

Chapter 3

Banach and Hilbert Spaces

Dr. von Neumann, ich möchte gern wissen,
was ist dann eigentlich ein Hilbertscher
Raum?

DAVID HILBERT

This chapter covers the necessary concepts from linear functional analysis on Hilbert and Banach spaces: in particular, we review here basic constructions such as orthogonality, direct sums and tensor products. Like Chapter 2, this chapter is intended as a review of material that should be understood as a prerequisite before proceeding; to an extent, Chapters 2 and 3 are interdependent and so can (and should) be read in parallel with one another.

3.1 Basic Definitions and Properties

In what follows, \mathbb{K} will denote either the real numbers \mathbb{R} or the complex numbers \mathbb{C} , and $|\cdot|$ denotes the absolute value function on \mathbb{K} . All the vector spaces considered in this book will be vector spaces over one of these two fields. In \mathbb{K} , notions of ‘size’ and ‘closeness’ are provided by the absolute value function $|\cdot|$. In a normed vector space, similar notions of ‘size’ and ‘closeness’ are provided by a function called a norm, from which we can build up notions of convergence, continuity, limits and so on.

Definition 3.1. A *norm* on a vector space \mathcal{V} over \mathbb{K} is a function $\|\cdot\|: \mathcal{V} \rightarrow \mathbb{R}$ that is

- (a) *positive semi-definite*: for all $x \in \mathcal{V}$, $\|x\| \geq 0$;
- (b) *positive definite*: for all $x \in \mathcal{V}$, $\|x\| = 0$ if and only if $x = 0$;

(c) *positively homogeneous*: for all $x \in \mathcal{V}$ and $\alpha \in \mathbb{K}$, $\|\alpha x\| = |\alpha|\|x\|$; and

(d) *sublinear*: for all $x, y \in \mathcal{V}$, $\|x + y\| \leq \|x\| + \|y\|$.

If the positive definiteness requirement is omitted, then $\|\cdot\|$ is said to be a *seminorm*. A vector space equipped with a norm (resp. seminorm) is called a *normed space* (resp. *seminormed space*).

In a normed vector space, we can sensibly talk about the ‘size’ or ‘length’ of a single vector, but there is no sensible notion of ‘angle’ between two vectors, and in particular there is no notion of orthogonality. Such notions are provided by an inner product:

Definition 3.2. An *inner product* on a vector space \mathcal{V} over \mathbb{K} is a function $\langle \cdot, \cdot \rangle: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{K}$ that is

(a) *positive semi-definite*: for all $x \in \mathcal{V}$, $\langle x, x \rangle \geq 0$;

(b) *positive definite*: for all $x \in \mathcal{V}$, $\langle x, x \rangle = 0$ if and only if $x = 0$;

(c) *conjugate symmetric*: for all $x, y \in \mathcal{V}$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$; and

(d) *sesquilinear*: for all $x, y, z \in \mathcal{V}$ and all $\alpha, \beta \in \mathbb{K}$, $\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle$.

A vector space equipped with an inner product is called an *inner product space*. In the case $\mathbb{K} = \mathbb{R}$, conjugate symmetry becomes symmetry, and sesquilinearity becomes bilinearity.

Many texts have sesquilinear forms be linear in the *first* argument, rather than the second as is done here; this is an entirely cosmetic difference that has no serious consequences, provided that one makes a consistent choice and sticks with it.

It is easily verified that every inner product space is a normed space under the *induced norm*

$$\|x\| := \sqrt{\langle x, x \rangle}.$$

The inner product and norm satisfy the *Cauchy-Schwarz inequality*

$$|\langle x, y \rangle| \leq \|x\|\|y\| \quad \text{for all } x, y \in \mathcal{V}, \quad (3.1)$$

where equality holds in (3.1) if and only if x and y are scalar multiples of one another. Every norm on \mathcal{V} that is induced by an inner product satisfies the *parallelogram identity*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \text{for all } x, y \in \mathcal{V}. \quad (3.2)$$

In the opposite direction, if $\|\cdot\|$ is a norm on \mathcal{V} that satisfies the parallelogram identity (3.2), then the unique inner product $\langle \cdot, \cdot \rangle$ that induces this norm is found by the *polarization identity*

$$\langle x, y \rangle = \frac{\|x + y\|^2 - \|x - y\|^2}{4} \quad (3.3)$$

in the real case, and

$$\langle x, y \rangle = \frac{\|x + y\|^2 - \|x - y\|^2}{4} + i \frac{\|ix - y\|^2 - \|ix + y\|^2}{4} \quad (3.4)$$

in the complex case.

The simplest examples of normed and inner product spaces are the familiar finite-dimensional Euclidean spaces:

Example 3.3. Here are some finite-dimensional examples of norms on \mathbb{K}^n :

- (a) The absolute value function $|\cdot|$ is a norm on \mathbb{K} .
- (b) The most familiar example of a norm is probably the *Euclidean norm* or *2-norm* on \mathbb{K}^n . The Euclidean norm of $v = (v_1, \dots, v_n) \in \mathbb{K}^n$ is given by

$$\|v\|_2 := \sqrt{\sum_{i=1}^n |v_i|^2} = \sqrt{\sum_{i=1}^n |v \cdot e_i|^2}. \quad (3.5)$$

The Euclidean norm is the induced norm for the inner product

$$\langle u, v \rangle := \sum_{i=1}^n \overline{u_i} v_i. \quad (3.6)$$

In the case $\mathbb{K} = \mathbb{R}$ this inner product is commonly called the *dot product* and denoted $u \cdot v$.

- (c) The analogous inner product and norm on $\mathbb{K}^{m \times n}$ of $m \times n$ matrices is the *Frobenius inner product*

$$\langle A, B \rangle \equiv A : B := \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \overline{a_{ij}} b_{ij}.$$

- (d) The *1-norm*, also known as the *Manhattan norm* or *taxicab norm*, on \mathbb{K}^n is defined by

$$\|v\|_1 := \sum_{i=1}^n |v_i|. \quad (3.7)$$

- (e) More generally, for $1 \leq p < \infty$, the *p-norm* on \mathbb{K}^n is defined by

$$\|v\|_p := \left(\sum_{i=1}^n |v_i|^p \right)^{1/p}. \quad (3.8)$$

- (f) Note, however, that the formula in (3.8) does *not* define a norm on \mathbb{K}^n if $p < 1$.
- (g) The analogous norm for $p = \infty$ is the ∞ -norm or *maximum norm* on \mathbb{K}^n :

$$\|v\|_\infty := \max_{i=1, \dots, n} |v_i|. \quad (3.9)$$

There are also many straightforward examples of infinite-dimensional normed spaces. In UQ applications, these spaces often arise as the solution spaces for ordinary or partial differential equations, spaces of random variables, or spaces for sequences of coefficients of expansions of random fields and stochastic processes.

Example 3.4. (a) An obvious norm to define for a sequence $v = (v_n)_{n \in \mathbb{N}}$ is the analogue of the maximum norm. That is, define the *supremum norm* by

$$\|v\|_\infty := \sup_{n \in \mathbb{N}} |v_n|. \quad (3.10)$$

Clearly, if v is not a bounded sequence, then $\|v\|_\infty = \infty$. Since norms are not allowed to take the value ∞ , the supremum norm is only a norm on the space of *bounded sequences*; this space is often denoted ℓ^∞ , or sometimes $\ell^\infty(\mathbb{K})$ if we wish to emphasize the field of scalars, or $\mathcal{B}(\mathbb{N}; \mathbb{K})$ if we want to emphasize that it is a space of bounded functions on some set, in this case \mathbb{N} .

