a finite-dimensional symplectic space (H, σ) is defined as

$$\det \alpha := \det_\sigma \alpha = \det A_\alpha^{(\sigma)} .$$

B.2 Symplectic bases and symplectic transformations

Lemma B.9. Let H be a finite-dimensional real vector space and σ be an anti-symmetric form on it. Then, $\operatorname{rk}\sigma = 2n$ is even, and there exists a basis $e_1, \dots, e_d$ such that

$$\sigma(e_i, e_j) = \begin{cases} 1, & i = 2k-1, j = 2k,\ k = 1, \dots, n, \\ -1, & i = 2k, j = 2k-1,\ k = 1, \dots, n, \\ 0, & \text{otherwise.} \end{cases} \tag{B.213}$$

Proof. If $\sigma = 0$ then $n = 0$ and any basis does the job. Assume that $\sigma \neq 0$. Then there exist $x, y \in H$ such that $\sigma(x, y) \neq 0$. By possibly interchanging x and y, we can assume that $\sigma(x, y) > 0$. Define

$$e_1 := \frac{1}{\sqrt{\sigma(x,y)}} x, \qquad e_2 := \frac{1}{\sqrt{\sigma(x,y)}} y, \qquad \text{and} \qquad H_1 := \operatorname{span}\{e_1, e_2\} .$$

Then,

$$\sigma(e_1, e_2) = 1, \quad \sigma(e_2, e_1) = -1, \quad \text{and} \quad \sigma(e_1, e_1) = \sigma(e_2, e_2) = 0 .$$

Let

$$H_1^\perp := \{x : \sigma(x, e_1) = \sigma(x, e_1) = 0\}$$

be the σ-orthocomplement of H_1. If $x = x_1 e_1 + x_2 e_2 \in H_1 \cap H_1^\perp$, then

$$x_1 = \sigma(x, e_2) = 0, \quad x_2 = -\sigma(x, e_1) = 0, \qquad \text{i.e.,} \qquad x = 0\,.$$

Hence, $H_1 \cap H_1^\perp = \{0\}$. Now we can restrict σ to $H_1^\perp \times H_1^\perp$ and continue the above process. In each step we obtain a new 2-dimensional subspace, spanned by e_{2k-1}, e_{2k} in the kth step, with

$$\sigma(e_{2k-1}, e_{2k}) = 1, \qquad \sigma(e_{2k}, e_{2k-1}) = -1, \quad \sigma(e_{2k-1}, e_{2k-1}) = \sigma(e_{2k}, e_{2k}) = 0$$

and

$$\sigma(e_{2k-1}, e_m) = \sigma(e_{2k}, e_m) = 0, \quad m \leq 2k - 1\,.$$

Let $H_k := \mathrm{span}\{e_{2k-1}, e_{2k}\}$. The process stops after the nth step if σ restricted to $H_0 \times H_0$, $H_0 := (H_1 \oplus \ldots \oplus H_n)^\perp$ is the zero form, whence we can choose any basis in H_0 to complete the basis of H. Note that if $\dim H_0 = 1$ then then σ restricted to $H_0 \times H_0$ is necessary 0, as there is no non-zero symplectic form on a one-dimensional space. The matrix of σ in the so-obtained basis is

$$[\sigma] = \left(\oplus_{k=1}^n \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right) \oplus 0_{d-2n \times d-2n}\,,$$

which immediately yields the assertion about the rank. □

Corollary B.10. A symplectic space has even or infinite dimension. In a finite-dimensional symplectic space there exists a basis $e_1, \ldots, e_{\dim H}$ such that the matrix of σ in this basis is

$$[\sigma]_e = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Definition B.11. A basis satisfying (B.213) is called a *symplectic basis* of H (for σ).

Example B.12. Let $H = \mathbb{R}^d \times \mathbb{R}^d$ and σ be its standard symplectic form. Then,

$$e_{2k-1} := (\mathbf{1}_{\{k\}}, 0), \qquad e_{2k} := (0, \mathbf{1}_{\{k\}}), \qquad k = 1, \ldots, d$$

is a symplectic basis, where $\mathbf{1}_{\{k\}}$, $k = 1, \ldots, d$ is the standard basis of $\mathbb{R}^d$.

Example B.13. Let $\mathcal{H}$ be a finite-dimensional complex Hilbert space, and $\sigma(x,y) := \mathfrak{Im}\,\langle x,y\rangle$, $x,y\in\mathcal{H}$ be its standard symplectic form. If $e_1,\dots,e_{\dim H}$ is an orthonormal basis then

$$f_{2k-1} := e_k, \qquad f_{2k} := ie_k, \qquad k=1,\dots,\dim\mathcal{H}$$

is a symplectic basis. On the other hand, not every symplectic basis is of this form. For instance, with the above notations, let

$$f'_{2k-1} := e_k, \qquad f'_{2k} := ie_k + c_k e_k, \qquad k=1,\dots,\dim\mathcal{H}$$

with some $c_1,\dots,c_{\dim\mathcal{H}}\in\mathbb{R}$. Then, $f_1,\dots,f_{\dim H}$ is a symplectic basis, but $\|f'_{2k}\|\neq 1$ unless $c_k=0$.

Remark B.14. The above definition of a symplectic basis only works for finite-dimensional spaces. To include infinite-dimensional spaces, one can modify the definition the following way. Let $\dim H$ denote the dimension of H, which in general is a cardinality. Assume that $\dim H$ is either infinite or if finite then it is even, and hence $\dim H/2$ is again a cardinality. A basis $\{e_k, e'_k : i\in I\}$ with $|I| = \dim H/2$ is a symplectic basis if

$$\sigma(e_k,e'_k)=1, \qquad \sigma(e_k,e_l)=\sigma(e'_k,e'_l)=0, \qquad k,l\in I.$$

For instance, if $\mathcal{H}$ is a complex Hilbert space with its standard symplectic form, and $\{e_k : k\in I\}$ is an orthonormal basis then $f_k := e_k,\ f'_k := ie_k,\ k\in I$ is a symplectic basis.

Lemma B.15. The dual basis of a symplectic basis $e_1,\dots,e_{\dim H}$, is

$$e^*_{2k-1} = \sigma^\flat(-e_{2k}) = -\sigma(e_{2k},.), \qquad e^*_{2k} = \sigma^\flat(e_{2k-1}) = \sigma(e_{2k-1},.), \qquad k=1,\dots,\dim H.$$

The coordinate expansion of $x\in H$ in the symplectic basis $e_1,\dots,e_{\dim H}$ is given by

$$x = \sum_{k=1}^{d} -\sigma(e_{2k},x)e_{2k-1} + \sigma(e_{2k-1},x)e_{2k}\,, \qquad x\in H.$$

Proof. By the above, for $x=\sum_{j=1}^{\dim H} x_j e_j$ we have

$$\sigma(e_{2k-1},x)=x_{2k}, \qquad \sigma(e_{2k},x)=-x_{2k-1},$$

from which the assertions follow. □

Similarly, the matrix elements of a linear operator A are given by

$$([A]_e)_{2k-1,m} = -\sigma(e_{2k}, Ae_m), \qquad ([A]_e)_{2k,m} = \sigma(e_{2k-1}, Ae_m).$$

In particular, if α is the bilinear form determined by $\alpha(x,y) = \sigma(Ax,y)$ (i.e., $A = \left(\sigma^\flat\right)^{-1} \circ \alpha^\flat$) then

$$\begin{aligned} ([A]_e)_{2k-1,m} &= \sigma(Ae_m, e_{2k}) = \alpha(e_m, e_{2k}) = ([\alpha]_e)_{m,2k} \\ ([A]_e)_{2k,m} &= -\sigma(Ae_m, e_{2k-1}) = -\alpha(e_m, e_{2k-1}) = -([\alpha]_e)_{m,2k-1}\,. \end{aligned} \tag{B.214}$$

As a consequence, we have the following:

Corollary B.16. Let α be a bilinear form in a symplectic space (H,σ). Then,

$$\det\alpha = \det[\alpha]_e$$

for any symplectic basis $e_1, \ldots, e_{\dim H}$.

