Hilbert Spaces

Jean Gallier and Jocelyn Quaintance
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@cis.upenn.edu

March 10, 2017

Contents

Chapter 1

A Glimpse at Hilbert Spaces

1.1 The Projection Lemma, Duality

Given a Hermitian space $\langle E, \varphi \rangle$, we showed in Section **??** that the function $\| \, \| \colon E \to \mathbb{R}$ defined such that $\|u\| = \sqrt{\varphi(u,u)}$, is a norm on E. Thus, E is a normed vector space. If E is also complete, then it is a very interesting space.

Recall that completeness has to do with the convergence of Cauchy sequences. A normed vector space $\langle E, \| \, \| \rangle$ is automatically a metric space under the metric d defined such that $d(u,v) = \|v - u\|$ (see Chapter **??** for the definition of a normed vector space and of a metric space, or Lang [5, 6], or Dixmier [3]). Given a metric space E with metric d, a sequence $(a_n)_{n \geq 1}$ of elements $a_n \in E$ is a *Cauchy sequence* iff for every $\epsilon > 0$, there is some $N \geq 1$ such that

$$d(a_m, a_n) < \epsilon \quad \text{for all} \quad m, n \geq N.$$

We say that E is *complete* iff every Cauchy sequence converges to a limit (which is unique, since a metric space is Hausdorff).

Every finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ is complete. For example, one can show by induction that given any basis $(e_1, \ldots, e_n)$ of E, the linear map $h \colon \mathbb{C}^n \to E$ defined such that

$$h((z_1, \ldots, z_n)) = z_1 e_1 + \cdots + z_n e_n$$

is a homeomorphism (using the *sup*-norm on $\mathbb{C}^n$). One can also use the fact that any two norms on a finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ are equivalent (see Chapter **??**, or Lang [6], Dixmier [3], Schwartz [9]).

However, if E has infinite dimension, it may not be complete. When a Hermitian space is complete, a number of the properties that hold for finite dimensional Hermitian spaces also hold for infinite dimensional spaces. For example, any closed subspace has an orthogonal complement, and in particular, a finite dimensional subspace has an orthogonal complement. Hermitian spaces that are also complete play an important role in analysis. Since they were first studied by Hilbert, they are called Hilbert spaces.

Definition 1.1. A (complex) Hermitian space $\langle E, \varphi \rangle$ which is a complete normed vector space under the norm $\| \ \|$ induced by φ is called a *Hilbert space*. A real Euclidean space $\langle E, \varphi \rangle$ which is complete under the norm $\| \ \|$ induced by φ is called a *real Hilbert space*.

All the results in this section hold for complex Hilbert spaces as well as for real Hilbert spaces. We state all results for the complex case only, since they also apply to the real case, and since the proofs in the complex case need a little more care.

Example 1.1. The space l^2 of all countably infinite sequences $x = (x_i)_{i \in \mathbb{N}}$ of complex numbers such that $\sum_{i=0}^{\infty} |x_i|^2 < \infty$ is a Hilbert space. It will be shown later that the map $\varphi \colon l^2 \times l^2 \to \mathbb{C}$ defined such that

$$\varphi\left((x_i)_{i \in \mathbb{N}}, (y_i)_{i \in \mathbb{N}}\right) = \sum_{i=0}^{\infty} x_i \overline{y_i}$$

is well defined, and that l^2 is a Hilbert space under φ. In fact, we will prove a more general result (Proposition 1.11).

Example 1.2. The set $\mathcal{C}^\infty[a, b]$ of smooth functions $f \colon [a, b] \to \mathbb{C}$ is a Hermitian space under the Hermitian form

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx,$$

but it is not a Hilbert space because it is not complete. It is possible to construct its completion $\mathrm{L}^2([a, b])$, which turns out to be the space of Lebesgue integrable functions on $[a, b]$.

Remark: Given a Hermitian space E (with Hermitian product $\langle -, - \rangle$), it is possible to construct a Hilbert space E_h (with Hermitian product $\langle -, - \rangle_h$) and a linear map $j \colon E \to E_h$, such that

$$\langle u, v \rangle = \langle j(u), j(v) \rangle_h$$

for all $u, v \in E$, and $j(E)$ is dense in E_h. Furthermore, E_h is unique up to isomorphism. For details, see Bourbaki [2].

One of the most important facts about finite-dimensional Hermitian (and Euclidean) spaces is that they have orthonormal bases. This implies that, up to isomorphism, every finite-dimensional Hermitian space is isomorphic to $\mathbb{C}^n$ (for some $n \in \mathbb{N}$) and that the inner product is given by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$

Furthermore, every subspace W has an orthogonal complement $W^\perp$, and the inner product induces a natural duality between E and E^*, where E^* is the space of linear forms on E.

When E is a Hilbert space, E may be infinite dimensional, often of uncountable dimension. Thus, we can't expect that E always have an orthonormal basis. However, if we modify the notion of basis so that a "Hilbert basis" is an orthogonal family that is also dense in E, i.e., every $v \in E$ is the limit of a sequence of finite combinations of vectors from the Hilbert basis, then we can recover most of the "nice" properties of finite-dimensional Hermitian spaces. For instance, if $(u_k)_{k\in K}$ is a Hilbert basis, for every $v \in E$, we can define the Fourier coefficients $c_k = \langle v, u_k\rangle/\|u_k\|$, and then, v is the "sum" of its Fourier series $\sum_{k\in K} c_k u_k$. However, the cardinality of the index set K can be very large, and it is necessary to define what it means for a family of vectors indexed by K to be summable. We will do this in Section 1.2. It turns out that every Hilbert space is isomorphic to a space of the form $l^2(K)$, where $l^2(K)$ is a generalization of the space of Example 1.1 (see Theorem 1.16, usually called the Riesz-Fischer theorem).