(b) Similarly, for $1 \leq p < \infty$, the *p-norm* of a sequence is defined by

$$\|v\|_p := \left(\sum_{n \in \mathbb{N}} |v_n|^p \right)^{1/p}. \quad (3.11)$$

The space of sequences for which this norm is finite is the space of *p-summable sequences*, which is often denoted $\ell^p(\mathbb{K})$ or just ℓ^p . The statement from elementary analysis courses that $\sum_{n=1}^{\infty} \frac{1}{n}$ (the harmonic series) diverges but that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges is the statement that

$$\left(1, \frac{1}{2}, \frac{1}{3}, \dots\right) \in \ell^2 \quad \text{but} \quad \left(1, \frac{1}{2}, \frac{1}{3}, \dots\right) \notin \ell^1.$$

(c) If S is any set, and $\mathcal{B}(S; \mathbb{K})$ denotes the vector space of all bounded \mathbb{K} -valued functions on S , then a norm on $\mathcal{B}(S; \mathbb{K})$ is the *supremum norm* (or *uniform norm*) defined by

$$\|f\|_\infty := \sup_{x \in S} |f(x)|.$$

(d) Since every continuous function on a closed and bounded interval is bounded, the supremum norm is also a norm on the space $\mathcal{C}^0([0, 1]; \mathbb{R})$ of continuous real-valued functions on the unit interval.

There is a natural norm to use for linear functions between two normed spaces:

Definition 3.5. Given normed spaces \mathcal{V} and \mathcal{W} , the *operator norm* of a linear map $A: \mathcal{V} \rightarrow \mathcal{W}$ is

$$\|A\| := \sup_{0 \neq v \in \mathcal{V}} \frac{\|A(v)\|_{\mathcal{W}}}{\|v\|_{\mathcal{V}}} \equiv \sup_{\substack{v \in \mathcal{V} \\ \|v\|_{\mathcal{V}}=1}} \|A(v)\|_{\mathcal{W}} \equiv \sup_{\substack{v \in \mathcal{V} \\ \|v\|_{\mathcal{V}} \leq 1}} \|A(v)\|_{\mathcal{W}}.$$

If $\|A\|$ is finite, then A is called a *bounded linear operator*. The operator norm of A will also be denoted $\|A\|_{\text{op}}$ or $\|A\|_{\mathcal{V} \rightarrow \mathcal{W}}$. There are many equivalent expressions for this norm: see Exercise 3.1.

Definition 3.6. Two inner product spaces $(\mathcal{V}, \langle \cdot, \cdot \rangle_{\mathcal{V}})$ and $(\mathcal{W}, \langle \cdot, \cdot \rangle_{\mathcal{W}})$ are said to be *isometrically isomorphic* if there is an invertible linear map $T: \mathcal{V} \rightarrow \mathcal{W}$ such that

$$\langle Tu, Tv \rangle_{\mathcal{W}} = \langle u, v \rangle_{\mathcal{V}} \quad \text{for all } u, v \in \mathcal{V}.$$

The two inner product spaces are then ‘the same up to relabelling’. Similarly, two normed spaces are isometrically isomorphic if there is an invertible linear map that preserves the norm.

Finally, normed spaces are examples of topological spaces, in that the norm structure induces a collection of open sets and (as will be revisited in the next section) a notion of convergence:

Definition 3.7. Let \mathcal{V} be a normed space:

(a) For $x \in \mathcal{V}$ and $r > 0$, the *open ball of radius r centred on x* is

$$\mathbb{B}_r(x) := \{y \in \mathcal{V} \mid \|x - y\| < r\} \quad (3.12)$$

and the *closed ball of radius r centred on x* is

$$\overline{\mathbb{B}}_r(x) := \{y \in \mathcal{V} \mid \|x - y\| \leq r\}. \quad (3.13)$$

(b) A subset $U \subseteq \mathcal{V}$ is called an *open set* if, for all $x \in U$, there exists $r = r(x) > 0$ such that $\mathbb{B}_r(x) \subseteq U$.

(c) A subset $F \subseteq \mathcal{V}$ is called a *closed set* if $\mathcal{V} \setminus F$ is an open set.

3.2 Banach and Hilbert Spaces

For the purposes of analysis, rather than pure algebra, it is convenient if normed spaces are complete in the same way that \mathbb{R} is complete and \mathbb{Q} is not:

Definition 3.8. Let $(\mathcal{V}, \|\cdot\|)$ be a normed space.

(a) A sequence $(x_n)_{n \in \mathbb{N}}$ in \mathcal{V} *converges* to $x \in \mathcal{V}$ if, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that, whenever $n \geq N$, $\|x_n - x\| < \varepsilon$.

(b) A sequence $(x_n)_{n \in \mathbb{N}}$ in \mathcal{V} is called *Cauchy* if, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that, whenever $m, n \geq N$, $\|x_m - x_n\| < \varepsilon$.

(c) A *complete space* is one in which each Cauchy sequence in \mathcal{V} converges to some element of \mathcal{V} . Complete normed spaces are called *Banach spaces*, and complete inner product spaces are called *Hilbert spaces*.

It is easily verified that a subset F of a normed space is closed (in the topological sense of being the complement of an open set) if and only if it is closed under the operation of taking limits of sequences (i.e. every convergent sequence in F has its limit also in F), and that closed linear subspaces of Banach (resp. Hilbert) spaces are again Banach (resp. Hilbert) spaces.

- Example 3.9.** (a) \mathbb{K}^n and $\mathbb{K}^{m \times n}$ are finite-dimensional Hilbert spaces with respect to their usual inner products.
 (b) The standard example of an infinite-dimensional Hilbert space is the space $\ell^2(\mathbb{K})$ of square-summable \mathbb{K} -valued sequences, which is a Hilbert space with respect to the inner product

$$\langle x, y \rangle_{\ell^2} := \sum_{n \in \mathbb{N}} \overline{x_n} y_n.$$

This space is the prototypical example of a separable Hilbert space, i.e. it has a countably infinite dense subset, and hence countably infinite dimension.

- (c) On the other hand, the subspace of ℓ^2 consisting of all sequences with only finitely many non-zero terms is a non-closed subspace of ℓ^2 , and not a Hilbert space. Of course, if the non-zero terms are restricted to lie in a predetermined finite range of indices, say $\{1, \dots, n\}$, then the subspace is an isomorphic copy of the Hilbert space \mathbb{K}^n .
 (d) Given a measure space $(\mathcal{X}, \mathcal{F}, \mu)$, the space $L^2(\mathcal{X}, \mu; \mathbb{K})$ of (equivalence classes modulo equality μ -almost everywhere of) square-integrable functions from \mathcal{X} to \mathbb{K} is a Hilbert space with respect to the inner product

$$\langle f, g \rangle_{L^2(\mu)} := \int_{\mathcal{X}} \overline{f(x)} g(x) \, d\mu(x). \quad (3.14)$$

Note that it is necessary to take the quotient by the equivalence relation of equality μ -almost everywhere since a function f that vanishes on a set of full measure but is non-zero on a set of zero measure is not the zero function but nonetheless has $\|f\|_{L^2(\mu)} = 0$. When $(\mathcal{X}, \mathcal{F}, \mu)$ is a probability space, elements of $L^2(\mathcal{X}, \mu; \mathbb{K})$ are thought of as random variables of finite variance, and the L^2 inner product is the covariance:

$$\langle X, Y \rangle_{L^2(\mu)} := \mathbb{E}_{\mu} [\overline{XY}] = \text{cov}(X, Y).$$

When $L^2(\mathcal{X}, \mu; \mathbb{K})$ is a separable space, it is isometrically isomorphic to $\ell^2(\mathbb{K})$ (see Theorem 3.24).