Proof. By definition, $\det\alpha = \det A$ for $A := A_\alpha^{(\sigma)}$, and by (B.214),

$$\det A = \det[A]_e = \det[\alpha]_e$$

in any symplectic basis $e_1, \ldots, e_{\dim H}$. □

It is easy to see that a linear map is symplectic if and only if it maps symplectic bases into symplectic bases. If S is a linear map then the matrix of the bilinear form $\sigma_S(x,y) := \sigma(Sx, Sy)$ is related to that of σ by

$$[\sigma_S] = [S]^{\mathrm{T}}[\sigma][S].$$

In particular, S is symplectic if and only if

$$[\sigma] = [S]^{\mathrm{T}}[\sigma][S]$$

in some (and hence any) basis. Choosing the basis to be symplectic, we get

$$1 = \det[\sigma] = \det\left([S]^{\mathrm{T}}[\sigma][S]\right) = \det\left([S]^{\mathrm{T}}\right)\det\left([S]\right) = \det\left(S\right)^2,$$

and hence the determinant of a symplectic transformation is 1 or -1. (We will see later that it is actually 1).

Note that in general the determinant of the matrix of a bilinear form α depends on the basis in which it is given, as the transformation rule (B.212) doesn't preserve the determinant. However, as symplectic transformations have unit determinant, the determinant of $[\alpha]$ is the same in any symplectic basis, as it also follows from Corollary B.16.

Lemma B.17. Let (H, σ) be a finite-dimensional symplectic space with $2d := \dim H$. Any symplectic basis defines a symplectic isomorphism with $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$ and another one with $(\mathbb{C}^d, \sigma_{\mathbb{C}^d})$.

Proof. Let $e_1, \dots, e_{2d}$ be a symplectic basis of H. One can check by a straightforward computation that the coordinate map

$$\begin{aligned} T: \quad x &\mapsto \left((-\sigma(e_2, x), \dots, -\sigma(e_{2d}, x)\right), \left(\sigma(e_1, x), \dots, \sigma(e_{2d-1}, x)\right)) \\ &= \left((x_1, \dots, x_{2d-1}), (x_2, \dots, x_{2d})\right), \qquad x = \textstyle\sum_{j=1}^{2d} x_j e_j \in H \end{aligned}$$

defines a symplectic isomorphism with $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$. Consequently,

$$S: x \mapsto (x_1 + ix_2, \dots, x_{2d-1} + ix_{2d}), \qquad x = \sum_{j=1}^{2d} x_j e_j \in H \tag{B.215}$$

is a symplectic isomorphism with $(\mathbb{C}^d, \sigma_{\mathbb{C}^d})$. Note that T and S are the real linear extensions of the maps

$$Te_m := \begin{cases} \left(\mathbf{1}_{\{k\}}, 0\right), & m = 2k-1, \\ \left(0, \mathbf{1}_{\{k\}}\right), & m = 2k, \end{cases} \qquad Se_m := \begin{cases} \mathbf{1}_{\{k\}}, & m = 2k-1, \\ i\mathbf{1}_{\{k\}}, & m = 2k, \end{cases}$$

respectively. □

Lemma B.18. Let (H, σ) be a finite-dimensional symplectic space with $2d := \dim H$. Any isomorphisms with $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$ or with $(\mathbb{C}^d, \sigma_{\mathbb{C}^d})$ arise the way described in Lemma B.17.

Proof. Let $T: H \to \mathbb{R}^d \times \mathbb{R}^d$ be a symplectic isomorphism, and let $\mathbf{1}_{\{1\}}, \dots, \mathbf{1}_{\{d\}}$ be the standard basis of $\mathbb{R}^d$. Obviously,

$$e_{2k-1} := T^{-1}\left(\mathbf{1}_{\{k\}}, 0\right), \qquad e_{2k} := T^{-1}\left(0, \mathbf{1}_{\{k\}}\right), \qquad k = 1, \dots, d$$

defines a symplectic basis in H, and T is easily seen to be the coordinate map corresponding to this basis.

Similarly, if $T: H \to \mathbb{C}^d$ is a symplectic isomorphism, and $\mathbf{1}_{\{1\}}, \dots, \mathbf{1}_{\{d\}}$ is the standard basis of $\mathbb{C}^d$, then

$$e_{2k-1} := T^{-1}\mathbf{1}_{\{k\}}, \qquad e_{2k} := T^{-1}\left(i\mathbf{1}_{\{k\}}\right), \qquad k = 1, \dots, d$$

defines a symplectic basis in H, and T is of the form given in (B.215). □

Remark B.19. Note that if $S, T : (H, \sigma) \to \left(\mathbb{C}^d, \sigma_{\mathbb{C}^d}\right)$ are symplectic isomorphisms then $S \circ T^{-1} : \mathbb{C}^d \to \mathbb{C}^d$ is also a symplectic isomorphism, but it is not a complex linear map in general. Indeed, let $(H, \sigma) := \left(\mathbb{C}^d, \sigma_{\mathbb{C}^d}\right)$ and T be the identity map. Let $e_1, \ldots, e_d$ by any orthonormal basis. Then, $f_{2k-1} := e_k,\ f_{2k} := e_k + ie_k,\ k = 1, \ldots, d$ is a symplectic basis, and

$$Sf_{2k-1} := e_k, \qquad Sf_{2k} := ie_k, \qquad k = 1, \ldots, d$$

defines a symplectic isomorphism, but $S \circ T^{-1} = S$ is not complex linear.

B.3 Complexification

Let H be a real vector space with finite dimension, define vector addition and multiplication by real scalars on $H \times H$ componentwise, and let $(x^{(1)}, x^{(2)}) := (-x^{(2)}, x^{(1)})$. One can easily see that the result is a complex vector space, which we denote by $H_{\mathbb{C}}$ and call the *standard complexification* of H. Note that H can be considered a real lineat subspace of $H_{\mathbb{C}}$ via $H \ni x \equiv (x, 0) \in H_{\mathbb{C}}$, and $ix \equiv i(x, 0) = (0, x)$, so that with this identification, $H_{\mathbb{C}} \ni (x^{(1)}, x^{(2)}) \equiv x^{(1)} + ix^{(2)}$, and therefore $H_{\mathbb{C}} \equiv H \oplus iH$. In the case of $H = \mathbb{R}^d$, $(\mathbb{R}^d)_{\mathbb{C}} \equiv \mathbb{R}^d \oplus i\mathbb{R}^d$ may be further identified with $\mathbb{C}^d$ via

$$(\mathbb{R}^d)_{\mathbb{C}} \ni (x^{(1)}, x^{(2)}) \equiv x^{(1)} + ix^{(2)} \in \mathbb{C}^d;$$

in particular, $\mathbb{R}_{\mathbb{C}} \equiv \mathbb{C}$.

If H, K are real vector spaces and $A : H \to K$ is a real linear map then

$$A_{\mathbb{C}} : H_{\mathbb{C}} \to K_{\mathbb{C}}, \qquad A_{\mathbb{C}}(x^{(1)}, x^{(2)}) := (Ax^{(1)}, Ax^{(2)})$$

defines a complex linear map. Indeed,

$$A_{\mathbb{C}} i(x^{(1)}, x^{(2)}) = A_{\mathbb{C}}(-x^{(2)}, x^{(1)}) = (-Ax^{(2)}, Ax^{(1)}) = i(Ax^{(1)}, Ax^{(2)}) = iA_{\mathbb{C}}(x^{(1)}, x^{(2)}).$$

Using the above identifications, the above can be rewritten as

$$A_{\mathbb{C}}(x^{(1)} + ix^{(2)}) = Ax^{(1)} + iAx^{(2)}.$$

In particular, for the complexification of a linear functional $\varphi \in H'$, we get

$$\varphi_{\mathbb{C}}\left(x^{(1)} + ix^{(2)}\right) = \varphi_{\mathbb{C}}((x^{(1)}, x^{(2)})) = (\varphi(x^{(1)}), \varphi(x^{(2)})) \equiv \varphi(x^{(1)}) + i\varphi(x^{(2)}).$$