Our first goal is to prove that a closed subspace of a Hilbert space has an orthogonal complement. We also show that duality holds if we redefine the dual E' of E to be the space of *continuous* linear maps on E. Our presentation closely follows Bourbaki [2]. We also were inspired by Rudin [7], Lang [5, 6], Schwartz [9, 8], and Dixmier [3]. In fact, we highly recommend Dixmier [3] as a clear and simple text on the basics of topology and analysis. We first prove the so-called projection lemma.

Recall that in a metric space E, a subset X of E is *closed* iff for every convergent sequence (x_n) of points $x_n \in X$, the limit $x = \lim_{n\to\infty} x_n$ also belongs to X. The *closure* $\overline{X}$ *of* X is the set of all limits of convergent sequences (x_n) of points $x_n \in X$. Obviously, $X \subseteq \overline{X}$. We say that the subset X of E is *dense in* E iff $E = \overline{X}$, the closure of X, which means that every $a \in E$ is the limit of some sequence (x_n) of points $x_n \in X$. Convex sets will again play a crucial role.

First, we state the following easy "parallelogram inequality", whose proof is left as an exercise.

Proposition 1.1. *If E is a Hermitian space, for any two vectors $u, v \in E$, we have*

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

From the above, we get the following proposition:

Proposition 1.2. *If E is a Hermitian space, given any $d, \delta \in \mathbb{R}$ such that $0 \le \delta < d$, let*

$$B = \{u \in E \mid \|u\| < d\} \quad \textit{and} \quad C = \{u \in E \mid \|u\| \le d + \delta\}.$$

For any convex set such A that $A \subseteq C - B$, we have

$$\|v - u\| \le \sqrt{12d\delta},$$

for all $u, v \in A$ (see Figure 1.1).

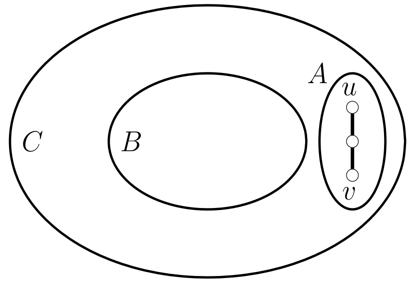

Figure 1.1: Inequality of Proposition 1.2

Proof. Since A is convex, $\frac{1}{2}(u+v) \in A$ if $u, v \in A$, and thus, $\|\frac{1}{2}(u+v)\| \geq d$. From the parallelogram inequality written in the form

$$\left\|\frac{1}{2}(u+v)\right\|^2 + \left\|\frac{1}{2}(u-v)\right\|^2 = \frac{1}{2}\left(\|u\|^2 + \|v\|^2\right),$$

since $\delta < d$, we get

$$\left\|\frac{1}{2}(u-v)\right\|^2 = \frac{1}{2}\left(\|u\|^2 + \|v\|^2\right) - \left\|\frac{1}{2}(u+v)\right\|^2 \leq (d+\delta)^2 - d^2 = 2d\delta + \delta^2 \leq 3d\delta,$$

from which

$$\|v - u\| \leq \sqrt{12d\delta}.$$

□

If X is a nonempty subset of a metric space (E, d), for any $a \in E$, recall that we define the *distance* $d(a, X)$ *of* a *to* X as

$$d(a, X) = \inf_{b \in X} d(a, b).$$

Also, the *diameter* $\delta(X)$ *of* X is defined by

$$\delta(X) = \sup\{d(a, b) \mid a, b \in X\}.$$

It is possible that $\delta(X) = \infty$. We leave the following standard two facts as an exercise (see Dixmier [3]):

Proposition 1.3. *Let E be a metric space.*

(1) For every subset $X \subseteq E$, $\delta(X) = \delta(\overline{X})$.

(2) If E is a complete metric space, for every sequence (F_n) of closed nonempty subsets of E such that $F_{n+1} \subseteq F_n$, if $\lim_{n\to\infty} \delta(F_n) = 0$, then $\bigcap_{n=1}^{\infty} F_n$ consists of a single point.

We are now ready to prove the crucial projection lemma.

Proposition 1.4. *(Projection lemma) Let E be a Hilbert space.*

(1) For any nonempty convex and closed subset $X \subseteq E$, for any $u \in E$, there is a unique vector $p_X(u) \in X$ such that

$$\|u - p_X(u)\| = \inf_{v \in X} \|u - v\| = d(u, X).$$

(2) The vector $p_X(u)$ is the unique vector $w \in E$ satisfying the following property (see Figure 1.2):

$$w \in X \quad \text{and} \quad \Re \langle u - w, z - w \rangle \leq 0 \quad \text{for all } z \in X. \tag{$*$}$$

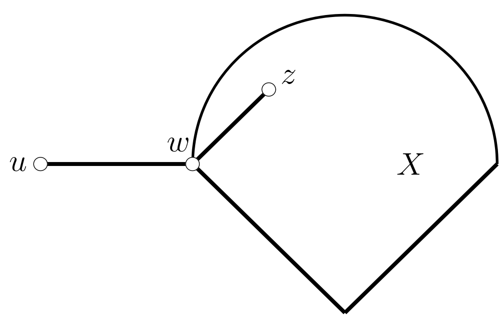

Figure 1.2: Inequality of Proposition 1.4

Proof. (1) Let $d = \inf_{v \in X} \|u - v\| = d(u, X)$. We define a sequence X_n of subsets of X as follows: for every $n \geq 1$,