- (e) Indeed, Hilbert spaces over a fixed field \mathbb{K} are classified by their dimension: whenever \mathcal{H} and \mathcal{K} are Hilbert spaces of the same dimension over \mathbb{K} , there is an invertible \mathbb{K} -linear map $T: \mathcal{H} \rightarrow \mathcal{K}$ such that $\langle Tx, Ty \rangle_{\mathcal{K}} = \langle x, y \rangle_{\mathcal{H}}$ for all $x, y \in \mathcal{H}$.

Example 3.10. (a) For a compact topological space \mathcal{X} , the space $\mathcal{C}^0(\mathcal{X}; \mathbb{K})$ of continuous functions $f: \mathcal{X} \rightarrow \mathbb{K}$ is a Banach space with respect to the *supremum norm*

$$\|f\|_\infty := \sup_{x \in \mathcal{X}} |f(x)|. \quad (3.15)$$

For non-compact \mathcal{X} , the supremum norm is only a bona fide norm if we restrict attention to bounded continuous functions, since otherwise it would take the inadmissible value $+\infty$.

(b) More generally, if \mathcal{X} is the compact closure of an open subset of a Banach space \mathcal{V} , and $r \in \mathbb{N}_0$, then the space $\mathcal{C}^r(\mathcal{X}; \mathbb{K})$ of all r -times continuously differentiable functions from \mathcal{X} to \mathbb{K} is a Banach space with respect to the norm

$$\|f\|_{\mathcal{C}^r} := \sum_{k=0}^r \|D^k f\|_\infty.$$

Here, $Df(x): \mathcal{V} \rightarrow \mathbb{K}$ denotes the first-order *Fréchet derivative* of f at x , the unique bounded linear map such that

$$\lim_{\substack{y \rightarrow x \\ \text{in } \mathcal{X}}} \frac{|f(y) - f(x) - Df(x)(y - x)|}{\|y - x\|} = 0,$$

$D^2 f(x) = D(Df)(x): \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{K}$ denotes the second-order Fréchet derivative, etc.

(c) For $1 \leq p \leq \infty$, the spaces $L^p(\mathcal{X}, \mu; \mathbb{K})$ from Definition 2.21 are Banach spaces, but only the L^2 spaces are Hilbert spaces. As special cases ($\mathcal{X} = \mathbb{N}$, and $\mu =$ counting measure), the sequence spaces ℓ^p are also Banach spaces, and are Hilbert if and only if $p = 2$.

Another family of Banach spaces that arises very often in PDE applications is the family of *Sobolev spaces*. For the sake of brevity, we limit the discussion to those Sobolev spaces that are also Hilbert spaces. To save space, we use multi-index notation for derivatives: for a multi-index $\alpha := (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$, with $|\alpha| := \alpha_1 + \dots + \alpha_n$,

$$\partial^\alpha u(x) := \frac{\partial^{|\alpha|} u}{\partial^{\alpha_1} x_1 \dots \partial^{\alpha_n} x_n}(x).$$

Sobolev spaces consist of functions¹ that have appropriately integrable *weak* derivatives, as defined by integrating by parts against smooth test functions:

¹ To be more precise, as with the Lebesgue L^p spaces, Sobolev spaces consist of *equivalence classes* of such functions, with equivalence being equality almost everywhere.

Definition 3.11. Let $\mathcal{X} \subseteq \mathbb{R}^n$, let $\alpha \in \mathbb{N}_0^n$, and consider $u: \mathcal{X} \rightarrow \mathbb{R}$. A *weak derivative of order α* for u is a function $v: \mathcal{X} \rightarrow \mathbb{R}$ such that

$$\int_{\mathcal{X}} u(x) \partial^\alpha \phi(x) \, dx = (-1)^{|\alpha|} \int_{\mathcal{X}} v(x) \phi(x) \, dx \quad (3.16)$$

for every smooth function $\phi: \mathcal{X} \rightarrow \mathbb{R}$ that vanishes outside a compact subset $\text{supp}(\phi) \subseteq \mathcal{X}$. Such a weak derivative is usually denoted $\partial^\alpha u$ as if it were a strong derivative, and indeed coincides with the classical (strong) derivative if the latter exists. For $s \in \mathbb{N}_0$, the *Sobolev space* $H^s(\mathcal{X})$ is

$$H^s(\mathcal{X}) := \left\{ u \in L^2(\mathcal{X}) \mid \begin{array}{l} \text{for all } \alpha \in \mathbb{N}_0^n \text{ with } |\alpha| \leq s, \\ u \text{ has a weak derivative } \partial^\alpha u \in L^2(\mathcal{X}) \end{array} \right\} \quad (3.17)$$

with the inner product

$$\langle u, v \rangle_{H^s} := \sum_{|\alpha| \leq s} \langle \partial^\alpha u, \partial^\alpha v \rangle_{L^2}. \quad (3.18)$$

The following result shows that smoothness in the Sobolev sense implies either a greater degree of integrability or even Hölder continuity. In particular, possibly after modification on sets of Lebesgue measure zero, Sobolev functions in H^s are continuous when $s > n/2$. Thus, such functions can be considered to have well-defined pointwise values.

Theorem 3.12 (Sobolev embedding theorem). *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a Lipschitz domain (i.e. a connected set with non-empty interior, such that $\partial\mathcal{X}$ can always be locally written as the graph of a Lipschitz function of $n - 1$ variables).*

(a) *If $s < n/2$, then $H^s(\mathcal{X}) \subseteq L^q(\mathcal{X})$, where $\frac{1}{q} = \frac{1}{2} - \frac{s}{n}$, and there is a constant $C = C(s, n, \mathcal{X})$ such that*

$$\|u\|_{L^q(\mathcal{X})} \leq C \|u\|_{H^s(\mathcal{X})} \quad \text{for all } u \in H^s(\mathcal{X}).$$

(b) *If $s > n/2$, then $H^s(\mathcal{X}) \subseteq C^{s - \lfloor n/2 \rfloor - 1, \gamma}(\mathcal{X})$, where*

$$\gamma = \begin{cases} \lfloor n/2 \rfloor + 1 - n/2, & \text{if } n \text{ is odd,} \\ \text{any element of } (0, 1), & \text{if } n \text{ is even,} \end{cases}$$

and there is a constant $C = C(s, n, \gamma, \mathcal{X})$ such that

$$\|u\|_{C^{s - \lfloor n/2 \rfloor - 1, \gamma}(\mathcal{X})} \leq C \|u\|_{H^s(\mathcal{X})} \quad \text{for all } u \in H^s(\mathcal{X}),$$

where the Hölder norm is defined (up to equivalence) by

$$\|u\|_{C^{k, \gamma}(\mathcal{X})} := \|u\|_{C^k} + \sup_{\substack{x, y \in \mathcal{X} \\ x \neq y}} \frac{|D^k u(x) - D^k u(y)|}{|x - y|}.$$

3.3 Dual Spaces and Adjoint

Dual Spaces. Many interesting properties of a vector space are encoded in a second vector space whose elements are the linear functions from the first space to its field. When the vector space is a normed space,² so that concepts like continuity are defined, it makes sense to study continuous linear functions:

Definition 3.13. The *continuous dual space* of a normed space \mathcal{V} over \mathbb{K} is the vector space \mathcal{V}' of all bounded (equivalently, continuous) linear functionals $\ell: \mathcal{V} \rightarrow \mathbb{K}$. The dual pairing between an element $\ell \in \mathcal{V}'$ and an element $v \in \mathcal{V}$ is denoted $\langle \ell | v \rangle$ or simply $\ell(v)$. For a linear functional ℓ on a seminormed space \mathcal{V} , being continuous is equivalent to being *bounded* in the sense that its *operator norm* (or *dual norm*)

$$\|\ell\|' := \sup_{0 \neq v \in \mathcal{V}} \frac{|\langle \ell | v \rangle|}{\|v\|} \equiv \sup_{\substack{v \in \mathcal{V} \\ \|v\|=1}} |\langle \ell | v \rangle| \equiv \sup_{\substack{v \in \mathcal{V} \\ \|v\| \leq 1}} |\langle \ell | v \rangle|$$

is finite.

