1 Vector Spaces

We recall first the definition of a vector space. A vector space is always attached to a set of numbers, called the field of scalars. We will denote this field by \( K \), but mostly we will have in mind the field of real numbers \( \mathbb{R} \) or the field of complex numbers \( \mathbb{C} \).

**Definition 1.1:** A vector space \( V \) over a field \( K \) is a set \( V \), whose elements are called vectors, equipped with two operations: vector addition \(+\), i.e. a map \( V \times V \to V \) and multiplication \( \cdot \) by scalars, i.e. a map \( K \times V \to V \), obeying the following set of axioms:

i) addition is commutative: \( v + w = w + v \), \( \forall v, w \in V \)

ii) addition is associative: \( (v + w) + u = v + (w + u) \), \( \forall u, v, w \in V \)

iii) there exists a distinguished element \( 0 \), called the zero vector, obeying: \( v + 0 = v \), \( \forall v \in V \)

iv) for each vector \( v \in V \) there exists an additive inverse, i.e. a vector \( -v \) such that \( (-v) + v = 0 \)

v) scalar multiplication is distributive over addition of scalars:
\[
(\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v , \quad \forall v \in V , \lambda, \mu \in K
\]

vi) scalar multiplication is distributive over addition of vectors:
\[
\lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w , \quad \forall v, w \in V , \lambda \in K
\]

vii) \( 1 \cdot v = v \), \( \forall v \in V \), where 1 is the unit element in \( K \)

viii) \( \lambda \cdot (\mu \cdot v) = (\lambda \mu) \cdot v \), \( \forall v \in V , \lambda, \mu \in K \).

We will from now on omit the \( \cdot \) in the multiplication by scalars and simply write \( \lambda v \) for the product. Properties (v) and (vi) imply the linearity of the vector operations. Thus, vector spaces are sometimes called *linear spaces*.

From the axioms a number of properties can be derived immediately:

- the zero vector \( 0 \) is *unique* (there is only one zero vector).
- for each vector \( v \in V \) its inverse \( -v \) is unique.
- \( (-1)v = -v \).
• if $\lambda v = 0$, then either $\lambda = 0$ or $v = 0$.

(Note that these properties are not trivial!). Most importantly the axioms allow us to make linear combinations in a vector space, i.e. expressions of the form:

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$$

A collection of vectors \(\{v_1, \ldots, v_k\}\) is called linearly independent if any relation of the form

$$\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k = 0$$

implies $\mu_1 = \mu_2 = \cdots = \mu_k = 0$. A basis of a vector space is any (finite or infinite) collection of vectors \(\{e_1, e_2, \ldots\}\), such that each vector in $V$ can be expressed in a unique way as a linear combination of vectors of the basis, i.e. for every $v$ we have a unique set of scalars $\nu_1, \nu_2, \ldots$ (called the components of $v$ in the given basis) such that

$$v = \nu_1 e_1 + \nu_2 e_2 + \cdots$$

If we can find a finite basis (i.e. a set with a finite number of elements), then we call the vector space $V$ finite-dimensional. In that case, any other basis will have the same number of elements, and that number is called the dimension of the vector space, denoted by $\dim V$. If the vector space is not finite-dimensional, then we call it infinite-dimensional (hence we cannot find a finite basis), and we denote this as: $\dim V = \infty$. There are many examples of infinite-dimensional vector spaces, and some of them are of importance in Quantum Mechanics.

Examples of vector spaces

Let us now consider some important examples of vector spaces.

1. Any subset of a given vector space that is closed under the operations of addition and multiplication by scalars, meaning that if $W \subset V$, with $V$ a vector space, such that:

$$v, w \in W \Rightarrow v + w \in W \text{ and } \lambda v \in W \forall \lambda \in \mathbb{K}$$

is again a vector space in its own right: a subspace of $V$.

2. $\mathbb{R}^n$, the $n$-dimensional Euclidean space. Elements can be represented by ordered $n$-tuples:

$$x = (x_1, x_2, \ldots, x_n), \quad y = (y_1, y_2, \ldots, y_n)$$

equipped with vector addition and multiplication by scalars given by

$$x + y = (x_1 + y_1, \ldots, x_n + y_n), \quad \lambda x = (\lambda x_1, \ldots, \lambda x_n)$$

It can be shown that every $n$-dimensional vector space over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$!

3. $\mathbb{C}^n$, the $n$-dimensional complex vector space of ordered $n$-tuples with entries in $\mathbb{C}$, the field of complex numbers. Addition and multiplication are defined in the same way as in the previous example.
4. We can also consider \( \mathbb{C}^n \) as a \( 2n \)-dimensional vector space, but now over the field of real numbers rather than over the field of complex numbers. This works as follows: to each \( z = (z_1, \ldots, z_n) \in \mathbb{C}^n \) we split the entries in its real and imaginary parts, i.e. \( z_j = x_j + iy_j \), and we can then identify the vectors \( z \) by \( 2n \)-tuples of the form \( (x_1, y_1, \ldots, x_n, y_n) \) with the usual rules. Strictly speaking, this gives you a different vector space from the previous one! (Even though the underlying set of elements is the same!).

5. \( C[a, b] \), the space of continuous real-valued functions over the interval \( [a, b] \). This is an infinite-dimensional vector space: its elements (vectors) are functions, and the addition and multiplication rules are defined through:

\[
\forall f, g \in C[a, b] \quad (f + g)(x) = f(x) + g(x) \quad , \quad (\lambda f)(x) = \lambda f(x) \quad \forall \lambda \in \mathbb{R}
\]

These are the usual point-wise addition and multiplication rules for functions\(^1\). The vector space \( C[a, b] \) is a very big space, and in many circumstances too big to be of much use.