Note that the standard inner product on $\mathbb{C}^d$ may be written as

$$\left\langle x^{(1)} + ix^{(2)}, y^{(1)} + iy^{(2)} \right\rangle_{\mathbb{C}^d} = \left\langle x^{(1)}, y^{(1)} \right\rangle_{\mathbb{R}^d} + i \left\langle x^{(1)}, y^{(2)} \right\rangle_{\mathbb{R}^d} - i \left\langle x^{(2)}, y^{(1)} \right\rangle_{\mathbb{R}^d} + \left\langle x^{(2)}, y^{(2)} \right\rangle_{\mathbb{R}^d},$$

where $x^{(k)}, y^{(k)} \in \mathbb{R}^d$, and $\langle ., . \rangle_{\mathbb{R}^d}$ is the standard inner product of $\mathbb{R}^d$. This motivates the extension of a real bilinear form $\omega : H \times H \to \mathbb{R}$ to a complex sesquilinear form $\omega_{\mathbb{C}} : H_{\mathbb{C}} \times H_{\mathbb{C}} \to \mathbb{C}$ as

$$\begin{aligned}\omega_{\mathbb{C}}\left((x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)})\right) &:= \omega(x^{(1)}, y^{(1)}) + i\omega(x^{(1)}, y^{(2)}) - i\omega(x^{(2)}, y^{(1)}) + \omega(x^{(2)}, y^{(2)}) \\ &= \left[\left(\omega^\flat(x^{(1)})\right)_{\mathbb{C}} - i\left(\omega^\flat(x^{(2)})\right)_{\mathbb{C}}\right]\left(y^{(1)} + iy^{(2)}\right), \quad \text{(B.216)}\end{aligned}$$

where the equality follows by a straightforward computation. Similarly to the real case, one can define

$$(\omega_{\mathbb{C}})^\flat : H_{\mathbb{C}} \to H_{\mathbb{C}}^*, \qquad (\omega_{\mathbb{C}})^\flat : x \mapsto \omega_{\mathbb{C}}(x, .), \qquad x \in H_{\mathbb{C}},$$

which is now a conjugate linear operator. By introducing the conjugate linear isomorphism

$$S : (H^*)_{\mathbb{C}} \to H_{\mathbb{C}}^*, \qquad S(\varphi_1, \varphi_2) := (\varphi_1)_{\mathbb{C}} - i(\varphi_2)_{\mathbb{C}},$$

we get

$$(\omega_{\mathbb{C}})^\flat = S \circ \left(\omega^\flat\right)_{\mathbb{C}}.$$

One can easily verify that $\omega_{\mathbb{C}}$ is non-degenerate if and only if ω is non-degenerate. If ω is non-degenerate then $(\omega_{\mathbb{C}})^\flat$ is bijective, and for any real bilinear form α on H,

$$A_{\alpha_{\mathbb{C}}}^{(\omega_{\mathbb{C}})} := \left(\omega_{\mathbb{C}}^\flat\right)^{-1} \circ \alpha_{\mathbb{C}}^\flat$$

is a complex linear map from $H_{\mathbb{C}}$ to $H_{\mathbb{C}}$, with

$$\alpha_{\mathbb{C}}(x, y) = \omega_{\mathbb{C}}\left(A_{\alpha_{\mathbb{C}}}^{(\omega_{\mathbb{C}})} x, y\right), \qquad x, y \in H_{\mathbb{C}}.$$

A straightforward computation shows that

$$A_{\alpha_{\mathbb{C}}}^{(\omega_{\mathbb{C}})} = \left(A_\alpha^{(\omega)}\right)_{\mathbb{C}},$$

that is,

$$\alpha_{\mathbb{C}}\left((x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)})\right) = \omega_{\mathbb{C}}\left((A_\alpha^{(\omega)} x^{(1)}, A_\alpha^{(\omega)} x^{(2)}), (y^{(1)}, y^{(2)})\right).$$

If α is a real inner product on H then $\alpha_{\mathbb{C}}$ is easily seen to be a complex inner product on $H_{\mathbb{C}}$, with induced norm $\left\|(x^{(1)}, x^{(2)})\right\|_\alpha^2 = \alpha(x^{(1)}, x^{(1)}) + \alpha(x^{(2)}, x^{(2)})$. Moreover, if $\{e_i\}_{i=1}^{\dim H}$ is an α-orthonormal basis in H then $\{(e_i, 0)\}_{i=1}^{\dim H}$ is an $\alpha_{\mathbb{C}}$-orthonormal basis in $H_{\mathbb{C}}$. Let A^{T} denote the transpose of A with respect to α,

defined by $\alpha(A^{\mathrm{T}}y, x) = \alpha(x, Ay)$, $x, y \in H$. For any $(x^{(1)}, x^{(2)})$, $(y^{(1)}, y^{(2)}) \in H_{\mathbb{C}}$ we have

$$\begin{aligned}
&\alpha_{\mathbb{C}}\left((x^{(1)}, x^{(2)}), A_{\mathbb{C}}(y^{(1)}, y^{(2)})\right) = \alpha_{\mathbb{C}}\left((x^{(1)}, x^{(2)}), (Ay^{(1)}, Ay^{(2)})\right) \\
&\quad = \alpha(x^{(1)}, Ay^{(1)}) + i\alpha(x^{(1)}, Ay^{(2)}) - i\alpha(x^{(2)}, Ay^{(1)}) + \alpha(x^{(2)}, Ay^{(2)}) \\
&\quad = \alpha(A^{\mathrm{T}}x^{(1)}, y^{(1)}) + i\alpha(A^{\mathrm{T}}x^{(1)}, y^{(2)}) - i\alpha(A^{\mathrm{T}}x^{(2)}, y^{(1)}) + \alpha(A^{\mathrm{T}}x^{(2)}, y^{(2)}) \\
&\quad = \alpha_{\mathbb{C}}\left((A^{\mathrm{T}}x^{(1)}, A^{\mathrm{T}}x^{(2)}), (y^{(1)}, y^{(2)})\right) \\
&\quad = \alpha_{\mathbb{C}}\left((A^{\mathrm{T}})_{\mathbb{C}}(x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)})\right) ,
\end{aligned}$$

and hence

$$A^*_{\mathbb{C}} = (A^{\mathrm{T}})_{\mathbb{C}} ,$$

where the adjoint $A^*_{\mathbb{C}}$ is taken with respect to the inner product $\alpha_{\mathbb{C}}$. In particular, if σ is a symplectic form then $\left(A^{(\alpha)}_{\sigma}\right)^{\mathrm{T}} = -A^{(\alpha)}_{\sigma}$, and hence

$$\left(A^{(\alpha)}_{\sigma}\right)^*_{\mathbb{C}} = -\left(A^{(\alpha)}_{\sigma}\right)_{\mathbb{C}} .$$

The real dimension of a complex vector space is even, and the above construction guarantees this necessary condition to be satisfied by doubling the original space. Hence, the complex dimension of the resulting complexification is the same as the real dimension of the original space. However, if the dimension of the original real vector space is even, one may follow a different way of complexification, that results in a complex vector space with half the dimension of the original real space.

Note that on a complex vector space H,

$$J : x \mapsto ix \quad \text{is real linear and} \quad J^2 = -I.$$

Definition B.20. A real linear map J on a real vector space H is a *complex structure* if

$$J^2 = -I.$$

One can easily see that if a complex structure J is given on a real vector space H then

$$ix := Jx$$

defines a multiplication between complex numbers and elements of H with respect to which H is a complex vector space.

Definition B.21. The resulting complex vector space is said to be the *J-complexification* of H and is denoted by H_J.

Note that different complex structures give rise to different complexifications. Indeed, if $J_1 \neq J_2$ then there exists some non-zero $x \in H$ for which $J_1 x \neq J_2 x$ and hence ix as meant in H_{J_1} is not equal to ix as meant in H_{J_2}.

Definition B.22. Let H be a real vector space and J be a complex structure on it. A real linear map $A : H \to H$ is *J-linear* if $AJ = JA$ holds.

Note that a real linear map is J-linear if and only if it is complex linear with respect to the J-complexification.

Example B.23. It is easy to see that $J(x, y) := (-y, x)$ is a complex structure on $H := \mathbb{R}^d \times \mathbb{R}^d$, and

$$\left(\mathbb{R}^d \times \mathbb{R}^d\right)_J = (\mathbb{R}^d)_{\mathbb{C}}.$$

Moreover, if A is a real linear map on $\mathbb{R}^d$ then

$$A(x^{(1)}, x^{(2)}) := \left(Ax^{(1)}, Ax^{(2)}\right) = A_{\mathbb{C}}(x^{(1)}, x^{(2)})$$

is J-linear.

Lemma B.24. A real vector space possesses a complex structure if and only if its dimension is even. To every complex structure there exists a basis $e_1, \ldots, e_{\dim H}$ in which

$$[J]_e = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

i.e.,

$$Je_{2k-1} = e_{2k}, \qquad Je_{2k} = -e_{2k-1}, \qquad k = 1, \ldots, \dim H. \tag{B.217}$$

Conversely, if $e_1, \ldots, e_{\dim H}$ is a basis then there exists a unique complex structure for which (B.217) holds.

Proof. As we have seen, a real vector space with a complex structure can be turned into a complex one, an hence its real dimension has to be even. On the other hand, let $e_1, \ldots, e_{\dim H}$ be a basis in an even-dimensional vector space and define J to be the unique real linear extension of

$$Je_{2k-1} := e_{2k}, \qquad Je_{2k} := -e_{2k-1}, \qquad k = 1, \ldots, \dim H.$$

One can easily verify that J is a complex structure.