Proposition 3.14. *For every normed space \mathcal{V} , the dual space \mathcal{V}' is a Banach space with respect to $\|\cdot\|'$.*

An important property of Hilbert spaces is that they are naturally *self-dual*: every continuous linear functional on a Hilbert space can be naturally identified with the action of taking the inner product with some element of the space:

Theorem 3.15 (Riesz representation theorem). *Let \mathcal{H} be a Hilbert space. For every continuous linear functional $f \in \mathcal{H}'$, there exists $f^\sharp \in \mathcal{H}$ such that $\langle f | x \rangle = \langle f^\sharp, x \rangle$ for all $x \in \mathcal{H}$. Furthermore, the map $f \mapsto f^\sharp$ is an isometric isomorphism between \mathcal{H} and its dual.*

The simplicity of the Riesz representation theorem for duals of Hilbert spaces stands in stark contrast to the duals of even elementary Banach spaces, which are identified on a more case-by-case basis:

- For $1 < p < \infty$, $L^p(\mathcal{X}, \mu)$ is isometrically isomorphic to the dual of $L^q(\mathcal{X}, \mu)$, where $\frac{1}{p} + \frac{1}{q} = 1$. This result applies to the sequence space ℓ^p , and indeed to the finite-dimensional Banach spaces \mathbb{R}^n and \mathbb{C}^n with the norm $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$.
- By the Riesz–Markov–Kakutani representation theorem, the dual of the Banach space $\mathcal{C}_c(\mathcal{X})$ of compactly supported continuous functions on a locally compact Hausdorff space \mathcal{X} is isomorphic to the space of regular signed measures on \mathcal{X} .

² Or even just a topological vector space.

The second example stands as another piece of motivation for measure theory in general and signed measures in particular. Readers interested in the details of these constructions should refer to a specialist text on functional analysis.

Adjoint Maps. Given a linear map $A: \mathcal{V} \rightarrow \mathcal{W}$ between normed spaces \mathcal{V} and \mathcal{W} , the *adjoint* of A is the linear map $A^*: \mathcal{W}' \rightarrow \mathcal{V}'$ defined by

$$\langle A^* \ell | v \rangle = \langle \ell | Av \rangle \quad \text{for all } v \in \mathcal{V} \text{ and } \ell \in \mathcal{W}'.$$

The following properties of adjoint maps are fundamental:

Proposition 3.16. *Let \mathcal{U} , \mathcal{V} and \mathcal{W} be normed spaces, let $A, B: \mathcal{V} \rightarrow \mathcal{W}$ and $C: \mathcal{U} \rightarrow \mathcal{V}$ be bounded linear maps, and let α and β be scalars. Then*

- (a) $A^*: \mathcal{W}' \rightarrow \mathcal{V}'$ is bounded, with operator norm $\|A^*\| = \|A\|$;
- (b) $(\alpha A + \beta B)^* = \bar{\alpha}A^* + \bar{\beta}B^*$;
- (c) $(AC)^* = C^*A^*$;
- (d) the kernel and range of A and A^* satisfy

$$\begin{aligned} \ker A^* &= (\text{ran } A)^\perp := \{\ell \in \mathcal{W}' \mid \langle \ell | Av \rangle = 0 \text{ for all } v \in \mathcal{V}\} \\ (\ker A^*)^\perp &= \overline{\text{ran } A}. \end{aligned}$$

When considering a linear map $A: \mathcal{H} \rightarrow \mathcal{K}$ between Hilbert spaces \mathcal{H} and \mathcal{K} , we can appeal to the Riesz representation theorem to identify \mathcal{H}' with \mathcal{H} , \mathcal{K}' with \mathcal{K} , and hence define the adjoint in terms of inner products:

$$\langle A^*k, h \rangle_{\mathcal{H}} = \langle k, Ah \rangle_{\mathcal{K}} \quad \text{for all } h \in \mathcal{H} \text{ and } k \in \mathcal{K}.$$

With this simplification, we can add to Proposition 3.16 the additional properties that $A^{**} = A$ and $\|A^*A\| = \|AA^*\| = \|A\|^2$. Also, in the Hilbert space setting, a linear map $A: \mathcal{H} \rightarrow \mathcal{H}$ is said to be *self-adjoint* if $A = A^*$. A self-adjoint map A is said to be *positive semi-definite* if

$$\inf_{\substack{x \in \mathcal{H} \\ x \neq 0}} \frac{\langle x, Ax \rangle}{\|x\|^2} \geq 0,$$

and *positive definite* if this inequality is strict.

Given a basis $\{e_i\}_{i \in I}$ of \mathcal{H} , the corresponding *dual basis* $\{e_i\}_{i \in I}$ of \mathcal{H} is defined by the relation $\langle e^i, e_j \rangle_{\mathcal{H}} = \delta_{ij}$. The matrix of A with respect to bases $\{e_i\}_{i \in I}$ of \mathcal{H} and $\{f_j\}_{j \in J}$ of \mathcal{K} and the matrix of A^* with respect to the corresponding dual bases are very simply related: the one is the conjugate transpose of the other, and so by abuse of terminology the conjugate transpose of a matrix is often referred to as the adjoint.

Thus, self-adjoint bounded linear maps are the appropriate generalization to Hilbert spaces of symmetric matrices over \mathbb{R} or Hermitian matrices over \mathbb{C} . They are also particularly useful in probability because the covariance operator of an \mathcal{H} -valued random variable is a self-adjoint (and indeed positive semi-definite) bounded linear operator on \mathcal{H} .

3.4 Orthogonality and Direct Sums

Orthogonal decompositions of Hilbert spaces will be fundamental tools in many of the methods considered later on.

Definition 3.17. A subset E of an inner product space \mathcal{V} is said to be *orthogonal* if $\langle x, y \rangle = 0$ for all distinct elements $x, y \in E$; it is said to be *orthonormal* if

$$\langle x, y \rangle = \begin{cases} 1, & \text{if } x = y \in E, \\ 0, & \text{if } x, y \in E \text{ and } x \neq y. \end{cases}$$

Lemma 3.18 (Gram–Schmidt). *Let $(x_n)_{n \in \mathbb{N}}$ be any sequence in an inner product space \mathcal{V} , with the first $d \in \mathbb{N}_0 \cup \{\infty\}$ terms linearly independent. Inductively define $(u_n)_{n \in \mathbb{N}}$ and $(e_n)_{n \in \mathbb{N}}$ by*

$$u_n := x_n - \sum_{k=1}^{n-1} \frac{\langle x_n, u_k \rangle}{\|u_k\|^2} u_k,$$

$$e_n := \frac{u_n}{\|u_n\|}$$

Then $(u_n)_{n \in \mathbb{N}}$ (resp. $(e_n)_{n \in \mathbb{N}}$) is a sequence of d orthogonal (resp. orthonormal) elements of \mathcal{V} , followed by zeros if $d < \infty$.

Definition 3.19. The *orthogonal complement* E^\perp of a subset E of an inner product space \mathcal{V} is

$$E^\perp := \{y \in \mathcal{V} \mid \text{for all } x \in E, \langle y, x \rangle = 0\}.$$

The orthogonal complement of $E \subseteq \mathcal{V}$ is always a closed linear subspace of \mathcal{V} , and hence if $\mathcal{V} = \mathcal{H}$ is a Hilbert space, then E^\perp is also a Hilbert space in its own right.