6. \( \ell^p \), (where \( p \) is a fixed real number, \( p \geq 1 \)), which is the space of \( p \)-summable semi-infinite sequences with real- or complex-valued entries:

\[ \xi = (\xi_1, \xi_2, \ldots) \text{ such that } \sum_{j=1}^{\infty} |\xi_j|^p < \infty \]

Introducing the usual addition and multiplication rules:

\[ \xi + \eta = (\xi_1, \xi_2, \ldots) + (\eta_1, \eta_2, \ldots) = (\xi_1 + \eta_1, \xi_2 + \eta_2, \ldots) \quad , \quad \lambda \xi = (\lambda \xi_1, \lambda \xi_2, \ldots) \]

we endow \( \ell^p \) with the structure of an (infinite-dimensional) vector space. A nontrivial aspect of this is that we need to prove that the above summation conditions on the entries remains satisfied when doing vector sums. To prove this we need the so-called Hölder inequality which states\(^2\):

\[
\sum_{j=1}^{\infty} |\xi_j \eta_j| \leq \left( \sum_{k=1}^{\infty} |\xi_k|^p \right)^{1/p} \left( \sum_{l=1}^{\infty} |\eta_l|^q \right)^{1/q} \quad \text{with } \frac{1}{p} + \frac{1}{q} = 1
\]

A consequence of Hölder’s inequality is Minkowski’s inequality which states:

\[
\left( \sum_{k=1}^{\infty} |\xi_j + \eta_j|^p \right)^{1/p} \leq \left( \sum_{k=1}^{\infty} |\xi_k|^p \right)^{1/p} + \left( \sum_{l=1}^{\infty} |\eta_l|^p \right)^{1/p}
\]

from which it follows directly that \( \xi + \eta \) is \( p \)-summable.

\(^1\)Note that these rules are nontrivial: In this context one should distinguish a function, \( f \) say, (regarded as an element of some space), from the evaluation of the same function at some given point \( x \) which is denoted by \( f(x) \).

An important special case of these spaces of sequences is the case $p = 2$, leading to the space $\ell^2$ which is called the *Hilbert sequence space*. The Hölder inequality in that case (i.e. taking $p = q = 2$) reduces to the *Cauchy-Schwarz inequality* which we will prove below. Another related space is $\ell^\infty$ consisting of all bounded sequences, i.e. all $\xi = (\xi_1, \xi_2, \ldots)$ such that there exists real positive number $c = c_\xi$ with $|\xi_j| \leq c_\xi$, (the bound $c$ depending possibly on the vector $\xi$ but not on $j$).

7. $\mathcal{P}_n$, the space of all $n$th order *polynomials*, i.e. expressions of the form

$$P_n(t) = p_n t^n + p_{n-1} t^{n-1} + \cdots + p_0$$

with real- or complex valued coefficients $p_j$, $t$ being an *indeterminate*. The space $\mathcal{P}_n$ becomes a vector space with the addition/multiplication rules;

$$(P_n + Q_n)(t) = (p_n + q_n) t^n + \cdots + (p_0 + q_0), \quad (\lambda P_n)(t) = (\lambda p_n) t^n + \cdots + (\lambda p_0)$$

It is easy to see that $\dim \mathcal{P}_n = n + 1$ and that it is isomorphic to $\mathbb{R}^{n+1}$, resp. $\mathbb{C}^{n+1}$, under the map

$$p_n t^n + \cdots + p_0 \mapsto (p_0, \ldots, p_n)$$

More interesting spaces are the space $\mathbb{R}[t]$, resp. $\mathbb{C}[t]$, which denote the spaces of *all* (i.e. without fixing the order) real- respectively complex-valued polynomials in the indeterminate $t$. These spaces are clearly infinite-dimensional.

8. $L^p(\mathbb{R})$, (with again $p \geq 1$ a fixed real value), which is the space of complex-valued $p$-integrable functions $f$ over the real line: $f : \mathbb{R} \to \mathbb{C}$. The $p$-integrability condition means that we restrict ourselves to functions obeying:

$$\int_{-\infty}^{\infty} |f(x)|^p \, dx < \infty$$

Strictly speaking, the integral here is understood to be a so-called *Lebesgue integral*, which is a generalisation of the more common Riemann-Stieltjes integral\(^3\). Without giving any explanation what a Lebesgue integral is, we just mention that it has the advantage that it is a wider integration concept: more functions are integrable in the Lebesgue sense than in the Riemann-Stieltjes sense which excludes many functions with erratic behaviour. The proof that the operations of addition and multiplication of functions in $L^p(\mathbb{R})$ are *closed* relies on an integral version of Hölder’s inequality:

$$\int |f(x)g(x)| \, dx \leq \left( \int |f(x)|^p \, dx \right)^{1/p} \left( \int |g(x)|^q \, dx \right)^{1/q}$$

and the corresponding Minkowski inequality

$$\int |f(x) + g(x)|^p \, dx \leq \left( \int |f(x)|^p \, dx \right)^{1/p} + \left( \int |g(x)|^q \, dx \right)^{1/q}$$

More generally one can consider so-called $L^p$-spaces on sets $\mathbb{X}$, i.e. $L^p(\mathbb{X})$, when $\mathbb{X}$ is a *measurable* set or forms a so-called *measure space*.