$$X_n = \left\{ v \in X \mid \|u - v\| \leq d + \frac{1}{n} \right\}.$$

It is immediately verified that each X_n is nonempty (by definition of d), convex, and that $X_{n+1} \subseteq X_n$. Also, by Proposition 1.2, we have

$$\sup\{\|w - v\| \mid v, w \in X_n\} \leq \sqrt{12d/n},$$

and thus, $\bigcap_{n \geq 1} X_n$ contains at most one point. We will prove that $\bigcap_{n \geq 1} X_n$ contains exactly one point, namely, $p_X(u)$. For this, define a sequence $(w_n)_{n \geq 1}$ by picking some $w_n \in X_n$ for every $n \geq 1$. We claim that $(w_n)_{n \geq 1}$ is a Cauchy sequence. Given any $\epsilon > 0$, if we pick N such that

$$N > \frac{12d}{\epsilon^2},$$

since $(X_n)_{n\geq 1}$ is a monotonic decreasing sequence, for all $m, n \geq N$, we have

$$\|w_m - w_n\| \leq \sqrt{12d/N} < \epsilon,$$

as desired. Since E is complete, the sequence $(w_n)_{n\geq 1}$ has a limit w, and since $w_n \in X$ and X is closed, we must have $w \in X$. Also observe that

$$\|u - w\| \leq \|u - w_n\| + \|w_n - w\|,$$

and since w is the limit of $(w_n)_{n\geq 1}$ and

$$\|u - w_n\| \leq d + \frac{1}{n},$$

given any $\epsilon > 0$, there is some n large enough so that

$$\frac{1}{n} < \frac{\epsilon}{2} \quad \text{and} \quad \|w_n - w\| \leq \frac{\epsilon}{2},$$

and thus

$$\|u - w\| \leq d + \epsilon.$$

Since the above holds for every $\epsilon > 0$, we have $\|u - w\| = d$. Thus, $w \in X_n$ for all $n \geq 1$, which proves that $\bigcap_{n\geq 1} X_n = \{w\}$. Now, any $z \in X$ such that $\|u - z\| = d(u, X) = d$ also belongs to every X_n, and thus $z = w$, proving the uniqueness of w, which we denote as $p_X(u)$.

(2) Let $w \in X$. Since X is convex, $z = (1 - \lambda)p_X(u) + \lambda w \in X$ for every λ, $0 \leq \lambda \leq 1$. Then, we have

$$\|u - z\| \geq \|u - p_X(u)\|$$

for all λ, $0 \leq \lambda \leq 1$, and since

$$\begin{aligned}\|u - z\|^2 &= \|u - p_X(u) - \lambda(w - p_X(u))\|^2 \\ &= \|u - p_X(u)\|^2 + \lambda^2\|w - p_X(u)\|^2 - 2\lambda\Re\left\langle u - p_X(u), w - p_X(u)\right\rangle,\end{aligned}$$

for all λ, $0 < \lambda \leq 1$, we get

$$\Re\left\langle u - p_X(u), w - p_X(u)\right\rangle = \frac{1}{2\lambda}\left(\|u - p_X(u)\|^2 - \|u - z\|^2\right) + \frac{\lambda}{2}\|w - p_X(u)\|^2,$$

and since this holds for every λ, $0 < \lambda \leq 1$ and

$$\|u - z\| \geq \|u - p_X(u)\|,$$

we have

$$\Re\left\langle u - p_X(u), w - p_X(u)\right\rangle \leq 0.$$

Conversely, assume that $w \in X$ satisfies the condition

$$\Re \langle u - w, z - w \rangle \leq 0$$

for all $z \in X$. For all $z \in X$, we have

$$\|u - z\|^2 = \|u - w\|^2 + \|z - w\|^2 - 2\Re \langle u - w, z - w \rangle \geq \|u - w\|^2,$$

which implies that $\|u - w\| = d(u, X) = d$, and from (1), that $w = p_X(u)$. $\square$

The vector $p_X(u)$ is called the *projection of* u *onto* X, and the map $p_X\colon E \to X$ is called the *projection of* E *onto* X. In the case of a real Hilbert space, there is an intuitive geometric interpretation of the condition

$$\langle u - p_X(u), z - p_X(u) \rangle \leq 0$$

for all $z \in X$. If we restate the condition as

$$\langle u - p_X(u), p_X(u) - z \rangle \geq 0$$

for all $z \in X$, this says that the absolute value of the measure of the angle between the vectors $u - p_X(u)$ and $p_X(u) - z$ is at most $\pi/2$. This makes sense, since X is convex, and points in X must be on the side opposite to the "tangent space" to X at $p_X(u)$, which is orthogonal to $u - p_X(u)$. Of course, this is only an intuitive description, since the notion of tangent space has not been defined!

The map $p_X\colon E \to X$ is continuous, as shown below.

Proposition 1.5. *Let E be a Hilbert space. For any nonempty convex and closed subset $X \subseteq E$, the map $p_X\colon E \to X$ is continuous.*

Proof. For any two vectors $u, v \in E$, let $x = p_X(u) - u$, $y = p_X(v) - p_X(u)$, and $z = v - p_X(v)$. Clearly,

$$v - u = x + y + z,$$

and from Proposition 1.4 (2), we also have

$$\Re \langle x, y \rangle \geq 0 \quad \text{and} \quad \Re \langle z, y \rangle \geq 0,$$

from which we get

$$\begin{aligned} \|v - u\|^2 &= \|x + y + z\|^2 = \|x + z + y\|^2 \\ &= \|x + z\|^2 + \|y\|^2 + 2\Re \langle x, y \rangle + 2\Re \langle z, y \rangle \\ &\geq \|y\|^2 = \|p_X(v) - p_X(u)\|^2. \end{aligned}$$

However, $\|p_X(v) - p_X(u)\| \leq \|v - u\|$ obviously implies that p_X is continuous. $\square$

We can now prove the following important proposition.