Theorem 3.20. *Let \mathcal{K} be a closed subspace of a Hilbert space \mathcal{H} . Then, for any $x \in \mathcal{H}$, there is a unique $\Pi_{\mathcal{K}}x \in \mathcal{K}$ that is closest to x in the sense that*

$$\|\Pi_{\mathcal{K}}x - x\| = \inf_{y \in \mathcal{K}} \|y - x\|.$$

Furthermore, x can be written uniquely as $x = \Pi_{\mathcal{K}}x + z$, where $z \in \mathcal{K}^\perp$. Hence, \mathcal{H} decomposes as the orthogonal direct sum

$$\mathcal{H} = \mathcal{K} \oplus^\perp \mathcal{K}^\perp.$$

Theorem 3.20 can be seen as a special case of closest-point approximation among convex sets: see Lemma 4.25 and Exercise 4.2. The operator $\Pi_{\mathcal{K}}: \mathcal{H} \rightarrow \mathcal{K}$ is called the *orthogonal projection* onto \mathcal{K} .

Theorem 3.21. Let \mathcal{K} be a closed subspace of a Hilbert space \mathcal{H} . The corresponding orthogonal projection operator $\Pi_{\mathcal{K}}$ is

(a) a continuous linear operator of norm at most 1;

(b) with $I - \Pi_{\mathcal{K}} = \Pi_{\mathcal{K}^\perp}$;

and satisfies, for every $x \in \mathcal{H}$,

(c) $\|x\|^2 = \|\Pi_{\mathcal{K}}x\|^2 + \|(I - \Pi_{\mathcal{K}})x\|^2$;

(d) $\Pi_{\mathcal{K}}x = x \iff x \in \mathcal{K}$;

(e) $\Pi_{\mathcal{K}}x = 0 \iff x \in \mathcal{K}^\perp$.

Example 3.22 (Conditional expectation). An important probabilistic application of orthogonal projection is the operation of conditioning a random variable. Let $(\Theta, \mathcal{F}, \mu)$ be a probability space and let $X \in L^2(\Theta, \mathcal{F}, \mu; \mathbb{K})$ be a square-integrable random variable. If $\mathcal{G} \subseteq \mathcal{F}$ is a σ -algebra, then the *conditional expectation* of X with respect to \mathcal{G} , usually denoted $\mathbb{E}[X|\mathcal{G}]$, is the orthogonal projection of X onto the subspace $L^2(\Theta, \mathcal{G}, \mu; \mathbb{K})$. In elementary contexts, \mathcal{G} is usually taken to be the σ -algebra generated by a single event E of positive μ -probability, i.e.

$$\mathcal{G} = \{\emptyset, [X \in E], [X \notin E], \Theta\};$$

or even the trivial σ -algebra $\{\emptyset, \Theta\}$, for which the only measurable functions are the constant functions, and hence the conditional expectation coincides with the usual expectation. The orthogonal projection point of view makes two important properties of conditional expectation intuitively obvious:

(a) Whenever $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$, $L^2(\Theta, \mathcal{G}_1, \mu; \mathbb{K})$ is a subspace of $L^2(\Theta, \mathcal{G}_2, \mu; \mathbb{K})$ and composition of the orthogonal projections onto these subspace yields the *tower rule* for conditional expectations:

$$\mathbb{E}[X|\mathcal{G}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1],$$

and, in particular, taking \mathcal{G}_1 to be the trivial σ -algebra $\{\emptyset, \Theta\}$,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]].$$

(b) Whenever $X, Y \in L^2(\Theta, \mathcal{F}, \mu; \mathbb{K})$ and X is, in fact, \mathcal{G} -measurable,

$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}].$$

Direct Sums. Suppose that \mathcal{V} and \mathcal{W} are vector spaces over a common field \mathbb{K} . The Cartesian product $\mathcal{V} \times \mathcal{W}$ can be given the structure of a vector space over \mathbb{K} by defining the operations componentwise:

$$\begin{aligned} (v, w) + (v', w') &:= (v + v', w + w'), \\ \alpha(v, w) &:= (\alpha v, \alpha w), \end{aligned}$$

for all $v, v' \in \mathcal{V}$, $w, w' \in \mathcal{W}$, and $\alpha \in \mathbb{K}$. The resulting vector space is called the (*algebraic*) *direct sum* of \mathcal{V} and \mathcal{W} and is usually denoted by $\mathcal{V} \oplus \mathcal{W}$, while elements of $\mathcal{V} \oplus \mathcal{W}$ are usually denoted by $v \oplus w$ instead of (v, w) .

If $\{e_i | i \in I\}$ is a basis of \mathcal{V} and $\{e_j | j \in J\}$ is a basis of \mathcal{W} , then $\{e_k | k \in K := I \uplus J\}$ is basis of $\mathcal{V} \oplus \mathcal{W}$. Hence, the dimension of $\mathcal{V} \oplus \mathcal{W}$ over \mathbb{K} is equal to the sum of the dimensions of \mathcal{V} and \mathcal{W} .

When \mathcal{H} and \mathcal{K} are Hilbert spaces, their (algebraic) direct sum $\mathcal{H} \oplus \mathcal{K}$ can be given a Hilbert space structure by defining

$$\langle h \oplus k, h' \oplus k' \rangle_{\mathcal{H} \oplus \mathcal{K}} := \langle h, h' \rangle_{\mathcal{H}} + \langle k, k' \rangle_{\mathcal{K}}$$

for all $h, h' \in \mathcal{H}$ and $k, k' \in \mathcal{K}$. The original spaces \mathcal{H} and \mathcal{K} embed into $\mathcal{H} \oplus \mathcal{K}$ as the subspaces $\mathcal{H} \oplus \{0\}$ and $\{0\} \oplus \mathcal{K}$ respectively, and these two subspaces are mutually orthogonal. For this reason, the orthogonality of the two summands in a Hilbert direct sum is sometimes emphasized by the notation $\mathcal{H} \overset{\perp}{\oplus} \mathcal{K}$. The Hilbert space projection theorem (Theorem 3.20) was the statement that whenever \mathcal{K} is a closed subspace of a Hilbert space \mathcal{H} , $\mathcal{H} = \mathcal{K} \overset{\perp}{\oplus} \mathcal{K}^\perp$.

It is necessary to be a bit more careful in defining the direct sum of countably many Hilbert spaces. Let \mathcal{H}_n be a Hilbert space over \mathbb{K} for each $n \in \mathbb{N}$. Then the Hilbert space direct sum $\mathcal{H} := \bigoplus_{n \in \mathbb{N}} \mathcal{H}_n$ is defined to be

$$\mathcal{H} := \overline{\left\{ x = (x_n)_{n \in \mathbb{N}} \mid \begin{array}{l} x_n \in \mathcal{H}_n \text{ for each } n \in \mathbb{N}, \text{ and} \\ x_n = 0 \text{ for all but finitely many } n \end{array} \right\}},$$

where the completion³ is taken with respect to the inner product

$$\langle x, y \rangle_{\mathcal{H}} := \sum_{n \in \mathbb{N}} \langle x_n, y_n \rangle_{\mathcal{H}_n},$$

which is always a finite sum when applied to elements of the generating set. This construction ensures that every element x of \mathcal{H} has finite norm $\|x\|_{\mathcal{H}}^2 = \sum_{n \in \mathbb{N}} \|x_n\|_{\mathcal{H}_n}^2$. As before, each of the summands \mathcal{H}_n is a subspace of \mathcal{H} that is orthogonal to all the others.

Orthogonal direct sums and orthogonal bases are among the most important constructions in Hilbert space theory, and will be very useful in what follows. Prototypical examples include the standard ‘Euclidean’ basis of ℓ^2 and the *Fourier basis* $\{e_n | n \in \mathbb{Z}\}$ of $L^2(\mathbb{S}^1; \mathbb{C})$, where

$$e_n(x) := \frac{1}{2\pi} \exp(inx).$$

³ Completions of normed spaces are formed in the same way as the completion of \mathbb{Q} to form \mathbb{R} : the completion is the space of equivalence classes of Cauchy sequences, with sequences whose difference tends to zero in norm being regarded as equivalent.