\(^3\)For a full treatment of the subject of Lebesgue integration we refer to the module MATH4011. For “ordinary” functions, like continuous functions, their Lebesgue integral coincides with the Riemann-Stieltjes integral.
Many more examples of, especially infinite-dimensional, vector spaces exist imposing different conditions on the vectors (e.g. functions) involved. They form the subject of what is called topological vector spaces which plays an important role in various areas of pure as well as applied mathematics.

Linear Transformations

In mathematics when one defines a new category of objects such as a space (whether it is a topological space, a vector space, an algebra, a manifold, a variety or whatever) one is interested in studying the class of transformations between the different objects that leave the main structure intact (i.e. that respect the axioms). In general such transformations are called morphisms, (homomorphisms in algebra, homeomorphisms in topology, diffeomorphisms in differential topology and geometry, etc.). The morphisms that are of interest in linear algebra are the so-called linear transformations between vector spaces, because they leave the linear structure of vector spaces intact. They are defined as follows:

Definition 1.2: A transformation \( T : V \to W \) between two vector spaces \( V \) and \( W \), defined over the same field of scalars \( \mathbb{K} \), is called a linear transformation if

\[
T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T(v_1) + \alpha_2 T(v_2) \quad \forall \ v_1, v_2 \in V , \ \alpha_1, \alpha_2 \in \mathbb{K}
\]

If \( V \) and \( W \) are finite-dimensional with \( \dim V = n \), \( \dim W = m \), then any linear transformation can be represented by a \( m \times n \) matrix. The matrix depends on the choice of the bases for \( V \) and \( W \): if \( \{e_1, \ldots, e_n\} \) is the basis of \( V \), and expanding in terms of a basis \( \{e'_1, \ldots, e'_m\} \) of \( W \) we can identify the matrix \( T \) of the transformation with entries

\[
T_{ij} = (T(e_j))_i , \quad i = 1, \ldots, m \ ; \ j = 1, \ldots, n
\]

In the case that \( \dim W = 1 \), i.e the case that \( W \) can be identified with the field \( \mathbb{K} \), we call the transformations linear functionals. The space of all linear functionals on \( V \) is called the dual space of \( V \) and is denoted by \( V^* \): it is a vector space in its own right. If \( V \) is \( n \)-dimensional (\( n \) finite), then \( V^* \) is \( n \)-dimensional as well. In fact, we can choose a basis \( \{E_1, \ldots, E_n\} \) in \( V^* \) whose elements are defined by the action on basis elements \( e_i \) of \( V \) through: \( E_j(e_i) = \delta_{ij} \), \( \delta_{ij} \) being the Kronecker delta-symbol, which induces an isomorphism \( V \to V^* \). This isomorphism depends on the choice of basis, so it is not a really universal construction. However, there does exist a canonical isomorphism between \( V \) and its double dual \( V^{**} = (V^*)^* \) (the dual of the dual of \( V \)) given by

\[
\iota : V \to V^{**} : \ v \mapsto [ \ T \mapsto T(v) ]
\]

(which means: to every \( v \in V \) we associate the functional on \( V^* \) which is the “evaluation map” at the argument \( v \)). The latter morphism also makes sense in the infinite-dimensional case, but is no longer always bijective (it is injective but not necessarily surjective).
Some important notions associated with linear transformations are the following.

- The **kernel** of a linear map $T : V \to W$ is the subspace of all vectors in $V$ that map onto the zero vector in $W$:
  \[
  \text{Ker } T = \{ v \in V \mid T(v) = 0 \}
  \]
  (it is easy to verify that Ker $T$ is a subspace). $T$ is injective iff Ker $T = \{0\}$.

- The **image** of a linear map $T : V \to W$ is the subspace of all vectors in $W$ that are obtained from the application of $T$ on all vectors in $V$, i.e.
  \[
  \text{Im } T = \{ w \in W \mid \exists v \in V : T(v) = w \} = T(V)
  \]
  (it is also easy to verify that Im $T$ is a subspace). $T$ is surjective iff Im $T = W$.

- If $V$ and $W$ are both finite-dimensional, then we have the equality:
  \[
  \dim(\text{Ker } T) + \dim(\text{Im } T) = \dim V
  \]
  however in the infinite-dimensional case there might occur a discrepancy between \( \dim V - \dim(\text{Im } T) \) and \( \dim(\text{Ker } T) \), which is related to the so-called **index** of the transformation $T$.

- Composition of linear transformations: suppose we have successive linear maps $T : V \to W$ and $S : W \to U$ between vector spaces $V$, $W$ and $U$, then the composition
  \[
  S \circ T : V \to U : \ v \mapsto S(T(v))
  \]
  is again a linear transformation. In the finite-dimensional case the matrix associated with the composition map $S \circ T$ is the **matrix product** of the matrices $T_{ij}$ resp. $S_{jk}$ of $T$ and $S$:
  \[
  (S \circ T)_{ik} = \sum_{j=1}^{m} S_{ij} T_{jk}, \quad i = 1, \ldots, l ; \ k = 1, \ldots, n
  \]
  where $n = \dim V$, $m = \dim W$, $l = \dim U$

- If $T : V \to W$ is injective (i.e. when $\dim(\text{Ker } T) = 0$) then one can define its left inverse $T_L^{-1} : W \to V$, defined through: $T_L^{-1} \circ T = \text{Id}_V$, where Id$_V$ is the identity map on $V$ (i.e. Id$_V$(v) = v). If $T$ is surjective then the right inverse $T_R^{-1}$ of $T$ exists which means: $T \circ T_R^{-1} = \text{Id}_W$, with Id$_W$ being the identity map on $W$. If $\dim V = \dim W = n$ finite, then both inverses coincide (if they exist) and can be represented by the matrix inverse of the $n \times n$ matrix representing $T$.