Proposition 1.6. *Let E be a Hilbert space.*

(1) For any closed subspace $V \subseteq E$, we have $E = V \oplus V^{\perp}$, and the map $p_V \colon E \to V$ is linear and continuous.

(2) For any $u \in E$, the projection $p_V(u)$ is the unique vector $w \in E$ such that

$$w \in V \quad \text{and} \quad \langle u - w, z\rangle = 0 \quad \text{for all } z \in V.$$

Proof. (1) First, we prove that $u - p_V(u) \in V^{\perp}$ for all $u \in E$. For any $v \in V$, since V is a subspace, $z = p_V(u) + \lambda v \in V$ for all $\lambda \in \mathbb{C}$, and since V is convex and nonempty (since it is a subspace), and closed by hypothesis, by Proposition 1.4 (2), we have

$$\Re(\overline{\lambda}\,\langle u - p_V(u), v\rangle) = \Re(\langle u - p_V(u), \lambda v\rangle = \Re\,\langle u - p_V(u), z - p_V(u)\rangle \leq 0$$

for all $\lambda \in \mathbb{C}$. In particular, the above holds for $\lambda = \langle u - p_V(u), v\rangle$, which yields

$$|\,\langle u - p_V(u), v\rangle\,| \leq 0,$$

and thus, $\langle u - p_V(u), v\rangle = 0$. As a consequence, $u - p_V(u) \in V^{\perp}$ for all $u \in E$. Since $u = p_V(u) + u - p_V(u)$ for every $u \in E$, we have $E = V + V^{\perp}$. On the other hand, since $\langle -, -\rangle$ is positive definite, $V \cap V^{\perp} = \{0\}$, and thus $E = V \oplus V^{\perp}$.

We already proved in Proposition 1.5 that $p_V \colon E \to V$ is continuous. Also, since

$$p_V(\lambda u + \mu v) - (\lambda p_V(u) + \mu p_V(v)) = p_V(\lambda u + \mu v) - (\lambda u + \mu v) + \lambda(u - p_V(u)) + \mu(v - p_V(v)),$$

for all $u, v \in E$, and since the left-hand side term belongs to V, and from what we just showed, the right-hand side term belongs to $V^{\perp}$, we have

$$p_V(\lambda u + \mu v) - (\lambda p_V(u) + \mu p_V(v)) = 0,$$

showing that p_V is linear.

(2) This is basically obvious from (1). We proved in (1) that $u - p_V(u) \in V^{\perp}$, which is exactly the condition

$$\langle u - p_V(u), z\rangle = 0$$

for all $z \in V$. Conversely, if $w \in V$ satisfies the condition

$$\langle u - w, z\rangle = 0$$

for all $z \in V$, since $w \in V$, every vector $z \in V$ is of the form $y - w$, with $y = z + w \in V$, and thus, we have

$$\langle u - w, y - w\rangle = 0$$

for all $y \in V$, which implies the condition of Proposition 1.4 (2):

$$\Re\,\langle u - w, y - w\rangle \leq 0$$

for all $y \in V$. By Proposition 1.4, $w = p_V(u)$ is the projection of u onto V. $\square$

Let us illustrate the power of Proposition 1.6 on the following "least squares" problem. Given a real $m \times n$-matrix A and some vector $b \in \mathbb{R}^m$, we would like to solve the linear system

$$Ax = b$$

in the least-squares sense, which means that we would like to find some solution $x \in \mathbb{R}^n$ that minimizes the Euclidean norm $\|Ax - b\|$ of the error $Ax - b$. It is actually not clear that the problem has a solution, but it does! The problem can be restated as follows: Is there some $x \in \mathbb{R}^n$ such that

$$\|Ax - b\| = \inf_{y \in \mathbb{R}^n} \|Ay - b\|,$$

or equivalently, is there some $z \in \mathrm{Im}\,(A)$ such that

$$\|z - b\| = d(b, \mathrm{Im}\,(A)),$$

where $\mathrm{Im}\,(A) = \{Ay \in \mathbb{R}^m \mid y \in \mathbb{R}^n\}$, the image of the linear map induced by A. Since $\mathrm{Im}\,(A)$ is a closed subspace of $\mathbb{R}^m$, because we are in finite dimension, Proposition 1.6 tells us that there is a unique $z \in \mathrm{Im}\,(A)$ such that

$$\|z - b\| = \inf_{y \in \mathbb{R}^n} \|Ay - b\|,$$

and thus, the problem always has a solution since $z \in \mathrm{Im}\,(A)$, and since there is at least some $x \in \mathbb{R}^n$ such that $Ax = z$ (by definition of $\mathrm{Im}\,(A)$). Note that such an x is not necessarily unique. Furthermore, Proposition 1.6 also tells us that $z \in \mathrm{Im}\,(A)$ is the solution of the equation

$$\langle z - b, w \rangle = 0 \quad \text{for all } w \in \mathrm{Im}\,(A),$$

or equivalently, that $x \in \mathbb{R}^n$ is the solution of

$$\langle Ax - b, Ay \rangle = 0 \quad \text{for all } y \in \mathbb{R}^n,$$

which is equivalent to

$$\langle A^\top (Ax - b), y \rangle = 0 \quad \text{for all } y \in \mathbb{R}^n,$$

and thus, since the inner product is positive definite, to $A^\top(Ax - b) = 0$, i.e.,

$$A^\top Ax = A^\top b.$$

Therefore, the solutions of the original least-squares problem are precisely the solutions of the the so-called *normal equations*

$$A^\top Ax = A^\top b,$$

discovered by Gauss and Legendre around 1800. We also proved that the normal equations always have a solution.