Assume now that J is a complex structure on H, let $f_1, \dots, f_{\dim H/2}$ be a basis of H_J and define $e_{2k-1} := f_k,\ e_{2k} := Jf_k,\ k = 1, \dots, \dim H/2$. Every $x \in H$ can be uniquely written in the form

$$x = \sum_{k=1}^{\dim H/2} (\mathfrak{Re}\, \lambda_k + i\, \mathfrak{Im}\, \lambda_k)\, f_k = \sum_{k=1}^{\dim H/2} (\mathfrak{Re}\, \lambda_k)\, f_k + (\mathfrak{Im}\, \lambda_k)\, Jf_k, \qquad \lambda_1, \dots, \lambda_{\dim H/2} \in \mathbb{C},$$

and hence $e_1, \dots, e_{\dim H}$ is a basis. $Je_{2k-1} = e_{2k}$ follows by definition, and $Je_{2k} = -e_{2k-1}$ because $J^2 = -I$. □

Definition B.25. We say that a basis $e_1, \dots, e_{\dim H}$ is *J-canonical* if (B.217) holds.

Note that a J-canonical basis is not unique.

Example B.26. Let $H := \mathbb{R}^d \times \mathbb{R}^d$. Then,

$$J(x, y) := (-y, x)$$

defines a complex structure on H, and

$$e_{2k-1} := \left(\mathbf{1}_{\{k\}}, 0\right), \qquad e_{2k} := \left(0, \mathbf{1}_{\{k\}}\right),$$

$$e'_{2k-1} := \left(\mathbf{1}_{\{k\}}, \mathbf{1}_{\{k\}}\right), \qquad e_{2k} := \left(-\mathbf{1}_{\{k\}}, \mathbf{1}_{\{k\}}\right),$$

are two different sets of J-canonical bases.

If $\mathcal{H}$ is a complex Hilbert space with inner product $\langle ., . \rangle$ then

$$\sigma_{\mathcal{H}}(x, y) := \mathfrak{Im}\, \langle x, y \rangle \tag{B.218}$$

defines a symplectic form on $\mathcal{H}_{\mathbb{R}}$, where $\mathcal{H}_{\mathbb{R}}$ is $\mathcal{H}$ with its real vector space structure. Note that a complex inner product can uniquely be recovered from its imaginary part, as

$$\mathfrak{Re}\, \langle x, y \rangle = \mathfrak{Im}\, i\, \langle x, y \rangle = \mathfrak{Im}\, \langle x, iy \rangle, \tag{B.219}$$

and hence

$$\langle x, y \rangle = \sigma_{\mathcal{H}}(x, iy) + i\sigma_{\mathcal{H}}(x, y), \qquad x, y \in \mathcal{H}.$$

Definition B.27. A complex structure J on a symplectic space (H, σ) is called *symplectic* if J is a symplectic map and is called *positive definite* if

$$\sigma(x, Jx) > 0, \qquad x \in H \setminus \{0\}. \tag{B.220}$$

Lemma B.28. A complex structure is symplectic if and only if the bilinear map

$$\alpha(x, y) := \sigma(x, Jy)$$

is symmetric, and positive definite if and only if the above bilinear map is positive definite.

Proof. The assertion about positive definitivity follows by definition. α is symmetric if and only if

$$\sigma(x, Jy) = -\sigma(Jx, y),$$

which implies

$$\sigma(Jx, Jy) = -\sigma(J^2x, y) = \sigma(x, y),$$

i.e., J is symplectic. On the other hand, if J is symplectic then

$$\sigma(x, Jy) = -\sigma(Jy, x) = \sigma(Jy, J^2x) = \sigma(y, Jx) = -\sigma(Jx, y),$$

i.e., α is symmetric. □

Remark B.29. Note that J is

$$\begin{aligned} \text{a complex structure} &\iff [J]^2 = -[I], \\ \text{symplectic} &\iff [J]^{\mathrm{T}}[\sigma][J] = [\sigma], \\ \text{positive definite} &\iff [\sigma][J] \text{ is a positive definite matrix} \end{aligned}$$

in some (and hence any) basis.

Example B.30. Consider $(\mathcal{H}, \sigma_{\mathcal{H}})$ for a complex inner product space and define

$$Jx := ix, \qquad x \in \mathcal{H}$$

as a real linear map. Obviously, $J^2 = -I$, and

$$\sigma_{\mathcal{H}}(x, Jy) = \mathfrak{Im}\,\langle x, iy\rangle = \mathfrak{Re}\,\langle x, y\rangle\,,$$

and hence $(x, y) \mapsto \sigma(x, Jy)$ is a positive definite symmetric bilinear form. By the above lemma, J is a positive definite symplectic complex structure. Note that

$$\langle x, y\rangle = \mathfrak{Re}\,\langle x, y\rangle + i\,\mathfrak{Im}\,\langle x, y\rangle = \sigma_{\mathcal{H}}(x, Jy) + i\sigma_{\mathcal{H}}(x, y), \qquad x, y \in \mathcal{H}.$$

Lemma B.31. A complex structure on a symplectic space (H, σ) is symplectic and positive definite if and only if there exists a complex inner product on the J-complexification H_J such that

$$\sigma(x, y) = \mathfrak{Im}\, \langle x, y \rangle\,, \qquad x, y \in H. \tag{B.221}$$

In this case, the inner product is uniquely determined by σ and J, as

$$\langle x, y \rangle = \sigma(x, Jy) + i\sigma(x, y), \qquad x, y \in H$$

holds.

Proof. A straightforward computation shows that

$$\langle x, y \rangle := \sigma(x, Jy) + i\sigma(x, y), \qquad x, y \in H$$

defines a sesquilinear form on the J-complexification H_J which is hermitian and positive definite due to J being symplectic and positive definite. Obviously, (B.221) holds.

On the other hand, if there exists a complex inner product on H_J such that (B.221) holds then $\mathfrak{Re}\, \langle x, y \rangle = \sigma(x, Jy)$ by (B.219), and Lemma B.28 yields the assertion. □

Lemma B.32. If $e_1, \ldots, e_{\dim H}$ is a symplectic basis in the symplectic space (H, σ) then

$$Je_{2k-1} := e_{2k}, \qquad Je_{2k} := -e_{2k-1}, \qquad k = 1, \ldots, \dim H \tag{B.222}$$

defines a positive definite symplectic complex structure, and e_{2k-1}, $k = 1, \ldots, \dim H/2$ is an orthonormal basis in H_J. Conversely, if J is a positive definite symplectic complex structure and $f_1, \ldots, f_{\dim H/2}$ is an orthonormal basis in H_J then

$$e_{2k-1} := f_k, \qquad e_{2k} := Jf_k, \qquad k = 1, \ldots, \dim H/2$$

is a symplectic basis in H.

Proof. First, let J as defined in (B.222). Then, the matrix of J in the given symplectic basis is

$$[J] = \bigoplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \text{and hence,} \quad [J]^2 = -[I], \qquad [J]^{\mathrm{T}}[\sigma][J] = [\sigma], \qquad [\sigma][J] = [I],$$

by which J is indeed a positive definite symplectic complex structure.

Now, $f_1, \ldots, f_{\dim H/2}$ be a J-orthonormal basis, i.e.,

$$\delta_{k,l} = \sigma\left(f_k, Jf_l\right) + i\sigma(f_k, f_l) \Longleftrightarrow \begin{cases} \sigma(f_k, f_l) = 0 \\ \sigma\left(f_k, Jf_l\right) = \delta_{k,l}, \end{cases} \qquad k, l = 1, \ldots, \dim H/2.$$

Since J is symplectic, we also have $\sigma(Jf_k, Jf_l) = 0$, $k, l = 1, \ldots, \dim H/2$. Hence, $e_{2k-1} := f_k$, $e_{2k} := Jf_k$, $k = 1, \ldots, \dim H/2$ is indeed a symplectic basis. □

Corollary B.33. There exist infinitely many different positive definite symplectic complex structures on a finite-dimensional symplectic space.