**Eigenvalues and Eigenvectors**

Here we restrict ourselves to linear transformations of a vector space to itself, i.e. $T : V \to V$, where $V$ is a vector space defined over an **algebraically closed field** $\mathbb{K}$ (like $\mathbb{C}$, but not $\mathbb{R}$).
Definition 1.3: An eigenvector of a linear transformation \( T : V \rightarrow V \) is a nonzero vector \( v \in V \) such that \( T(v) = \lambda v \). The corresponding value \( \lambda \in K \) is called an eigenvalue of \( T \). The set of all eigenvalues of \( T \) is called the spectrum of \( T \), denoted as Spec \( T \).

It is well-known that if \( \dim V = n \), i.e. \( V \) is finite-dimensional, the eigenvalues are determined as the roots of the characteristic polynomial
\[
P(\lambda) = \det(T - \lambda I) = 0
\]
in which \( T \) denotes the matrix representing the linear transformation \( T \), \( I \) denoting the \( n \times n \) unit matrix.

We assert the following:

- The characteristic polynomial does **not** depend on the choice of basis in \( V \) (although the matrix \( T \) itself does).
- The following statement holds (Caley-Hamilton’s Theorem): The characteristic polynomial \( P(\lambda) \) of a linear transformation \( T \) annihilates \( T \).
  This statement can be expressed in a formula as \( P(T) = 0_V \), i.e. replacing \( \lambda \) by the transformation \( T \) itself, thereby taking that \( T^k \) means the \( k \)-fold composition of \( T \), \( (T^0 \) is equal to the identity map on \( V \), and \( 0_V \) denotes the zero mapping on \( V \): \( 0_V(v) = 0 \), \( \forall v \in V \). A simpler way of saying this is that \( P(T) = 0 \), the latter being the \( n \times n \) zero matrix, which is a simple matrix identity.
- A change of basis \( \{e_1, \ldots, e_n\} \mapsto \{\tilde{e}_1, \ldots, \tilde{e}_n\} \) given by: \( \tilde{e}_i = \sum_j U_{ji} e_j \) induced by an invertible matrix \( U \) changes the matrix \( T \) representing a linear transformation \( T : V \rightarrow V \) into a matrix \( \tilde{T} = U^{-1}TU \), (i.e. by a so-called similarity transformation).

Definition 1.4: A transformation \( T : V \rightarrow V \) is called diagonalizable if the vectors space \( V \) admits a basis of eigenvectors of the transformation \( T \).

This is the same as saying that if \( T \) is diagonalizable there exists a choice of basis in \( V \) such that in that basis the matrix \( T \) representing \( T \) is a diagonal matrix. Even in the finite-dimensional case not all transformations are diagonalisable. If \( T \) is not diagonalisable then the eigenvectors of \( T \) no longer span the entire vector space and we need to supplement them by additional vectors (the generalised eigenvectors) to find a basis of \( V \). The construction leads in this case to the so-called Jordan normal form for the matrix \( T \) of \( T \), i.e. we can find a basis such that the matrix \( T \) takes the block-diagonal form:

\[
T = \begin{pmatrix}
J_{r_1}(\lambda_1) & \emptyset & \cdots \\
\emptyset & J_{r_2}(\lambda_2) & \cdots \\
\vdots & \vdots & \ddots & \emptyset \\
\emptyset & \cdots & \cdots & \emptyset & J_{r_k}(\lambda_k)
\end{pmatrix}
\]
where the blocks $J_r(\lambda_i)$ are so-called *Jordan blocks* which are either $r_i \times r_i$ diagonal matrices with eigenvalues $\lambda_i$ on the diagonal, or matrices of the form

$$
\begin{pmatrix}
\lambda_i & 1 &  &  \\
& \lambda_i & 1 & \\
&  & \ddots & \ddots \\
&  &  & 1 & \\
&  &  &  & \lambda_i
\end{pmatrix}
$$

2 Inner Product Spaces and Hilbert Spaces

Now that we know what are vector spaces we can add more structure to them by introducing some novel concepts. These new structures will allow us to study some topological properties of the vector spaces. For the purpose of Quantum Mechanics one of the most important notions is that of an *inner product* on a vector space. From now on, unless stated otherwise, we will assume that we work with vector spaces over the field of complex numbers $\mathbb{C}$.

**Definition 2.1:** An inner product on a complex vector space $V$ is a bilinear map: $V \times V \rightarrow \mathbb{C}$ which associates to two vectors $v_1$ and $v_2$ a complex number denoted by $(v_1, v_2)$, having the following properties:

i) $(v_1, v_2) = (v_2, v_1)^*$, $\forall v_1, v_2 \in V$ (Symmetry)

ii) $(w, \alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 (w, v_1) + \alpha_2 (w, v_2)$, $\forall v_1, v_2, w \in V$, $\alpha_1, \alpha_2 \in \mathbb{C}$ (Linearity)

iii) $(v, v) \geq 0$, and $(v, v) = 0 \iff v = 0$ (Positivity)

A vector space equipped with an inner product is called an *inner product space*, (or also a *pre-Hilbert space*). We will usually work with complex vector spaces. However, the definition is easily adapted in a natural way to the case of real vector spaces (i.e. vector spaces defined over the real numbers).