Computationally, it is best not to solve the normal equations directly, and instead, to use methods such as the QR-decomposition (applied to A) or the SVD-decomposition (in the form of the pseudo-inverse). We will come back to this point later on.

As another corollary of Proposition 1.6, for any continuous nonnull linear map $h\colon E \to \mathbb{C}$, the null space

$$H = \mathrm{Ker}\, h = \{u \in E \mid h(u) = 0\} = h^{-1}(0)$$

is a closed hyperplane H, and thus, $H^{\perp}$ is a subspace of dimension one such that $E = H \oplus H^{\perp}$. This suggests defining the dual space of E as the set of all continuous maps $h\colon E \to \mathbb{C}$.

Remark: If $h\colon E \to \mathbb{C}$ is a linear map which is **not** continuous, then it can be shown that the hyperplane $H = \mathrm{Ker}\, h$ is dense in E! Thus, $H^{\perp}$ is reduced to the trivial subspace $\{0\}$. This goes against our intuition of what a hyperplane in $\mathbb{R}^n$ (or $\mathbb{C}^n$) is, and warns us not to trust our "physical" intuition too much when dealing with infinite dimensions. As a consequence, the map $\flat\colon E \to E^*$ introduced in Section **??** (see just after Definition 1.2 below) is not surjective, since the linear forms of the form $u \mapsto \langle u, v\rangle$ (for some fixed vector $v \in E$) are continuous (the inner product is continuous).

We now show that by redefining the dual space of a Hilbert space as the set of continuous linear forms on E, we recover Theorem **??**.

Definition 1.2. Given a Hilbert space E, we define the *dual space E' of E* as the vector space of all continuous linear forms $h\colon E \to \mathbb{C}$. Maps in E' are also called *bounded linear operators, bounded linear functionals, or simply, operators or functionals*.

As in Section **??**, for all $u, v \in E$, we define the maps $\varphi^l_u\colon E \to \mathbb{C}$ and $\varphi^r_v\colon E \to \mathbb{C}$ such that

$$\varphi^l_u(v) = \overline{\langle u, v\rangle},$$

and

$$\varphi^r_v(u) = \langle u, v\rangle\,.$$

In fact, $\varphi^l_u = \varphi^r_u$, and because the inner product $\langle -, -\rangle$ is continuous, it is obvious that φ^r_v is continuous and linear, so that $\varphi^r_v \in E'$. To simplify notation, we write φ_v instead of φ^r_v.

Theorem **??** is generalized to Hilbert spaces as follows.

Proposition 1.7. *(Riesz representation theorem) Let E be a Hilbert space. Then, the map $\flat\colon E \to E'$ defined such that*

$$\flat(v) = \varphi_v,$$

is semilinear, continuous, and bijective.

Proof. The proof is basically identical to the proof of Theorem **??**, except that a different argument is required for the surjectivity of $\flat\colon E \to E'$, since E may not be finite dimensional. For any nonnull linear operator $h \in E'$, the hyperplane $H = \mathrm{Ker}\, h = h^{-1}(0)$ is a closed subspace of E, and by Proposition 1.6, $H^{\perp}$ is a subspace of dimension one such that $E =$

$H \oplus H^{\perp}$. Then, picking any nonnull vector $w \in H^{\perp}$, observe that H is also the kernel of the linear operator φ_w, with

$$\varphi_w(u) = \langle u, w \rangle ,$$

and thus, since any two nonzero linear forms defining the same hyperplane must be proportional, there is some nonzero scalar $\lambda \in \mathbb{C}$ such that $h = \lambda \varphi_w$. But then, $h = \varphi_{\overline{\lambda} w}$, proving that $\flat \colon E \to E'$ is surjective. $\square$

Proposition 1.7 is known as the *Riesz representation theorem, or "Little Riesz Theorem."* It shows that the inner product on a Hilbert space induces a natural linear isomorphism between E and its dual E'.

Remarks:

(1) Actually, the map $\flat \colon E \to E'$ turns out to be an isometry. To show this, we need to recall the notion of norm of a linear map, which we do not want to do right now.

(2) Many books on quantum mechanics use the so-called Dirac notation to denote objects in the Hilbert space E and operators in its dual space E'. In the Dirac notation, an element of E is denoted as $|x\rangle$, and an element of E' is denoted as $\langle t|$. The scalar product is denoted as $\langle t| \cdot |x\rangle$. This uses the isomorphism between E and E', except that the inner product is assumed to be semi-linear on the left, rather than on the right.

Proposition 1.7 allows us to define the adjoint of a linear map, as in the Hermitian case (see Proposition **??**). The proof is unchanged.

Proposition 1.8. *Given a Hilbert space E, for every linear map $f \colon E \to E$, there is a unique linear map $f^* \colon E \to E$, such that*

$$\langle f^*(u), v \rangle = \langle u, f(v) \rangle$$

for all $u, v \in E$. The map f^* is called the adjoint of f.

It is easy to show that if f is continuous, then f^* is also continuous. As in the Hermitian case, given two Hilbert spaces E and F, for any linear map $f \colon E \to F$, such that

$$\langle f(u), v \rangle_2 = \langle u, f^*(v) \rangle_1$$

for all $u \in E$ and all $v \in F$. The linear map f^* is also called the adjoint of f.

1.2 Total Orthogonal Families (Hilbert Bases), Fourier Coefficients

We conclude our quick tour of Hilbert spaces by showing that the notion of orthogonal basis can be generalized to Hilbert spaces. However, the useful notion is not the usual notion of a basis, but a notion which is an abstraction of the concept of Fourier series. Every element of a Hilbert space is the "sum" of its Fourier series.