Proof. By the previous lemma, any symplectic basis $e_1, \ldots, e_{\dim H}$ defines a positive definite symplectic complex structure. Now, $e_{c,k} := e_k$, $k \neq 2$ and $e_{c,2} := e_2 + ce_1$ defines a symplectic basis for any $c \in \mathbb{R}$, and by the above lemma, there exist positive definite symplectic complex structures such that

$$J_c e_{c,2k-1} = e_{c,2k}, \qquad J_c e_{c,2k} = -e_{c,2k-1}, \qquad k = 1, \ldots, \dim H.$$

Thus,

$$J_c e_2 = J_c\left(e_{c,2} - ce_1\right) = -e_{c,1} - cJ_c e_1 = -e_1 - ce_2,$$

and hence $J_c \neq J_{c'}$ unless $c = c'$. □

Example B.34. On the symplectic space $(\mathbb{R}^d \times \mathbb{R}^d, \sigma_{\mathbb{R}^{2d}})$,

$$J(x, y) := (-y, x), \qquad x, y \in \mathbb{R}^d$$

defines a complex structure. Since

$$\left((x^{(1)}, x^{(2)}), J(y^{(1)}, y^{(2)})\right) \mapsto \sigma_{\mathbb{R}^{2d}}\left((x^{(1)}, x^{(2)}), J(y^{(1)}, y^{(2)})\right) = x^{(1)}y^{(1)} + x^{(2)}y^{(2)}$$

is a positive definite symmetric bilinear form, J is a positive definite symplectic complex structure. The corresponding inner product is

$$\begin{aligned} \left\langle (x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)}) \right\rangle &:= \sigma((x^{(1)}, x^{(2)}), i(y^{(1)}, y^{(2)})) + i\sigma((x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)})) \\ &= x^{(1)}y^{(1)} + x^{(2)}y^{(2)} + i\left(x^{(1)}y^{(2)} - x^{(2)}y^{(1)}\right). \end{aligned}$$

Note that $Jx = T^{-1}\left(iTx\right)$, where T is the symplectic isomorphism given in Example B.6.

B.4 Inner products in symplectic spaces

It is well-known in linear algebra that two positive semi-definite symmetric forms can simultaneously be diagonzalized, if at least one of them is strictly positive definite, i.e., an inner product. The following lemma establishes a similar result for a symplectic form and an inner product, by showing that they can be brought to a canonical form in the same basis.

Lemma B.35. Let σ be an antisymmetric form and α an inner product. There exists a symplectic basis in H which is also α-orthogonal.

Proof. If $\operatorname{rk}\sigma = 0$ then any α-orthogonal basis does the job, hence we assume for the rest that $\operatorname{rk}\sigma > 0$. This also implies that $\dim H > 1$, since one can easily see that on a one-dimensional space there is no non-zero symplectic form.

Let $T := A_\sigma^{(\alpha)} = \left(\alpha^\flat\right)^{-1} \circ \sigma^\flat$, i.e., $\sigma(x,y) = \alpha(Tx,y)\,, \quad x,y \in H\,.$ Then, for any $x, y \in H$,

$$\alpha(x, T^{\mathrm{T}}y) := \alpha(Tx,y) = \sigma(x,y) = -\sigma(y,x) = -\alpha(Ty,x) = \alpha(x,(-T)y)\,.$$

Hence, $T^{\mathrm{T}} = -T$, where the transpose is taken with respect to α, as defined above. Obviously, $\operatorname{rk} T = \operatorname{rk}\sigma^\flat = \operatorname{rk}\sigma$.

Consider now the operator $T_{\mathbb{C}}$ on the complexification $H_{\mathbb{C}}$, equipped with the inner product $\alpha_{\mathbb{C}}$. Since $T_{\mathbb{C}}^* = (T^{\mathrm{T}})_{\mathbb{C}} = (-T)_{\mathbb{C}} = -T_{\mathbb{C}}$, $T_{\mathbb{C}}$ is a normal operator, and all its eigenvalues are purely imaginary. Let it be a non-zero eigenvalue with corresponding eigenvector $(v^{(1)}, v^{(2)})$, i.e.,

$$(Tv^{(1)}, Tv^{(2)}) = T_{\mathbb{C}}(v^{(1)}, v^{(2)}) = it(v^{(1)}, v^{(2)}) = (-tv^{(2)}, tv^{(1)})\,,$$

or equivalently,

$$Tv^{(1)} = -tv^{(2)}, \qquad Tv^{(2)} = tv^{(1)}\,.$$

As a consequence, neither of $v^{(1)}$ or $v^{(2)}$ can be equal to zero. Further,

$$T_{\mathbb{C}}(v^{(1)}, -v^{(2)}) = (Tv^{(1)}, -Tv^{(2)}) = (-tv^{(2)}, -tv^{(1)}) = -t(v^{(2)}, v^{(1)}) = -it(v^{(1)}, -v^{(2)})\,,$$

and thus $-it$ is an eigenvalue with eigenvector $(v^{(1)}, -v^{(2)})$. Since $it \neq -it$, the corresponding eigensubspaces are orthogonal, and hence

$$\begin{aligned} 0 &= \left\langle (v^{(1)}, v^{(2)}), (v^{(1)}, -v^{(2)}) \right\rangle \\ &= \alpha(v^{(1)}, v^{(1)}) - i\alpha(v^{(1)}, v^{(2)}) - i\alpha(v^{(2)}, v^{(1)}) - \alpha(v^{(2)}, v^{(2)}) \\ &= \alpha(v^{(1)}, v^{(1)}) - \alpha(v^{(2)}, v^{(2)}) - 2i\alpha(v^{(1)}, v^{(2)})\,, \end{aligned}$$

i.e.,

$$\alpha(v^{(1)}, v^{(1)}) = \alpha(v^{(2)}, v^{(2)}) \qquad \text{and} \qquad \alpha(v^{(1)}, v^{(2)}) = 0\,.$$

As a consequence, $v^{(1)}$ and $v^{(2)}$ are α-orthogonal, and can be assumed to be normalized, i.e., $\alpha(v^{(1)}, v^{(1)}) = \alpha(v^{(2)}, v^{(2)}) = 1$. The above computation also shows that the map $(x^{(1)}, x^{(2)}) \mapsto (x^{(1)}, -x^{(2)})$ establishes an isomorphism between the (orthogonal) eigensubspaces corresponding to the eigenvalues it and $-it$. As a consequence, the geometric multiplicity of it and $-it$ are the same, and hence one can list the non-zero eigenvalues with multiplicities as $it_1, -it_1, \ldots, it_n, -it_n,\ t_1, \ldots, t_n > 0$. Moreover, there exist α-orthonormal vectors $f_1, \ldots, f_{2n}$, such that

$$Tf_{2k-1} = -t_k f_{2k}, \qquad Tf_{2k} = t_k f_{2k-1}\,.$$

Let $e_{2k-1} := \frac{1}{\sqrt{t_k}} f_{2k}$, $e_{2k} := \frac{1}{\sqrt{t_k}} f_{2k-1}$, and choose an arbitrary α-orthonormal basis $e_{2n+1}, \ldots, e_{\dim H}$ in the α-orthocomplement of $\operatorname{span}\{e_1, \ldots, e_{2n}\}$ (this can be done e.g. by Gram-Schmidt orthogonalization). Then $e_1, \ldots, e_{\dim H}$ is a basis with the desired properties. □

Corollary B.36. Let σ be a symplectic form and α an inner product on H. There exist positive numbers $a_1, \ldots, a_{\dim H/2}$, uniquely determined by α, and a symplectic basis which is also α-orthogonal such that the matrices of α and σ in this basis are

$$[\alpha] = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} a_k & 0 \\ 0 & a_k \end{bmatrix}, \qquad [\sigma] = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{B.223}$$

Proof. Let $A := T^{-1} = \left(\sigma^\flat\right)^{-1} \circ \alpha^\flat$ and $a_k := \frac{1}{t_k}$ from the proof of the previous lemma. One can easily see that the matrix of A in the basis constructed in the lemma is

$$[A] = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & a_k \\ -a_k & 0 \end{bmatrix}, \tag{B.224}$$

and by (B.214), the matrix of α is as given in (B.223). As a consequence, the characteristic polynomial of A is

$$p(\lambda) = \prod_{k=1}^{d} (\lambda^2 + a_k^2)\,,$$

with complex roots $\pm i a_1, \ldots, \pm i a_d$. Since T and its characteristic polynomial is independent of the choice of the basis, so are the numbers $a_1, \ldots, a_{\dim H/2}$. □

Definition B.37. The numbers $a_1, \ldots, a_{\dim H/2}$ in the above lemma are called the *symplectic eigenvalues* of α (with respect to σ). We denote the symplectic spectrum by $\Sigma(\alpha)$. We call a symplectic basis α-*canonical* if the matrix of α in this basis is of the form (B.223), with the symplectic eigenvalues in the diagonal.

Remark B.38. In general, there may be more than one canonical basis for a given α. On the other hand, the defining operator $A = \left(\sigma^\flat\right)^{-1} \circ \alpha^\flat$ has the form (B.224) in any α-canonical basis, due to (B.214). Hence we call the form (B.224) the canonical form of A.