**Examples of inner product spaces**

1. $\mathbb{C}^n$ can be equipped with the natural inner product:

$$
(w, z) = w_1^* z_1 + \cdots + w_n^* z_n
$$

for $w = (w_1, \ldots, w_n)$, $z = (z_1, \ldots, z_n)$ in $\mathbb{C}^n$. In $\mathbb{R}^n$ the analogous inner product is the standard scalar product in Euclidean space (i.e. the usual “dot product”).

---

*For a more complete account on inner product spaces there is a special module MATH3181.*
2. $\ell^2$ can be equipped with the inner product:

$$(\xi, \eta) = \sum_{j=1}^{\infty} \xi_j^* \eta_j$$

3. $C[a,b]$ can be equipped with the inner product

$$(f, g) = \int_{a}^{b} f(x)g(x) \, dx$$

4. $L^2(\mathbb{R})$ can be equipped with the inner product

$$(f, g) = \int_{-\infty}^{\infty} f^*(x)g(x) \, dx$$

to make it into an inner product space. In the latter case we need Cauchy-Schwarz inequality to prove that the integral in the inner product is convergent, and hence that the inner product is well-defined (see below).

Once we have an inner product on a vector space we can start introducing notions from topology such as a metric, limits and convergence properties, continuity, compactness, etc. This arises via the notion of a norm.

**Definition 2.2:** A norm on a real or complex vector space is a map $V \to \mathbb{R}_+ \times (\mathbb{R}_+^\times$ denoting the set of nonnegative real numbers), which associates $v \mapsto \|v\|$, such that the following properties hold

i) $\|v\| \geq 0$, and $\|v\| = 0 \iff v = 0$

ii) $\|\lambda v\| = |\lambda| \|v\|$ \quad $\forall v \in V$, $\lambda \in \mathbb{C}$

(with $|\lambda| = \sqrt{\lambda \lambda^*}$ denoting the usual absolute value of the complex number $\lambda$)

iii) $\|v + w\| \leq \|v\| + \|w\|$ \quad $\forall v, w \in V$ \quad (Triangle inequality)

A vector space equipped with a norm is called a normed space. Each inner product space is also a normed space, because if we have an inner product, we can associate with it a norm through the formula:

$$\|v\| = \sqrt{(v, v)}$$

However, not every norm derives in this way from an inner product. Thus, a normed space is a slightly more general concept than an inner product space. Once we have a norm, we have also a metric (i.e. a distance function) on the space through the formula:

$$d(v, w) = \|v - w\|$$

In general, we don’t need to have vector space in order to define a metric on a set. A metric on an arbitrary set $X$ is a map $d : X \times X \to \mathbb{R}_+^\times$ with the following properties:
\text{i)} \ d(x, y) = 0 \text{ iff } x = y, \ \text{ii)} \ d(x, y) = d(y, x), \ (\text{symmetry}), \ \text{iii)} \ d(x, z) \leq d(x, y) + d(y, z) \quad \text{(triangle inequality)}. \ A \text{ set equipped with a metric is called a metric space, which is a notion from topology: its definition does not rely on any properties of a vector space. If, however, we have a vector space and a norm defined on it we also have automatically a metric space. In particular, an inner product space is a metric space through the above construction.}

**Schwarz and triangle inequality:** An inner product and corresponding norm satisfy the following inequalities:

\begin{enumerate}
\item \[ |(v, w)| \leq \|v\| \|w\| \quad \text{(Schwarz inequality)} \]
The equal sign holds iff \( v \) and \( w \) are linearly dependent.

\item \[ \|v + w\| \leq \|v\| + \|w\| \quad \text{(triangle inequality)} \]
The equality sign holds iff either \( v \) or \( w \) are equal to the zero vector or if they are related through \( v = \alpha w \), where \( \alpha \) is a positive real number.
\end{enumerate}

The proof of the Schwarz inequality is simple. Consider the combination \( v - \alpha w \), then we have:

\[
0 \leq \|v - \alpha w\|^2 = (v - \alpha w, v - \alpha w) = (v, v) - \alpha (v, w) - \alpha^*(w, v) + |\alpha|^2 (w, w)
\]

\[
= \|v\|^2 + |\alpha|^2 \|w\|^2 - 2 \text{Re}(\alpha (v, w))
\]

using the bilinear property of the inner product. Now, either \( \|w\| = 0 \), in which case \( w = 0 \) and the Schwarz inequality is trivially satisfied, or \( \|w\| \neq 0 \) in which case we can choose \( \alpha = (w, v)/\|w\|^2 \). Using the symmetry property of the inner product we then get

\[
0 \leq \|v\|^2 + \frac{|(w, v)|^2}{\|w\|^4} \|w\|^2 - 2 \frac{|(w, v)|^2}{\|w\|^2}
\]

\[
= \|v\|^2 - \frac{|(w, v)|^2}{\|w\|^2}
\]

Moving the last term in the second line to the left-hand side and multiplying by \( \|w\|^2 \) we obtain \( |(w, v)|^2 \leq \|w\|^2 \|v\|^2 \), from which the Schwarz inequality follows immediately.