Definition 1.3. Given a Hilbert space E, a family $(u_k)_{k\in K}$ of nonnull vectors is an *orthogonal family* iff the u_k are pairwise orthogonal, i.e., $\langle u_i, u_j\rangle = 0$ for all $i \neq j$ $(i, j \in K)$, and an *orthonormal family* iff $\langle u_i, u_j\rangle = \delta_{i,j}$, for all $i, j \in K$. A *total orthogonal family (or system)* or *Hilbert basis* is an orthogonal family that is dense in E. This means that for every $v \in E$, for every $\epsilon > 0$, there is some finite subset $I \subseteq K$ and some family $(\lambda_i)_{i\in I}$ of complex numbers, such that

$$\left\| v - \sum_{i\in I} \lambda_i u_i \right\| < \epsilon.$$

Given an orthogonal family $(u_k)_{k\in K}$, for every $v \in E$, for every $k \in K$, the scalar $c_k = \langle v, u_k\rangle / \|u_k\|^2$ is called the *k-th Fourier coefficient of v over $(u_k)_{k\in K}$*.

Remark: The terminology Hilbert basis is misleading, because a Hilbert basis $(u_k)_{k\in K}$ is not necessarily a basis in the algebraic sense. Indeed, in general, $(u_k)_{k\in K}$ does not span E. Intuitively, it takes linear combinations of the u_k's with infinitely many nonnull coefficients to span E. Technically, this is achieved in terms of limits. In order to avoid the confusion between bases in the algebraic sense and Hilbert bases, some authors refer to algebraic bases as *Hamel bases* and to total orthogonal families (or Hilbert bases) as *Schauder bases*.

Given an orthogonal family $(u_k)_{k\in K}$, for any finite subset I of K, we often call sums of the form $\sum_{i\in I} \lambda_i u_i$ *partial sums of Fourier series*, and if these partial sums converge to a limit denoted as $\sum_{k\in K} c_k u_k$, we call $\sum_{k\in K} c_k u_k$ a *Fourier series*.

However, we have to make sense of such sums! Indeed, when K is unordered or uncountable, the notion of limit or sum has not been defined. This can be done as follows (for more details, see Dixmier [3]):

Definition 1.4. Given a normed vector space E (say, a Hilbert space), for any nonempty index set K, we say that a family $(u_k)_{k\in K}$ of vectors in E is *summable with sum $v \in E$* iff for every $\epsilon > 0$, there is some finite subset I of K, such that,

$$\left\| v - \sum_{j\in J} u_j \right\| < \epsilon$$

for every finite subset J with $I \subseteq J \subseteq K$. We say that the family $(u_k)_{k\in K}$ is *summable* iff there is some $v \in E$ such that $(u_k)_{k\in K}$ is summable with sum v. A family $(u_k)_{k\in K}$ is a

Cauchy family iff for every $\epsilon > 0$, there is a finite subset I of K, such that,

$$\Big\|\sum_{j\in J} u_j\Big\| < \epsilon$$

for every finite subset J of K with $I \cap J = \emptyset$,

If $(u_k)_{k\in K}$ is summable with sum v, we usually denote v as $\sum_{k\in K} u_k$. The following technical proposition will be needed:

Proposition 1.9. *Let E be a complete normed vector space (say, a Hilbert space).*

(1) For any nonempty index set K, a family $(u_k)_{k\in K}$ is summable iff it is a Cauchy family.

(2) Given a family $(r_k)_{k\in K}$ of nonnegative reals $r_k \geq 0$, if there is some real number $B > 0$ such that $\sum_{i\in I} r_i < B$ for every finite subset I of K, then $(r_k)_{k\in K}$ is summable and $\sum_{k\in K} r_k = r$, where r is least upper bound of the set of finite sums $\sum_{i\in I} r_i$ $(I \subseteq K)$.

Proof. (1) If $(u_k)_{k\in K}$ is summable, for every finite subset I of K, let

$$u_I = \sum_{i\in I} u_i \quad \text{and} \quad u = \sum_{k\in K} u_k$$

For every $\epsilon > 0$, there is some finite subset I of K such that

$$\|u - u_L\| < \epsilon/2$$

for all finite subsets L such that $I \subseteq L \subseteq K$. For every finite subset J of K such that $I \cap J = \emptyset$, since $I \subseteq I \cup J \subseteq K$ and $I \cup J$ is finite, we have

$$\|u - u_{I\cup J}\| < \epsilon/2 \quad \text{and} \quad \|u - u_I\| < \epsilon/2,$$

and since

$$\|u_{I\cup J} - u_I\| \leq \|u_{I\cup J} - u\| + \|u - u_I\|$$

and $u_{I\cup J} - u_I = u_J$ since $I \cap J = \emptyset$, we get

$$\|u_J\| = \|u_{I\cup J} - u_I\| < \epsilon,$$

which is the condition for $(u_k)_{k\in K}$ to be a Cauchy family.