Indeed, Fourier's claim⁴ that any periodic function f could be written as

$$f(x) = \sum_{n \in \mathbb{Z}} \widehat{f}_n e_n(x),$$

$$\widehat{f}_n := \int_{\mathbb{S}^1} f(y) \overline{e_n(y)} \, dy,$$

can be seen as one of the historical drivers behind the development of much of analysis. For the purposes of this book's treatment of UQ, key examples of an orthogonal bases are given by *orthogonal polynomials*, which will be considered at length in Chapter 8.

Some important results about orthogonal systems are summarized below; classically, many of these results arose in the study of Fourier series, but hold for any orthonormal basis of a general Hilbert space.

Lemma 3.23 (Bessel's inequality). *Let \mathcal{V} be an inner product space and $(e_n)_{n \in \mathbb{N}}$ an orthonormal sequence in \mathcal{V} . Then, for any $x \in \mathcal{V}$, the series $\sum_{n \in \mathbb{N}} |\langle e_n, x \rangle|^2$ converges and satisfies*

$$\sum_{n \in \mathbb{N}} |\langle e_n, x \rangle|^2 \leq \|x\|^2. \quad (3.19)$$

Theorem 3.24 (Parseval identity). *Let $(e_n)_{n \in \mathbb{N}}$ be an orthonormal sequence in a Hilbert space \mathcal{H} , and let $(\alpha_n)_{n \in \mathbb{N}}$ be a sequence in \mathbb{K} . Then the series $\sum_{n \in \mathbb{N}} \alpha_n e_n$ converges in \mathcal{H} if and only if the series $\sum_{n \in \mathbb{N}} |\alpha_n|^2$ converges in \mathbb{R} , in which case*

$$\left\| \sum_{n \in \mathbb{N}} \alpha_n e_n \right\|^2 = \sum_{n \in \mathbb{N}} |\alpha_n|^2. \quad (3.20)$$

Hence, for any $x \in \mathcal{H}$, the series $\sum_{n \in \mathbb{N}} \langle e_n, x \rangle e_n$ converges.

Theorem 3.25. *Let $(e_n)_{n \in \mathbb{N}}$ be an orthonormal sequence in a Hilbert space \mathcal{H} . Then the following are equivalent:*

- (a) $\{e_n \mid n \in \mathbb{N}\}^\perp = \{0\}$;
- (b) $\mathcal{H} = \text{span}\{e_n \mid n \in \mathbb{N}\}$;
- (c) $\mathcal{H} = \bigoplus_{n \in \mathbb{N}} \mathbb{K}e_n$ as a direct sum of Hilbert spaces;
- (d) for all $x \in \mathcal{H}$, $\|x\|^2 = \sum_{n \in \mathbb{N}} |\langle e_n, x \rangle|^2$;
- (e) for all $x \in \mathcal{H}$, $x = \sum_{n \in \mathbb{N}} \langle e_n, x \rangle e_n$.

If one (and hence all) of these conditions holds true, then $(e_n)_{n \in \mathbb{N}}$ is called a complete orthonormal basis for \mathcal{H}

⁴ Of course, Fourier did not use the modern notation of Hilbert spaces! Furthermore, if he had, then it would have been 'obvious' that his claim could only hold true for L^2 functions and in the L^2 sense, not pointwise for arbitrary functions.

Corollary 3.26. *Let $(e_n)_{n \in \mathbb{N}}$ be a complete orthonormal basis for a Hilbert space \mathcal{H} . For every $x \in \mathcal{H}$, the truncation error $x - \sum_{n=1}^N \langle e_n, x \rangle e_n$ is orthogonal to $\text{span}\{e_1, \dots, e_N\}$.*

Proof. Let $v := \sum_{m=1}^N v_m e_m \in \text{span}\{e_1, \dots, e_N\}$ be arbitrary. By completeness,

$$x = \sum_{n \in \mathbb{N}} \langle e_n, x \rangle e_n.$$

Hence,

$$\begin{aligned} \left\langle x - \sum_{n=1}^N \langle e_n, x \rangle e_n, v \right\rangle &= \left\langle \sum_{n > N} \langle e_n, x \rangle e_n, \sum_{m=1}^N v_m e_m \right\rangle \\ &= \sum_{\substack{n > N \\ m \in \{0, \dots, N\}}} \langle \langle e_n, x \rangle e_n, v_m e_m \rangle \\ &= \sum_{\substack{n > N \\ m \in \{0, \dots, N\}}} \langle x, e_n \rangle v_m \langle e_n, e_m \rangle \\ &= 0 \end{aligned}$$

since $\langle e_n, e_m \rangle = \delta_{nm}$, and $m \neq n$ in the double sum. \square

Remark 3.27. The results cited above (in particular, Theorems 3.20, 3.21, and 3.25, and Corollary 3.26) imply that if we wish to find the closest point of $\text{span}\{e_1, \dots, e_N\}$ to some $x = \sum_{n \in \mathbb{N}} \langle e_n, x \rangle e_n$, then this is a simple matter of series truncation: the optimal approximation is $x \approx x^{(N)} := \sum_{n=1}^N \langle e_n, x \rangle e_n$. Furthermore, this operation is a continuous linear operation as a function of x , and if it is desired to improve the quality of an approximation $x \approx x^{(N)}$ in $\text{span}\{e_1, \dots, e_N\}$ to an approximation in, say, $\text{span}\{e_1, \dots, e_{N+1}\}$, then the improvement is a simple matter of calculating $\langle e_{N+1}, x \rangle$ and adjoining the new term $\langle e_{N+1}, x \rangle e_{N+1}$ to form a new norm-optimal approximation

$$x \approx x^{(N+1)} := \sum_{n=1}^{N+1} \langle e_n, x \rangle e_n = x^{(N)} + \langle e_{N+1}, x \rangle e_{N+1}.$$

However, in Banach spaces (even finite-dimensional ones), closest-point approximation is not as simple as series truncation, and the improvement of approximations is not as simple as adjoining new terms: see Exercise 3.4.

3.5 Tensor Products

The heuristic definition of the tensor product $\mathcal{V} \otimes \mathcal{W}$ of two vector spaces \mathcal{V} and \mathcal{W} over a common field \mathbb{K} is that it is the vector space over \mathbb{K} with basis given by the formal symbols $\{e_i \otimes f_j \mid i \in I, j \in J\}$, where $\{e_i \mid i \in I\}$ is a basis of \mathcal{V} and $\{f_j \mid j \in J\}$ is a basis of \mathcal{W} . Alternatively, we might say that elements of $\mathcal{V} \otimes \mathcal{W}$ are elements of \mathcal{W} with \mathcal{V} -valued rather than \mathbb{K} -valued coefficients (or elements of \mathcal{V} with \mathcal{W} -valued coefficients). However, it is not immediately clear that this definition is independent of the bases chosen for \mathcal{V} and \mathcal{W} . A more thorough definition is as follows.