The triangle inequality follows by considering

\[
\|v + w\|^2 = (v + w, v + w) = (v, v) + (v, w) + (w, v) + (w, w)
\]

\[
= \|v\|^2 + \|w\|^2 + 2 \text{Re}(v, w) \leq \|v\|^2 + \|w\|^2 + 2 |(v, w)| = \|v\|^2 + \|w\|^2 + 2 \|v\| \|w\| = (\|v\| + \|w\|)^2
\]

from which the triangle inequality follows directly. In the first step we have used the fact that \( \text{Re} \lambda \leq |\lambda| \) for any complex number \( \lambda \) and in the second step the Schwarz inequality was used. The equality sign applies if both \( \text{Im} (w, v) = 0 \), \( \text{Re} (w, v) \geq 0 \) and \( v = \alpha w \) with \( \alpha \) as in the proof of the Schwarz inequality. Thus, the equal sign happens iff \( v = \alpha w \) with \( \alpha \) real positive.
Orthonormality & Gram-Schmidt orthogonalisation

In an inner product space we can define the notions of orthogonality and orthonormality:

**Definition 2.3:** Two vectors \( v, w \) are said to be orthogonal to each other if their inner product is zero:

\[
(v, w) = 0 \quad \Leftrightarrow \quad v \perp w
\]

A nonzero vector \( v \) is said to be normalised if its norm is equal to unity: \( \|v\| = 1 \). A set of vectors \( \{v_1, v_2, \ldots \} \) is said to be an orthonormal set if all its vectors are normalised and mutually orthogonal.

An inner product space has the advantage that we can look for an orthonormal basis, i.e. a basis which is an orthonormal set of vectors. This is particularly convenient for the purpose of calculations. An orthonormal basis can be obtained from an arbitrary basis via a constructive procedure which is the so-called Gram-Schmidt orthogonalisation process. It works as follows.

Suppose we have a set of nonzero vectors \( \{v_1, v_2, \ldots \} \), not necessarily orthogonal to each other. We will create step by step an orthogonal set. First, we select one of the vectors, say \( v_1 \), and normalise it by taking

\[
e_1 = \frac{1}{\|v_1\|} v_1
\]

Next, we select a second vector, \( v_2 \) say, and use it to find a vector orthogonal to \( e_1 \) by taking

\[
v'_2 = v_2 - (e_1, v_2) e_1 \quad , \quad e_2 = \frac{1}{\|v'_2\|} v'_2
\]

It is easy to verify that the intermediate vector \( v'_2 \) is orthogonal to \( e_1 \) and we normalise it to get \( e_2 \). Carrying on, we select a third vector \( v_3 \) and create a vector orthogonal to both \( e_1 \) and \( e_2 \) by taking

\[
v'_3 = v_3 - (e_1, v_3) e_1 - (e_2, v_3) e_2
\]

It is easily checked that \( v'_3 \) is orthogonal by taking inner products with \( e_1 \) resp. \( e_2 \). After normalisation, \( e_3 = \frac{v'_3}{\|v'_3\|} \) we have now an orthonormal set of three vectors. We can carry on like this: after \( k - 1 \) steps we have an orthonormal set \( \{e_1, \ldots, e_{k-1}\} \) and the next step is performed by making the \( k \)-th vector \( v_k \) orthogonal to the existing set by taking

\[
v'_k = v_k - \sum_{j=1}^{k-1} (e_j, v_k) e_j \quad , \quad e_k = \frac{1}{\|v'_k\|} v'_k
\]

If along the way we will encounter a vector \( v'_j \) which turns out to be the zero vector, then we throw this vector out since it is then by construction linearly dependent of the existing set. After performing this procedure on a large enough set of initial vectors we thus obtain an orthonormal set, which (in the finite-dimensional case at least) forms an orthonormal basis of the vector space.
**Bessel Inequality:** Once we have a (possibly infinite) orthonormal set of vectors \{e_1, e_2, \ldots \} in an inner product space, we have the inequality:

\[
\sum_{j=1}^{\infty} |(v, e_j)|^2 \leq \|v\|^2 \\
\text{(Bessel inequality)}
\]

**Hilbert and Banach Spaces**

The difference between general inner product spaces and what we call a Hilbert spaces depends on a rather subtle refinement of the definition. The property on which this refinement is based, namely the *completeness* of a space, is needed for a number of technical reasons. To be able to appreciate what this property means, let us consider certain sequences of elements in a space. The main definition of completeness only requires the notion of a metric, so we can formulate it in terms of general metric spaces.

**Definition 2.4:** A sequence of elements \(x_1, x_2, \ldots\) in a metric space \(X\) is called a Cauchy sequence if for each \(\varepsilon > 0\) there exists an integer \(N = N(\varepsilon)\) such that

\[
d(x_n, x_m) < \varepsilon \quad \forall \quad n, m > N
\]

The notion of a Cauchy sequence is weaker than that of a convergent sequence. A sequence \((x_n)\) of elements in the metric space \(X\) is said to converge to a limit \(x\) if for every \(\varepsilon > 0\), there is an integer \(N = N(\varepsilon)\) such that \(d(x_n, x) < \varepsilon\) for all \(n > N\). Every convergent sequence is a Cauchy sequence, but not all Cauchy sequences converge! It depends on the space \(X\) in which we are working: the space \(X\) is called *complete* if in \(X\) all Cauchy sequences converge (i.e. to a limit in \(X\)). An example is the set of all real numbers \(\mathbb{R}\), in contrast to for example \(\mathbb{Q}\) (the set of all rational numbers): whereas \(\mathbb{R}\) is complete, \(\mathbb{Q}\) is not. However if we add to \(\mathbb{Q}\) all limit points of Cauchy sequences in it, which is the procedure of *completion*, then we actually get in this way the set of all real number \(\mathbb{R}\). It is this notion of completeness that we need to finally give the full definition of a Hilbert space:

**Definition 2.5:** A Hilbert space is a *complete* inner product space.

Here the completeness is defined through the metric that arises from the norm, which in its turn arises from the inner product. If we don’t have an inner product, but only a norm, we call the vector space a *Banach space*.