Corollary B.39. The determinant of α, defined in Definition B.8, is the product of its symplectic values on the square:

$$\det \alpha = \left(a_1 \cdot \ldots \cdot a_{\dim H/2}\right)^2 .$$

Remark B.40. If $f_1, \ldots, f_{\dim H}$ is a symplectic basis which is also α-orthogonal then

$$[\alpha]_f = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} b_k & 0 \\ 0 & c_k \end{bmatrix}, \qquad [\sigma]_f = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{B.225}$$

with some numbers b_k, c_k, $k = 1, \ldots, \dim H/2$, that are positive by

$$b_k = \alpha(f_{2k-1}, f_{2k-1}) > 0, \qquad c_k = \alpha(f_{2k}, f_{2k}) > 0.$$

By (B.214), the matrix of $A = \left(\sigma^\flat\right)^{-1} \circ \alpha^\flat$ in this basis is

$$[A]_f = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & c_k \\ -b_k & 0 \end{bmatrix}$$

and hence the characteristic polynomial of A is

$$p(\lambda) = \prod_{k=1}^{d} (\lambda^2 + b_k c_k) ,$$

from which there has to exist a permutation $\pi \in S_{\dim H/2}$ such that

$$a_k = \sqrt{b_{\pi(k)} c_{\pi(k)}}, \qquad k = 1, \ldots, \dim H/2.$$

In an arbitrary symplectic basis $f_1, \ldots, f_{\dim H}$, one can define an inner product by giving its matrix as in (B.225). Obviously, the basis is then α-orthogonal for the so defined α, but if $b_k \neq c_k$ for some k then it is not α-canonical. Hence α-orthogonality of a symplectic basis doesn't imply it being α-canonical in general.

Example B.41. Let $(H, \sigma) = (\mathcal{H}, \sigma_{\mathcal{H}})$ for some complex Hilbert space $\mathcal{H}$. Any complex linear operator A defines a real bilinear form α by

$$\alpha(x, y) := \sigma(Ax, y) = \mathfrak{Im} \langle Ax, y \rangle .$$

By definition, $A_{\mathbb{R}} = \left(\sigma^{\flat}\right)^{-1} \circ \alpha^{\flat}$, where $A_{\mathbb{R}}$ is simply A as a real linear operator. Then,

$$\alpha(x, y) = \mathfrak{Im} \langle Ax, y \rangle = \mathfrak{Im} \langle x, A^* y \rangle = \mathfrak{Im} \overline{\langle A^* y, x \rangle} = \mathfrak{Im} \langle (-A^*) y, x \rangle ,$$

and hence α is symmetric if and only if $A^* = -A$. In this case, the operator $Q := iA$ is self-adjoint, and

$$\alpha(x, y) = \mathfrak{Im} \langle (-iQ)x, y \rangle = \mathfrak{Im}\, i \langle Qx, y \rangle = \mathfrak{Re} \langle Qx, y \rangle = \mathfrak{Re} \langle x, Qy \rangle .$$

As a consequence, α is an inner product if and only if Q is (strictly) positive definite.

Q has an eigen-decomposition

$$Q = \sum_{k=1}^{d} q_k \, |v_k\rangle\langle v_k| ,$$

and

$$e_{2k-1} := v_k, \qquad e_{2k} := i v_k, \qquad k = 1, \ldots, d$$

is an α-canonical basis. In particular, the symplectic eigenvalues of α coincide with the eigenvalues of Q.

Moreover, if $e_1, \ldots, e_{\dim H}$, is an α-canonical basis then e_{2k-1}, $k = 1, \ldots, \dim H/2$, is an orthonormal eigen-basis of Q, and $e_{2k} = i e_{2k-1}$ for all k. Indeed, $A e_{2k-1} = -q_k e_{2k}$, $A e_{2k} = q_k e_{2k-1}$ and thus $Q e_{2k-1} = -i q_k e_{2k}$, $Q e_{2k} = i q_k e_{2k-1}$ for all k. Hence,

$$Q \left(e_{2k-1} + i e_{2k}\right) = -i q_k e_{2k} - q_k e_{2k-1} = -q_k \left(e_{2k-1} + i e_{2k}\right) ,$$

and, since $-q_k$ is not an eigenvalue of Q, we get $e_{2k-1} + i e_{2k} = 0$, i.e., $e_{2k} = i e_{2k-1}$. Now, for any $m \neq 2k, 2k-1$,

$$\mathfrak{Im} \langle e_m, e_{2k-1} \rangle = 0 = \mathfrak{Im} \langle e_m, e_{2k} \rangle = \mathfrak{Re} \langle e_m, e_{2k-1} \rangle ,$$

and therefore $e_{2k-1} \perp e_{2l-1}$ for $k \neq l$. Finally,

$$1 = \mathfrak{Im} \langle e_{2k-1}, e_{2k} \rangle = \mathfrak{Re} \langle e_{2k-1}, e_{2k-1} \rangle = \left\| e_{2k-1} \right\|^2 .$$

Now let σ be a fixed symplectic form on H and α an inner product with corresponding $A := \left(\sigma^\flat\right)^{-1} \circ \alpha^\flat$. In an α-canonical basis we have the canonical matrix forms given in (B.223) and (B.224), and one may then be tempted to define functions of α, either through

$$[f(\alpha)] := \oplus_{k=1}^{\dim H/2} \begin{bmatrix} f(a_k) & 0 \\ 0 & f(a_k) \end{bmatrix}$$

or

$$f(\alpha)(x,y) := \sigma(f(A)x, y), \qquad x, y \in H,$$

for positive-valued functions on $\Sigma(\alpha)$ in the first case, or entire analytic functions in the second. The second idea has some inherent problems, namely that $f(A)$ doesn't in general define an inner product even for the simplest functions. Indeed,

$$\sigma(A^2x, y) = -\sigma(Ax, Ay) = \sigma(x, A^2y),$$

and hence the bilinear form $(x,y) \mapsto \sigma(A^2x, y)$ is not symmetric. (This is heuristically related to the fact that $[A^2] = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} -a_k^2 & 0 \\ 0 & -a_k^2 \end{bmatrix}$ is not of the canonical form (B.224).) One can easily verify by induction that odd powers $n = 2m+1$ yield symmetric bilinear forms; however, they are positive definite only for $n = 4m+1$, and negative definite for the rest, in accordance with

$$\left[A^{2m+1}\right] = (-1)^m \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & a_k^n \\ -a_k^n & 0 \end{bmatrix}. \tag{B.226}$$

Hence in order to get symplectic eigenvalues $a_1^{2m+1}, \ldots, a_{\dim H/2}^{2m+1}$, one has to apply $f(t) := (-1)^m t^{2m+1}$ to A instead of the naively expected $f(t) = t^{2m+1}$.

The first definition, on the other hand, clearly defines an inner product; what is not clear is whether it is independent of the α-canonical basis in which it is defined. The following lemma shows that both of the above ideas can be used to obtained a well-behaved functional calculus for inner products. We give two different proofs, corresponding to the different approaches outlined above.

Lemma B.42. For any real-valued function f on $\Sigma(\alpha)$, there exists a bilinear form $f(\alpha)$ such that in any α-canonical basis

$$[f(\alpha)] = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} f(a_k) & 0 \\ 0 & f(a_k) \end{bmatrix}, \qquad \left[A_{f(\alpha)}^{(\sigma)}\right] = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & f(a_k) \\ -f(a_k) & 0 \end{bmatrix}. \tag{B.227}$$

In particular, $f(\alpha)$ is an inner product if and only if f is positive-valued, and in this case any α-canonical basis is also $f(\alpha)$-canonical.

Proof 1: Let $H_{\mathbb{C}}$ be the standard complexification of H and $\sigma_{\mathbb{C}}, \alpha_{\mathbb{C}}$ and $A_{\mathbb{C}}$ be the standard complex extensions of α, σ and $A = A_{\alpha}^{(\sigma)}$, respectively. For any α-canonical basis $e_1, \ldots, e_{\dim H}$,

$$Ae_{2k-1} = -a_k e_{2k}, \quad Ae_{2k} = a_k e_{2k-1},$$

and hence,

$$\begin{aligned} A_{\mathbb{C}}(e_{2k}, e_{2k-1}) &= (Ae_{2k}, Ae_{2k-1}) = a_k(e_{2k-1}, -e_{2k}) = -ia_k(e_{2k}, e_{2k-1}) \\ A_{\mathbb{C}}(e_{2k}, -e_{2k-1}) &= (Ae_{2k}, -Ae_{2k-1}) = a_k(e_{2k-1}, e_{2k}) = ia_k(e_{2k}, -e_{2k-1}). \end{aligned}$$

Since $v_{k,\pm} := (e_{2k}, \mp e_{2k-1})$, $k = 1, \ldots, \dim H/2$ is a basis of $H_{\mathbb{C}}$, the spectrum of $A_{\mathbb{C}}$ is $-i\Sigma(\alpha) \cup i\Sigma(\alpha)$.