Conversely, assume that $(u_k)_{k\in K}$ is a Cauchy family. We define inductively a decreasing sequence (X_n) of subsets of E, each of diameter at most $1/n$, as follows: For $n = 1$, since $(u_k)_{k\in K}$ is a Cauchy family, there is some finite subset J_1 of K such that

$$\|u_J\| < 1/2$$

for every finite subset J of K with $J_1 \cap J = \emptyset$. We pick some finite subset J_1 with the above property, and we let $I_1 = J_1$ and

$$X_1 = \{u_I \mid I_1 \subseteq I \subseteq K,\ I \text{ finite}\}.$$

For $n \geq 1$, there is some finite subset J_{n+1} of K such that

$$\|u_J\| < 1/(2n+2)$$

for every finite subset J of K with $J_{n+1} \cap J = \emptyset$. We pick some finite subset J_{n+1} with the above property, and we let $I_{n+1} = I_n \cup J_{n+1}$ and

$$X_{n+1} = \{u_I \mid I_{n+1} \subseteq I \subseteq K,\ I \text{ finite}\}.$$

Since $I_n \subseteq I_{n+1}$, it is obvious that $X_{n+1} \subseteq X_n$ for all $n \geq 1$. We need to prove that each X_n has diameter at most $1/n$. Since J_n was chosen such that

$$\|u_J\| < 1/(2n)$$

for every finite subset J of K with $J_n \cap J = \emptyset$, and since $J_n \subseteq I_n$, it is also true that

$$\|u_J\| < 1/(2n)$$

for every finite subset J of K with $I_n \cap J = \emptyset$ (since $I_n \cap J = \emptyset$ and $J_n \subseteq I_n$ implies that $J_n \cap J = \emptyset$). Then, for every two finite subsets J, L such that $I_n \subseteq J, L \subseteq K$, we have

$$\|u_{J-I_n}\| < 1/(2n) \quad \text{and} \quad \|u_{L-I_n}\| < 1/(2n),$$

and since

$$\|u_J - u_L\| \leq \|u_J - u_{I_n}\| + \|u_{I_n} - u_L\| = \|u_{J-I_n}\| + \|u_{L-I_n}\|,$$

we get

$$\|u_J - u_L\| < 1/n,$$

which proves that $\delta(X_n) \leq 1/n$. Now, if we consider the sequence of closed sets $(\overline{X_n})$, we still have $\overline{X_{n+1}} \subseteq \overline{X_n}$, and by Proposition 1.3, $\delta(\overline{X_n}) = \delta(X_n) \leq 1/n$, which means that $\lim_{n\to\infty} \delta(\overline{X_n}) = 0$, and by Proposition 1.3, $\bigcap_{n=1}^{\infty} \overline{X_n}$ consists of a single element u. We claim that u is the sum of the family $(u_k)_{k\in K}$.

For every $\epsilon > 0$, there is some $n \geq 1$ such that $n > 2/\epsilon$, and since $u \in \overline{X_m}$ for all $m \geq 1$, there is some finite subset J_0 of K such that $I_n \subseteq J_0$ and

$$\|u - u_{J_0}\| < \epsilon/2,$$

where I_n is the finite subset of K involved in the definition of X_n. However, since $\delta(X_n) \leq 1/n$, for every finite subset J of K such that $I_n \subseteq J$, we have

$$\|u_J - u_{J_0}\| \leq 1/n < \epsilon/2,$$

and since

$$\|u - u_J\| \leq \|u - u_{J_0}\| + \|u_{J_0} - u_J\|,$$

we get

$$\|u - u_J\| < \epsilon$$

for every finite subset J of K with $I_n \subseteq J$, which proves that u is the sum of the family $(u_k)_{k\in K}$.

(2) Since every finite sum $\sum_{i\in I} r_i$ is bounded by the uniform bound B, the set of these finite sums has a least upper bound $r \leq B$. For every $\epsilon > 0$, since r is the least upper bound of the finite sums $\sum_{i\in I} r_i$ (where I finite, $I \subseteq K$), there is some finite $I \subseteq K$ such that

$$\left| r - \sum_{i\in I} r_i \right| < \epsilon,$$

and since $r_k \geq 0$ for all $k \in K$, we have

$$\sum_{i\in I} r_i \leq \sum_{j\in J} r_j$$

whenever $I \subseteq J$, which shows that

$$\left| r - \sum_{j\in J} r_j \right| \leq \left| r - \sum_{i\in I} r_i \right| < \epsilon$$

for every finite subset J such that $I \subseteq J \subseteq K$, proving that $(r_k)_{k\in K}$ is summable with sum $\sum_{k\in K} r_k = r$. □

Remark: The notion of summability implies that the sum of a family $(u_k)_{k\in K}$ is independent of any order on K. In this sense, it is a kind of "commutative summability". More precisely, it is easy to show that for every bijection $\varphi\colon K \to K$ (intuitively, a reordering of K), the family $(u_k)_{k\in K}$ is summable iff the family $(u_l)_{l\in\varphi(K)}$ is summable, and if so, they have the same sum.

The following proposition gives some of the main properties of Fourier coefficients. Among other things, at most countably many of the Fourier coefficient may be nonnull, and the partial sums of a Fourier series converge. Given an orthogonal family $(u_k)_{k\in K}$, we let $U_k = \mathbb{C}u_k$, and $p_{U_k}\colon E \to U_k$ is the projection of E onto U_k.

Proposition 1.10. *Let E be a Hilbert space, $(u_k)_{k\in K}$ an orthogonal family in E, and V the closure of the subspace generated by $(u_k)_{k\in K}$. The following properties hold:*

(1) For every $v \in E$, for every finite subset $I \subseteq K$, we have

$$\sum_{i\in I} |c_i|^2 \leq \|v\|^2,$$

where the c_k are the Fourier coefficients of v.

(2) For every vector $v \in E$, if $(c_k)_{k\in K}$ are the Fourier coefficients of v, the following conditions are equivalent:

(2a) $v \in V$

(2b) The family $(c_k u_k)_{k\in K}$ is summable and $v = \sum_{k\in K} c_k u_k$.