**Definition 2.6:** A Banach space is a *complete* normed space.

Although the notion of a Banach space is more general, we need in Quantum Mechanics the extra structure provided by the inner product. So, in QM we work primarily with Hilbert spaces as the prime spaces of functions in terms of which the states of quantum-mechanical systems are described.
Examples of Hilbert spaces

1. As mentioned above, the set of real numbers \( \mathbb{R} \) as a metric space is complete (for a proof, see any textbook on Analysis). Similarly \( \mathbb{C} \) is complete. As a consequence, since they are also inner product spaces they are Hilbert spaces. Furthermore, \( \mathbb{R}^n \) as well as \( \mathbb{C}^n \) are complete as a consequence of the completeness of \( \mathbb{R} \) resp. \( \mathbb{C} \), and since they are inner product spaces they are Hilbert spaces.

2. The spaces \( C[a,b] \) is not complete, hence it is not a Hilbert space.

3. The space of sequences \( \ell^2 \) is complete, and since it is also an inner product space, it is a Hilbert space.

4. The space \( L^2(\mathbb{R}) \) is complete, and hence a Hilbert space, given the inner product structure given before. This is probably the most important Hilbert space for the purpose of Quantum Mechanics.

Separable Hilbert spaces

The notion of separability is a somewhat technical notion which we need to guarantee the existence of a countable orthonormal basis in a Hilbert space. In the more mathematically rigorous approaches to Quantum Mechanics the requirement of separability is often included in the basic postulates.

Separability is actually a notion from topology: a metric space is called separable if it has a countable dense subset. These words mean the following: a subset of a metric space is dense in the space if any neighborhood of any element of the space contains an element from the subset (in other words if the closure of the subset coincides with the full set, or alternatively if any element of the space can be obtained as a limit of a sequence of elements in the subset). An example is the set of real numbers \( \mathbb{R} \) with the usual metric (i.e. the absolute value of the difference of two numbers) in which the set of rational numbers \( \mathbb{Q} \) lies dense: every neighborhood of any real number contains a rational number, (even if the real number in question is irrational like \( \sqrt{2} \) or \( \pi \)).

An important theorem in functional analysis\(^5\) states that if a Hilbert space \( \mathcal{H} \) is separable, then every orthonormal set is countable, and conversely if \( \mathcal{H} \) contains a complete orthonormal sequence (i.e. an orthonormal set whose span is dense in \( \mathcal{H} \)) then \( \mathcal{H} \) is separable. In these statements the completeness of \( \mathcal{H} \) is crucial: if \( \mathcal{H} \) were not complete such a complete orthonormal set may not exist.

Examples: The Hilbert spaces which we will encounter most commonly in QM are separable: the space \( \ell^2 \) is separable and the space \( L^2([a,b]) \) is separable.

\(^5\)See, e.g. E. Kreyszig, loc. cit. p. 171.
3 Linear Operators on a Hilbert Space

The reason why Hilbert spaces are important for Quantum Mechanics resides in the fact that the completeness ensures many convenient properties for the linear transformations defined on Hilbert spaces. From now on we will call these linear transformations linear operators, or just operators on the Hilbert space. We need to investigate a number of properties of such operators.

First we need the notion of boundedness of operators for which can be defined already in general normed spaces (at this point we don’t yet need inner products):

**Definition 2.7:** A linear transformation \( T : V \rightarrow W \) between two normed spaces is called bounded if there exists a positive real number \( c \) such that

\[
\|T(v)\| \leq c\|v\|, \quad \forall \ v \in V
\]

The collection of all bounded operators \( \mathcal{B}(V, W) = \{T : V \rightarrow W \mid T \text{ linear & bounded}\} \) forms a vector space in its own right with the following rules of addition/multiplication:

\[
T, S \in \mathcal{B}(V, W) : \begin{align*}
(T + S)(v) &= T(v) + T(v) \\
(\lambda T)(v) &= \lambda T(v)
\end{align*} \quad \forall \ \lambda \in \mathbb{C} , \ v, w \in V
\]

It is actually also a normed space itself, by introducing the following operator norm:

\[
\|T\| = \sup_{v \neq 0} \frac{\|T(v)\|}{\|v\|}
\]

If \( W = V \), then we have, in addition to a vector space structure, on the collection of operators an even richer structure provided by the operation of composition \( S \circ T \) of operators. This composition can be viewed as a multiplication rule of operators. Then, if \( V \) is a Banach space, it can be shown that also the normed space of bounded operators \( \mathcal{B}(V, V) \) on \( V \) is also a Banach space, and taking into account the product rule given by \( \circ \) it becomes a so-called Banach algebra.