Choose a polynomial p such that $p(a_k) = f(a_k)$ and $p(-a_k) = -f(a_k)$, $k = 1, \ldots, \dim H$, and define $\tilde{f}(z) := ip(-iz)$, $z \in \mathbb{C}$. Then,

$$\tilde{f}(ia_k) = if(a_k), \quad \tilde{f}(-ia_k) = -if(a_k), \quad k = 1, \ldots, \dim H,$$

and hence

$$\begin{aligned} \tilde{f}\,(A_{\mathbb{C}})\, v_{k,-} &= \tilde{f}(-ia_k) v_{k,-} = -if(a_k) v_{k,-}, \\ \tilde{f}\,(A_{\mathbb{C}})\, v_{k,+} &= \tilde{f}(ia_k) v_{k,+} = if(a_k) v_{k,+}. \end{aligned}$$

Consequently,

$$\begin{aligned} \tilde{f}\,(A_{\mathbb{C}})\,(e_{2k}, 0) &= \tilde{f}\,(A_{\mathbb{C}})\,\frac{1}{2}\,(v_{k,-} + v_{k,+}) = if(a_k)\frac{1}{2}\,(v_{k,+} - v_{k,-}) \\ &= -if(a_k)(0, e_{2k-1}) = f(a_k)(e_{2k-1}, 0), \\ \tilde{f}\,(A_{\mathbb{C}})\,(e_{2k-1}, 0) &= \tilde{f}\,(A_{\mathbb{C}})\,\frac{i}{2}\,(v_{k,+} - v_{k,-}) = \frac{i}{2} if(a_k)\,(v_{k,+} + v_{k,-}) \\ &= -f(a_k)(e_{2k}, 0). \end{aligned}$$

Therefore, $\tilde{f}\,(A_{\mathbb{C}})$ leaves the real subspace $H \times \{0\}$ invariant, and hence one can define the real linear operator A_f on H as the restriction of $\tilde{f}\,(A_{\mathbb{C}})$ onto $H \times \{0\}$. Define

$$f(\alpha)(x, y) := \sigma(A_f x, y).$$

By the above,

$$[A_f]_e = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & f(a_k) \\ -f(a_k) & 0 \end{bmatrix}$$

and hence by (B.214),

$$[f(\alpha)]_e = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} f(a_k) & 0 \\ 0 & f(a_k) \end{bmatrix},$$

which yields that $f(\alpha)$ is an inner product. Since $e_1, \ldots, e_{\dim H}$ was an arbitrary α-canonical basis, the assertion is proven. □

Remark B.43. Note that $A_f \neq f(A)$ in general. The case $f(x) = \frac{1}{x}$ is of special importance; in this case $A_f = -f(A)$, i.e., the inverse of an inner product is defined through

$$\alpha^{-1}(x,y) = \sigma\left(-A^{-1}x, y\right).$$

Proof 2: For $f : \Sigma(\alpha) \to \mathbb{R}$ define

$$\tilde{f} : -\Sigma(\alpha) \cup \Sigma(\alpha) \to \mathbb{R}, \qquad \tilde{f}(x) := \begin{cases} f(x), & x \in \Sigma(\alpha), \\ -f(x), & x \in -\Sigma(\alpha). \end{cases}$$

Choose any odd polynomial $\tilde{p}$ for which $\tilde{p}(x) = \tilde{f}(x)$, $x \in -\Sigma(\alpha) \cup \Sigma(\alpha)$. For instance, the Lagrange interpolation polynomial

$$\tilde{p}(x) := \sum_{k=1}^{\dim H} f(a_k) \prod_{\substack{m=1 \\ m\neq k}}^{d} \frac{x - a_m}{a_k - a_m} \prod_{m=1}^{k} \frac{x + a_m}{a_k + a_m} - f(a_k) \prod_{m=1}^{d} \frac{x - a_m}{-a_k - a_m} \prod_{\substack{m=1 \\ m\neq k}}^{k} \frac{x + a_m}{-a_k + a_m}$$

will do. Being odd, $\tilde{p}$ has the form $\tilde{p}(x) = x \sum_{k=1}^{r} c_k x^{2k}$. Let $p(x) := x \sum_{k=1}^{r} (-1)^k c_k x^{2k}$ and define

$$f(\alpha)(x,y) := \sigma\left(p\left(A_\alpha^{(\sigma)}\right)x, y\right), \quad x, y \in H, \qquad \text{i.e.,} \qquad A_{f(\alpha)}^{(\sigma)} := p\left(A_\alpha^{(\sigma)}\right).$$

By (B.226),

$$\left[p\left(A_\alpha^{(\sigma)}\right)\right] = \bigoplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & p(a_k) \\ -p(a_k) & 0 \end{bmatrix} = \bigoplus_{k=1}^{\dim H/2} \begin{bmatrix} 0 & f(a_k) \\ -f(a_k) & 0 \end{bmatrix}$$

in any α-canonical basis, from which (B.227) follows, and it is also clear that $A_{f(\alpha)}^{(\sigma)} = p\left(A_\alpha^{(\sigma)}\right)$ is independent of the concrete polynomial chosen. □

Example B.44. In the setting of Example B.41, one can define

$$f(\alpha)(x,y) := \mathfrak{Re}\left\langle f(Q)x, y\right\rangle, \qquad x, y \in H$$

for any function f on $\Sigma(\alpha)$, and the result is an inner product whenever f takes strictly positive values on $\Sigma(\alpha)$. If $\mathcal{H}$ is finite-dimensional then the matrix of $f(\alpha)$ in any α-canonical basis is

$$[f(\alpha)] = \oplus_{k=1}^{\dim H/2} \begin{bmatrix} f(a_k) & 0 \\ 0 & f(a_k) \end{bmatrix},$$

and hence this definition of $f(\alpha)$ coincides with the one provided by Lemma B.42. Note that

$$\det f(\alpha) = (\det f(Q))^2 . \tag{B.228}$$

Lemma B.45. Let α be an inner product on a finite-dimensional symplectic space (H,σ). There exists a complex structure J_α on H such that $A := \left(\sigma^\flat\right)^{-1} \circ \alpha^\flat$ is J_α-linear and $Q := iA := J_\alpha A$ is positive definite with

$$\alpha(x,y) = \mathfrak{Re}\,\langle x, Qy\rangle\,, \qquad \langle x,y\rangle = \sigma(x,J_\alpha y) + i\sigma(x,y), \qquad x,y\in H.$$

Moreover, if $e_1,\dots,e_{\dim H}$ is an α-canonical basis then e_{2k-1}, $k=1,\dots,\dim H/2$ is a J_α-orthonormal eigenbasis of Q, and hence

$$Q = \sum_{k=1}^{\dim H/2} a_k\,|e_{2k-1}\rangle\langle e_{2k-1}|\,, \qquad a_1,\dots,a_{\dim H/2}\in\Sigma(\alpha).$$

Vice versa, any J_α-orthonormal eigenbasis $f_1,\dots,f_{\dim H/2}$ of Q defines an α-canonical basis through $e_{2k-1} := f_k$, $e_{2k} := J_\alpha f_k$.

Proof. Let $f(x) := -1$, $x>0$ and $J_\alpha := A^{(\sigma)}_{f(\alpha)}$. Then

$$[J_\alpha] = \oplus_k \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

in any α-canonical basis, and hence

$$\left[J_\alpha^2\right] = -[I], \qquad [J_\alpha]^{\mathrm{T}}\,[\sigma]\,[J_\alpha] = [\sigma], \qquad [\sigma]\,[J_\alpha] = [I], \qquad [A]\,[J_\alpha] = [J_\alpha]\,[A],$$

i.e., J_α is a positive definite symplectic complex structure by Remark B.29, and A is J_α-linear. Moreover,

$$\mathfrak{Re}\,\langle x,Qy\rangle = \sigma\left(x,J_\alpha Qx\right) = \sigma(x,-Ay) = \sigma(Ax,y) = \alpha(x,y), \qquad x,y\in H.$$

Now, if $e_1,\dots,e_{\dim H}$ is an α-canonical basis then

$$\langle e_{2k-1},e_{2l-1}\rangle = \sigma\left(e_{2k-1},J_\alpha e_{2l-1}\right) + i\sigma\left(e_{2k-1},e_{2l-1}\right) = \sigma\left(e_{2k-1},e_{2l}\right) = \delta_{k,l},$$

showing that e_{2k-1}, $k=1,\dots,\dim H/2$, is J_α-orthonormal. Moreover,

$$Qe_{2k-1} = J_\alpha A e_{2k-1} = J_\alpha\left(-a_k e_{2k}\right) = a_k e_{2k-1},$$

and hence e_{2k-1}, $k=1,\dots,\dim H/2$, is an eigenbasis of Q.