Definition 3.28. The *free vector space* $F_{\mathcal{V} \times \mathcal{W}}$ on the Cartesian product $\mathcal{V} \times \mathcal{W}$ is defined by taking the vector space in which the elements of $\mathcal{V} \times \mathcal{W}$ are a basis:

$$F_{\mathcal{V} \times \mathcal{W}} := \left\{ \sum_{i=1}^n \alpha_i e_{(v_i, w_i)} \mid \begin{array}{l} n \in \mathbb{N} \text{ and, for } i = 1, \dots, n, \\ \alpha_i \in \mathbb{K}, (v_i, w_i) \in \mathcal{V} \times \mathcal{W} \end{array} \right\}.$$

The ‘freeness’ of $F_{\mathcal{V} \times \mathcal{W}}$ is that the elements $e_{(v, w)}$ are, by definition, linearly independent for distinct pairs $(v, w) \in \mathcal{V} \times \mathcal{W}$; even $e_{(v, 0)}$ and $e_{(-v, 0)}$ are linearly independent. Now define an equivalence relation \sim on $F_{\mathcal{V} \times \mathcal{W}}$ such that

$$\begin{aligned} e_{(v+v', w)} &\sim e_{(v, w)} + e_{(v', w)}, \\ e_{(v, w+w')} &\sim e_{(v, w)} + e_{(v, w')}, \\ \alpha e_{(v, w)} &\sim e_{(\alpha v, w)} \sim e_{(v, \alpha w)} \end{aligned}$$

for arbitrary $v, v' \in \mathcal{V}$, $w, w' \in \mathcal{W}$, and $\alpha \in \mathbb{K}$. Let R be the subspace of $F_{\mathcal{V} \times \mathcal{W}}$ generated by these equivalence relations, i.e. the equivalence class of $e_{(0, 0)}$.

Definition 3.29. The (*algebraic*) *tensor product* $\mathcal{V} \otimes \mathcal{W}$ is the quotient space

$$\mathcal{V} \otimes \mathcal{W} := \frac{F_{\mathcal{V} \times \mathcal{W}}}{R}.$$

One can easily check that $\mathcal{V} \otimes \mathcal{W}$, as defined in this way, is indeed a vector space over \mathbb{K} . The subspace R of $F_{\mathcal{V} \times \mathcal{W}}$ is mapped to the zero element of $\mathcal{V} \otimes \mathcal{W}$ under the quotient map, and so the above equivalences become equalities in the tensor product space:

$$\begin{aligned} (v + v') \otimes w &= v \otimes w + v' \otimes w, \\ v \otimes (w + w') &= v \otimes w + v \otimes w', \\ \alpha(v \otimes w) &= (\alpha v) \otimes w = v \otimes (\alpha w) \end{aligned}$$

for all $v, v' \in \mathcal{V}$, $w, w' \in \mathcal{W}$, and $\alpha \in \mathbb{K}$.

One can also check that the heuristic definition in terms of bases holds true under the formal definition: if $\{e_i | i \in I\}$ is a basis of \mathcal{V} and $\{f_j | j \in J\}$ is a basis of \mathcal{W} , then $\{e_i \otimes f_j | i \in I, j \in J\}$ is a basis of $\mathcal{V} \otimes \mathcal{W}$. Hence, the dimension of the tensor product is the product of dimensions of the original spaces.

Definition 3.30. The *Hilbert space tensor product* of two Hilbert spaces \mathcal{H} and \mathcal{K} over the same field \mathbb{K} is given by defining an inner product on the algebraic tensor product $\mathcal{H} \otimes \mathcal{K}$ by

$$\langle h \otimes k, h' \otimes k' \rangle_{\mathcal{H} \otimes \mathcal{K}} := \langle h, h' \rangle_{\mathcal{H}} \langle k, k' \rangle_{\mathcal{K}} \quad \text{for all } h, h' \in \mathcal{H} \text{ and } k, k' \in \mathcal{K},$$

extending this definition to all of the algebraic tensor product by sesquilinearity, and defining the Hilbert space tensor product $\mathcal{H} \otimes \mathcal{K}$ to be the completion of the algebraic tensor product with respect to this inner product and its associated norm.

Tensor products of Hilbert spaces arise very naturally when considering spaces of functions of more than one variable, or spaces of functions that take values in other function spaces. A prime example of the second type is a space of stochastic processes.

Example 3.31. (a) Given two measure spaces $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$, consider $L^2(\mathcal{X} \times \mathcal{Y}, \mu \otimes \nu; \mathbb{K})$, the space of functions on $\mathcal{X} \times \mathcal{Y}$ that are square integrable with respect to the product measure $\mu \otimes \nu$. If $f \in L^2(\mathcal{X}, \mu; \mathbb{K})$ and $g \in L^2(\mathcal{Y}, \nu; \mathbb{K})$, then we can define a function $h: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{K}$ by $h(x, y) := f(x)g(y)$. The definition of the product measure ensures that $h \in L^2(\mathcal{X} \times \mathcal{Y}, \mu \otimes \nu; \mathbb{K})$, so this procedure defines a bilinear mapping $L^2(\mathcal{X}, \mu; \mathbb{K}) \times L^2(\mathcal{Y}, \nu; \mathbb{K}) \rightarrow L^2(\mathcal{X} \times \mathcal{Y}, \mu \otimes \nu; \mathbb{K})$. It turns out that the span of the range of this bilinear map is dense in $L^2(\mathcal{X} \times \mathcal{Y}, \mu \otimes \nu; \mathbb{K})$ if $L^2(\mathcal{X}, \mu; \mathbb{K})$ and $L^2(\mathcal{Y}, \nu; \mathbb{K})$ are separable. This shows that

$$L^2(\mathcal{X}, \mu; \mathbb{K}) \otimes L^2(\mathcal{Y}, \nu; \mathbb{K}) \cong L^2(\mathcal{X} \times \mathcal{Y}, \mu \otimes \nu; \mathbb{K}),$$

and it also explains why it is necessary to take the completion in the construction of the Hilbert space tensor product.

(b) Similarly, $L^2(\mathcal{X}, \mu; \mathcal{H})$, the space of functions $f: \mathcal{X} \rightarrow \mathcal{H}$ that are square integrable in the sense that

$$\int_{\mathcal{X}} \|f(x)\|_{\mathcal{H}}^2 d\mu(x) < +\infty,$$

is isomorphic to $L^2(\mathcal{X}, \mu; \mathbb{K}) \otimes \mathcal{H}$ if this space is separable. The isomorphism maps $f \otimes \varphi \in L^2(\mathcal{X}, \mu; \mathbb{K}) \otimes \mathcal{H}$ to the \mathcal{H} -valued function $x \mapsto f(x)\varphi$ in $L^2(\mathcal{X}, \mu; \mathcal{H})$.

(c) Combining the previous two examples reveals that

$$L^2(\mathcal{X}, \mu; \mathbb{K}) \otimes L^2(\mathcal{Y}, \nu; \mathbb{K}) \cong L^2(\mathcal{X} \times \mathcal{Y}, \mu \otimes \nu; \mathbb{K}) \cong L^2(\mathcal{X}, \mu; L^2(\mathcal{Y}, \nu; \mathbb{K})).$$

Similarly, one can consider a *Bochner space* $L^p(\mathcal{X}, \mu; \mathcal{V})$ of functions (random variables) taking values in a Banach space \mathcal{V} that are p^{th} -power-integrable in the sense that $\int_{\mathcal{X}} \|f(x)\|_{\mathcal{V}}^p d\mu(x)$ is finite, and identify this space with a suitable tensor product $L^p(\mathcal{X}, \mu; \mathbb{R}) \otimes \mathcal{V}$. However, several subtleties arise in doing this, as there is no single ‘natural’ Banach tensor product of Banach spaces as there is for Hilbert spaces.

3.6 Bibliography

Reference texts on elementary functional analysis, including Banach and Hilbert space theory, include the books of Reed and Simon (1972), Rudin (1991), and Rynne and Youngson (2008). The article of Deutsch (1982) gives a good overview of closest-point approximation properties for subspaces of Banach spaces. Further discussion of the relationship between tensor products and spaces of vector-valued integrable functions can be found in the books of Ryan (2002) and Hackbusch (2012); the former is essentially a pure mathematical text, whereas the latter also includes significant treatment of numerical and computational matters. The Sobolev embedding theorem (Theorem 3.12) and its proof can be found in Evans (2010, Section 5.6, Theorem 6).