**Definition 2.8:** An algebra over a field \( \mathbb{K} \) is a vector space \( A \) such that (apart from the usual axioms of a vector space as given in Def. 1.1) we have an additional bilinear operation: \( A \times A \rightarrow A \) which we call the product obeying the axioms:

i) \( (ab)c = a(bc) \), \quad \forall \ a, b, c \in A

ii) \( (a + b)c = ac + ab \), \quad a(b + c) = ab + ac \), \quad \forall \ a, b, c \in A

iii) \( \lambda(ab) = (\lambda a)b = a(\lambda b) \), \quad \forall \ a, b \in A , \ \lambda \in \mathbb{K}

The algebra \( A \) is said to be an algebra with identity if there is an element \( e \), such that \( ea = ae = a \).
Probably the most natural example of an algebra is the vector space of all $n \times n$ matrices over a field $\mathbb{K}$ equipped with the product of matrix multiplication. It is easy to check that this algebra, which is sometimes denoted by $Mat_n(\mathbb{K})$, indeed obeys the axioms i)-iii) given above. This example also demonstrates that in general the product rule in an algebra need not be commutative! (Indeed matrix multiplication is generally not commutative). There are many other examples of algebras, but we will here only be concerned with algebras of linear operators, where the product is given by the composition $\circ$ of operators:

$$S, T \in \mathcal{B}(V, V) : \quad ST(v) = S \circ T(v) = S(T(v))$$

It is straightforward to check that the axioms (i)–(iii) of Def. 2.8 are satisfied (note that the linearity of the operators is crucial). Since we can represent linear operators by matrices (see section 1), the algebra of operators is very similar to the algebra of matrices.

The algebra of bounded operators on a Banach space is a normed algebra, meaning that it is a normed space by virtue of the operator norm, but importantly the norm also respects the algebra structure, namely through the identity:

$$\|ST\| \leq \|S\| \|T\|$$

Together with the completeness of the space of linear operators we have then a so-called Banach algebra. It is an algebra with identity in view of the fact that the identity operator $Id_V$ obeys the proper rules for a unit, and furthermore its norm $\|Id_V\| = 1$.

The operators that we will encounter in Quantum Mechanics also have an algebra structure. However, we will not exclusively deal with bounded operators: there are some important unbounded operators as well (for example the momentum operator $-i\hbar \nabla$ which is not a bounded operator).

Let us now turn to operators on a Hilbert space, and see how linear transformations interplay with the inner product structure. Denoting the Hilbert space in question by $\mathcal{H}$, we are interested in linear operators $T : \mathcal{H} \to \mathcal{H}$.

**Definition 2.9:** The Hermitian conjugate (or: Hilbert-adjoint) of a bounded linear operator $T : \mathcal{H} \to \mathcal{H}$ is the operator $T^\dagger$ defined by

$$(T^\dagger v, w) = (v, Tw) \quad \forall \, v, w \in \mathcal{H}$$

If the operator and its Hermitian conjugate coincide, i.e. $T = T^\dagger$, then $T$ is called a self-adjoint or Hermitian operator.

Important are the following well-known facts:

1. All eigenvalues (if they exist, which may not be always the case) of a Hermitian operator on a (complex) Hilbert space are real.
2. The eigenvectors of a Hermitian operator corresponding to different eigenvalues are orthogonal.

These properties are not difficult to prove:

1. Let $\lambda$ be an eigenvalue of $T$, (T Hermitian), then $Tv = \lambda v$ for a corresponding eigenvector. Thus,

$$\lambda \|v\|^2 = \lambda (v, v) = (v, \lambda v) = (v, Tv) =$$

$$= (Tv, v) = \lambda^* (v, v) = \lambda^* \|v\|^2$$

and since $\|v\| \neq 0$ (the eigenvector not being equal to the zero vector), we have $\lambda = \lambda^*$, hence $\lambda$ is real.

2. Let $\lambda$ and $\mu$ be two different eigenvalues of $T$, associated with eigenvectors $v$, resp. $w$, then

$$\mu (v, w) = (v, \mu w) = (v, Tw) = (Tv, w)$$

$$= \lambda (v, w) \Rightarrow (\mu - \lambda)(v, w) = 0$$

from which it follows that either $\lambda = \mu$ or $(v, w) = 0$, i.e. $v \perp w$.

The properties are of crucial importance to Quantum Mechanics: hermitian operators (since they have real eigenvalues) can be thought to represent physical quantities. The corresponding bases of eigenvectors are automatically orthogonal.

In Def. 2.9 we used bounded operators and for bounded operators many general results hold true. However, in QM we often have to deal also with unbounded operators— in fact, the operator of differentiation is an unbounded operator– which have to be handled with much greater care. Especially, in the basic definitions the proper domains on which such operators are defined need to be given special attention. In this module MATH3383 it is impossible to go into the mathematical details of such matters which implies that in calculations often we have to rely on arguments that will not always have the appropriate mathematical rigour. For those who are interested there are good texts available treating some aspects of the theory in a mathematically rigorous way (see e.g. the booklist on the information sheet). Finally, as a historical remark one should acknowledge that, in fact, Quantum Mechanics has served as an inspiration for many of the developments that gave shape to a rigorous theory of unbounded operators on Hilbert spaces\(^6\), (e.g. the theory developed by J. von Neumann and M.H. Stone in 1929–1932).

A more extensive account of Hilbert spaces and linear operators is also presented in the module MATH3263.

\(^6\)For a good introduction, see e.g E. Kreyszig, loc. cit., chapter 10.