Assume now that $f_1,\dots,f_{\dim H/2}$ is a J_α-orthonormal eigenbasis of Q. By Lemma B.32, $e_{2k-1} := f_k$, $e_{2k} := J_\alpha f_k$ defines a symplectic basis. Moreover,

$$\begin{aligned} Ae_{2k-1} &= (-iQ)\,f_k = -ia_k f_k = -a_k e_{2k}, \\ Ae_{2k} &= (-iQ)\,if_k = a_k f_k = a_k e_{2k-1}, \end{aligned}$$

by which $e_1,\dots,e_{\dim H}$ is indeed an α-canonical basis. □

Definition B.46. The above complexification is called *the α-canonical complexification* of H, and Q is the *symbol* of α.

B.5 Gauge-invariant inner products

In the previous section we have seen that for any inner product α there exists a complexification of the symplectic space and a positive definite complex linear operator Q such that

$$\sigma(x,y) = \mathfrak{Im}\,\langle x,y\rangle\,, \qquad \alpha(x,y) = \mathfrak{Re}\,\langle Qx,y\rangle\,, \qquad x,y\in H.$$

Obviously, this complexification depends on α, and we may get different complexifications for different inner products. As later we would like to treat more than one inner products together, it is important to know when there exists a common complexification for them. As a slightly less ambitious goal, in this section we consider the situation when the symplectic space is $(\mathcal{H},\sigma_{\mathcal{H}})$ for some Hilbert space $\mathcal{H}$, and investigate the conditions for an inner product to be compatible with the complex structure of the Hilbert space.

Hence, let $H=\mathcal{H}_{\mathbb{R}}$ and $\sigma(x,y)=\sigma_{\mathcal{H}}(x,y)=\mathfrak{Im}\,\langle x,y\rangle$ for some Hilbert space $\mathcal{H}$. Let A be a real linear map on $\mathcal{H}_{\mathbb{R}}$ and

$$\alpha(x,y) := \sigma(Ax,y) = \mathfrak{Im}\,\langle Ax,y\rangle\,, \qquad x,y\in\mathcal{H},$$

This α is symmetric if and only if

$$\mathfrak{Im}\,\langle Ax,y\rangle = \alpha(x,y) = \alpha(y,x) = \mathfrak{Im}\,\langle Ay,x\rangle\,, \qquad x,y\in\mathcal{H}.$$

Replacing y with iy we get the equivalent condition

$$\mathfrak{Re}\,\langle Ax,y\rangle = \mathfrak{Im}\,i\,\langle Ax,y\rangle = \mathfrak{Im}\,\langle Ax,iy\rangle = \mathfrak{Im}\,\langle Aiy,x\rangle = \mathfrak{Re}\,\langle iAiy,x\rangle\,, \qquad x,y\in\mathcal{H}.$$

Finally, symmetricity of α is equivalent to

$$\begin{aligned}\langle Ax,y\rangle &= \mathfrak{Re}\,\langle Ax,y\rangle + i\,\mathfrak{Im}\,\langle Ax,y\rangle = \mathfrak{Re}\,\langle iAiy,x\rangle - i\,\mathfrak{Im}\,\langle x,Ay\rangle\\ &= \mathfrak{Re}\,\langle x,iAiy\rangle - \mathfrak{Re}\,\langle x,-Ay\rangle + \mathfrak{Re}\,\langle x,-Ay\rangle - i\,\mathfrak{Im}\,\langle x,Ay\rangle\\ &= \langle x,(-A)\,y\rangle + \mathfrak{Re}\,\langle x,(iAi+A)\,y\rangle\,, \qquad x,y\in\mathcal{H}. \qquad \text{(B.229)}\end{aligned}$$

As a consequence, we have the following:

Lemma B.47. For a symmetric bilinear form $\alpha(x,y)=\mathfrak{Im}\,\langle Ax,y\rangle$, we have

$$A \text{ is complex linear} \iff \langle Ax,y\rangle = \langle x,(-A)\,y\rangle\,, \quad x,y\in\mathcal{H}.$$

Proof. Assume first that A is complex linear. Then $iAi=-A$, and hence $\mathfrak{Re}\,\langle x,(iAi+A)\,y\rangle = 0$ in (B.229), from which the assertion follows. Vice versa, if $\langle Ax,y\rangle = \langle x,(-A)\,y\rangle\,$, $x,y\in\mathcal{H}$ then $\mathfrak{Re}\,\langle x,(iAi+A)\,y\rangle = 0$, $x,y\in\mathcal{H}$. Replacing y with iy, we get $\langle x,(iAi+A)\,y\rangle = 0$, $x,y\in\mathcal{H}$, which yields $iAi=-A$, and therefore A is complex linear. □

The map $z \mapsto zI$ gives a unitary representation of the complex unit circle $\mathbb{T}$ on $\mathcal{H}$, which is called the *gauge group* of $\mathcal{H}$. We say that a bilinear form α is *gauge-invariant* if $\alpha(zx, zy) = \alpha(x, y)$ for all $x, y \in \mathcal{H}$ and $z \in \mathbb{T}$.

Lemma B.48. A bilinear form α is gauge-invariant if and only if $A = A_\alpha^{(\sigma)}$ is complex linear.

Proof. Gauge-invariance of α is equivalent to

$$\begin{aligned}\alpha(x,y) &= \alpha\left((\cos t + i\sin t)\,x, (\cos t + i\sin t)\,y\right)\\ &= \cos^2 t\,\alpha(x,y) + \cos t\sin t\,\alpha(x,iy) + \cos t\sin t\,\alpha(ix,y) + \sin^2 t\,\alpha(ix,iy)\end{aligned}$$

for all $x, y \in \mathcal{H}$, $t \in [0, 2\pi)$. By rearranging, we get

$$\frac{1-\cos 2t}{2}[\alpha(ix,iy) - \alpha(x,y)] + \frac{\sin 2t}{2}[\alpha(x,iy) + \alpha(ix,y)] = 0$$

for all t, by which

$$\alpha(ix,iy) = \alpha(x,y), \qquad \alpha(x,iy) = -\alpha(ix,y), \qquad x, y \in \mathcal{H}.$$

Writing out,

$$\begin{aligned}\mathfrak{Re}\,\langle iAx, y\rangle &= \mathfrak{Re}(-i)\,\langle Ax, y\rangle = \mathfrak{Im}\,\langle Ax, y\rangle = \alpha(x,y)\\ &= \alpha(ix,iy) = \mathfrak{Im}\,\langle Aix, iy\rangle = \mathfrak{Im}\,i\,\langle Aix, y\rangle\\ &= \mathfrak{Re}\,\langle Aix, y\rangle\end{aligned}$$

and

$$\begin{aligned}\mathfrak{Im}\,\langle iAx, y\rangle &= \mathfrak{Im}(-i)\,\langle Ax, y\rangle = -\,\mathfrak{Im}\,\langle Ax, iy\rangle = -\alpha(x,iy)\\ &= \alpha(ix,y) = \mathfrak{Im}\,\langle Aix, y\rangle\,.\end{aligned}$$

Finally, α is gauge-invariant if and only if $\langle iAx, y\rangle = \langle Aix, y\rangle$, $x, y \in \mathcal{H}$, i.e., A is complex linear. □

Corollary B.49. A real bilinear form α is a gauge-invariant real inner product if and only if $A = A_\alpha^{(\sigma)}$ is complex linear, $A^* = -A$ and $Q := iA$ is positive definite.

Corollary B.50. A real inner product α is gauge-invariant if and only if $\mathcal{H}$ coincides with the α-canonical complexification of $\mathcal{H}_\mathbb{R}$, i.e., $J_\alpha x = ix$, $x \in \mathcal{H}$.

Definition B.51. Let α be a gauge-invariant real inner product on $(\mathcal{H}, \sigma_\mathcal{H})$ with symbol Q. For any bounded measurable function f on the spectrum of Q, we define

$$f(\alpha)(x,y) := \mathfrak{Re}\,\langle f(Q)x, y\rangle\,, \qquad x, y \in \mathcal{H}.$$

By Example B.44, this definition coincides with that of Lemma B.42 when $\mathcal{H}$ is finite-dimensional and f is real-valued.