Intrepid students may wish to consult Bourbaki (1987), but the standard warnings about Bourbaki texts apply: the presentation is comprehensive but often forbiddingly austere, and so it is perhaps better as a reference text than a learning tool. On the other hand, the *Hitchhiker’s Guide* of Aliprantis and Border (2006) is a surprisingly readable encyclopaedic text.

3.7 Exercises

Exercise 3.1 (Formulae for the operator norm). Let $A: \mathcal{V} \rightarrow \mathcal{W}$ be a linear map between normed vector spaces $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ and $(\mathcal{W}, \|\cdot\|_{\mathcal{W}})$. Show that the operator norm $\|A\|_{\mathcal{V} \rightarrow \mathcal{W}}$ of A is equivalently defined by any of the following expressions:

$$\begin{aligned} \|A\|_{\mathcal{V} \rightarrow \mathcal{W}} &= \sup_{0 \neq v \in \mathcal{V}} \frac{\|Av\|_{\mathcal{W}}}{\|v\|_{\mathcal{V}}} \\ &= \sup_{\|v\|_{\mathcal{V}}=1} \frac{\|Av\|_{\mathcal{W}}}{\|v\|_{\mathcal{V}}} = \sup_{\|v\|_{\mathcal{V}}=1} \|Av\|_{\mathcal{W}} \\ &= \sup_{0 < \|v\|_{\mathcal{V}} \leq 1} \frac{\|Av\|_{\mathcal{W}}}{\|v\|_{\mathcal{V}}} = \sup_{\|v\|_{\mathcal{V}} \leq 1} \|Av\|_{\mathcal{W}} \\ &= \sup_{0 < \|v\|_{\mathcal{V}} < 1} \frac{\|Av\|_{\mathcal{W}}}{\|v\|_{\mathcal{V}}} = \sup_{\|v\|_{\mathcal{V}} < 1} \|Av\|_{\mathcal{W}}. \end{aligned}$$

Exercise 3.2 (Properties of the operator norm). Suppose that \mathcal{U} , \mathcal{V} , and \mathcal{W} are normed vector spaces, and let $A: \mathcal{U} \rightarrow \mathcal{V}$ and $B: \mathcal{V} \rightarrow \mathcal{W}$ be bounded linear maps. Prove that the operator norm is

(a) *compatible* (or *consistent*) with $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$: for all $x \in \mathcal{U}$,

$$\|Au\|_{\mathcal{V}} \leq \|A\|_{\mathcal{U} \rightarrow \mathcal{V}} \|u\|_{\mathcal{U}}.$$

(b) *sub-multiplicative*: $\|B \circ A\|_{\mathcal{U} \rightarrow \mathcal{W}} \leq \|B\|_{\mathcal{V} \rightarrow \mathcal{W}} \|A\|_{\mathcal{U} \rightarrow \mathcal{V}}$.

Exercise 3.3 (Definiteness of the Gram matrix). Let \mathcal{V} be a vector space over \mathbb{K} , equipped with a semi-definite inner product $\langle \cdot, \cdot \rangle$ (i.e. one satisfying all the requirements of Definition 3.2 except possibly positive definiteness). Given vectors $v_1, \dots, v_n \in \mathcal{V}$, the associated *Gram matrix* is

$$G(v_1, \dots, v_n) := \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{bmatrix}.$$

- (a) Show that, in the case that $\mathcal{V} = \mathbb{K}^n$ with its usual inner product, $G(v_1, \dots, v_n) = V^*V$, where V is the matrix with the vectors v_i as its columns, and V^* denotes the conjugate transpose of V .
- (b) Show that $G(v_1, \dots, v_n)$ is a conjugate-symmetric (a.k.a. Hermitian) matrix, and hence is symmetric in the case $\mathbb{K} = \mathbb{R}$.
- (c) Show that $\det G(v_1, \dots, v_n) \geq 0$. Show also that $\det G(v_1, \dots, v_n) = 0$ if v_1, \dots, v_n are linearly dependent, and that this is an ‘if and only if’ if $\langle \cdot, \cdot \rangle$ is positive definite.
- (d) Using the case $n = 2$, prove the Cauchy–Schwarz inequality (3.1).

Exercise 3.4 (Closest-point approximation in Banach spaces). Let $R_\theta: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denote the linear map that is rotation of the Euclidean plane about the origin through a fixed angle $-\frac{\pi}{4} < \theta < \frac{\pi}{4}$. Define a Banach norm $\|\cdot\|_\theta$ on \mathbb{R}^2 in terms of R_θ and the usual 1-norm by

$$\|(x, y)\|_\theta := \|R_\theta(x, y)\|_1.$$

Find the closest point of the x -axis to the point $(1, 1)$, i.e. find $x' \in \mathbb{R}$ to minimize $\|(x', 0) - (1, 1)\|_\theta$; in particular, show that the closest point is *not* $(1, 0)$. Hint: sketch some norm balls centred on $(1, 1)$.

Exercise 3.5 (Series in normed spaces). Many UQ methods involve series expansions in spaces of deterministic functions and/or random variables, so it is useful to understand when such series converge. Let $(v_n)_{n \in \mathbb{N}}$ be a sequence in a normed space \mathcal{V} . As in \mathbb{R} , we say that the series $\sum_{n \in \mathbb{N}} v_n$ converges to $v \in \mathcal{V}$ if the sequence of partial sums converges to v , i.e. if, for all $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that

$$N \geq N_\varepsilon \implies \left\| v - \sum_{n=1}^N v_n \right\| < \varepsilon.$$

- (a) Suppose that $\sum_{n \in \mathbb{N}} v_n$ converges *absolutely* to $v \in \mathcal{V}$, i.e. the series converges and also $\sum_{n \in \mathbb{N}} \|v_n\|$ is finite. Prove the infinite triangle inequality

$$\|v\| \leq \sum_{n \in \mathbb{N}} \|v_n\|.$$

- (b) Suppose that $\sum_{n \in \mathbb{N}} v_n$ converges absolutely to $v \in \mathcal{V}$. Show that $\sum_{n \in \mathbb{N}} v_n$ converges *unconditionally* to $v \in \mathcal{V}$, i.e. $\sum_{n \in \mathbb{N}} v_{\pi(n)}$ converges to $x \in \mathcal{V}$ for every bijection $\pi: \mathbb{N} \rightarrow \mathbb{N}$. Thus, the order of summation ‘does not matter’. (Note that the converse of this result is false: Dvoretzky and Rogers (1950) showed that every infinite-dimensional Banach space contains series that converge unconditionally but not absolutely.)



- (c) Suppose that \mathcal{V} is a Banach space and that $\sum_{n \in \mathbb{N}} \|v_n\|$ is finite. Show that $\sum_{n \in \mathbb{N}} v_n$ converges to some $v \in \mathcal{V}$.

Exercise 3.6 (Weierstrass M -test). Let S be any set, let \mathcal{V} be a Banach space, and, for each $n \in \mathbb{N}$, let $f_n: S \rightarrow \mathcal{V}$. Suppose that M_n is such that

$$\|f_n(x)\| \leq M_n \quad \text{for all } x \in S \text{ and } n \in \mathbb{N},$$

and that $\sum_{n \in \mathbb{N}} M_n$ is finite. Show that the series $\sum_{n \in \mathbb{N}} f_n$ converges *uniformly* on S , i.e. there exists $f: S \rightarrow \mathcal{V}$ such that, for all $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ so that

$$N \geq N_\varepsilon \implies \sup_{x \in S} \left\| f(x) - \sum_{n=1}^N f_n(x) \right\| < \varepsilon.$$