(2c) The family $(|c_k|^2)_{k\in K}$ is summable and $\|v\|^2 = \sum_{k\in K} |c_k|^2$;

(3) The family $(|c_k|^2)_{k\in K}$ is summable, and we have the Bessel inequality:

$$\sum_{k\in K} |c_k|^2 \leq \|v\|^2.$$

As a consequence, at most countably many of the c_k may be nonzero. The family $(c_k u_k)_{k\in K}$ forms a Cauchy family, and thus, the Fourier series $\sum_{k\in K} c_k u_k$ converges in E to some vector $u = \sum_{k\in K} c_k u_k$. Furthermore, $u = p_V(v)$.

Proof. (1) Let

$$u_I = \sum_{i\in I} c_i u_i$$

for any finite subset I of K. We claim that $v - u_I$ is orthogonal to u_i for every $i \in I$. Indeed,

$$\begin{aligned}\langle v - u_I, u_i\rangle &= \left\langle v - \sum_{j\in I} c_j u_j, u_i \right\rangle \\ &= \langle v, u_i\rangle - \sum_{j\in I} c_j \langle u_j, u_i\rangle \\ &= \langle v, u_i\rangle - c_i \|u_i\|^2 \\ &= \langle v, u_i\rangle - \langle v, u_i\rangle = 0,\end{aligned}$$

since $\langle u_j, u_i\rangle = 0$ for all $i \neq j$ and $c_i = \langle v, u_i\rangle / \|u_i\|^2$. As a consequence, we have

$$\begin{aligned}\|v\|^2 &= \left\| v - \sum_{i\in I} c_i u_i + \sum_{i\in I} c_i u_i \right\|^2 \\ &= \left\| v - \sum_{i\in I} c_i u_i \right\|^2 + \left\| \sum_{i\in I} c_i u_i \right\|^2 \\ &= \left\| v - \sum_{i\in I} c_i u_i \right\|^2 + \sum_{i\in I} |c_i|^2,\end{aligned}$$

since the u_i are pairwise orthogonal, that is,

$$\|v\|^2 = \left\| v - \sum_{i\in I} c_i u_i \right\|^2 + \sum_{i\in I} |c_i|^2.$$

Thus,

$$\sum_{i\in I} |c_i|^2 \leq \|v\|^2,$$

as claimed.

(2) We prove the chain of implications $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

$(a) \Rightarrow (b)$: If $v \in V$, since V is the closure of the subspace spanned by $(u_k)_{k\in K}$, for every $\epsilon > 0$, there is some finite subset I of K and some family $(\lambda_i)_{i\in I}$ of complex numbers, such that

$$\left\| v - \sum_{i\in I} \lambda_i u_i \right\| < \epsilon.$$

Now, for every finite subset J of K such that $I \subseteq J$, we have

$$\begin{aligned}\left\| v - \sum_{i\in I} \lambda_i u_i \right\|^2 &= \left\| v - \sum_{j\in J} c_j u_j + \sum_{j\in J} c_j u_j - \sum_{i\in I} \lambda_i u_i \right\|^2 \\ &= \left\| v - \sum_{j\in J} c_j u_j \right\|^2 + \left\| \sum_{j\in J} c_j u_j - \sum_{i\in I} \lambda_i u_i \right\|^2,\end{aligned}$$

since $I \subseteq J$ and the u_j (with $j \in J$) are orthogonal to $v - \sum_{j\in J} c_j u_j$ by the argument in (1), which shows that

$$\left\| v - \sum_{j\in J} c_j u_j \right\| \leq \left\| v - \sum_{i\in I} \lambda_i u_i \right\| < \epsilon,$$

and thus, that the family $(c_k u_k)_{k\in K}$ is summable with sum v, so that

$$v = \sum_{k\in K} c_k u_k.$$

$(b) \Rightarrow (c)$: If $v = \sum_{k\in K} c_k u_k$, then for every $\epsilon > 0$, there some finite subset I of K, such that

$$\left\| v - \sum_{j\in J} c_j u_j \right\| < \sqrt{\epsilon},$$

for every finite subset J of K such that $I \subseteq J$, and since we proved in (1) that

$$\|v\|^2 = \left\| v - \sum_{j\in J} c_j u_j \right\|^2 + \sum_{j\in J} |c_j|^2,$$

we get

$$\|v\|^2 - \sum_{j\in J} |c_j|^2 < \epsilon,$$

which proves that $(|c_k|^2)_{k\in K}$ is summable with sum $\|v\|^2$.

$(c) \Rightarrow (a)$: Finally, if $(|c_k|^2)_{k\in K}$ is summable with sum $\|v\|^2$, for every $\epsilon > 0$, there is some finite subset I of K such that

$$\|v\|^2 - \sum_{j\in J} |c_j|^2 < \epsilon^2$$

for every finite subset J of K such that $I \subseteq J$, and again, using the fact that

$$\|v\|^2 = \Big\|v - \sum_{j\in J} c_j u_j\Big\|^2 + \sum_{j\in J} |c_j|^2,$$

we get

$$\Big\|v - \sum_{j\in J} c_j u_j\Big\| < \epsilon,$$

which proves that $(c_k u_k)_{k\in K}$ is summable with sum $\sum_{k\in K} c_k u_k = v$, and $v \in V$.

(3) Since $\sum_{i\in I} |c_i|^2 \leq \|v\|^2$ for every finite subset I of K, by Proposition 1.9, the family $(|c_k|^2)_{k\in K}$ is summable. The Bessel inequality

$$\sum_{k\in K} |c_k|^2 \leq \|v\|^2$$

is an obvious consequence of the inequality $\sum_{i\in I} |c_i|^2 \leq \|v\|^2$ (for every finite $I \subseteq K$). Now, for every natural number $n \geq 1$, if K_n is the subset of K consisting of all c_k such that $|c_k| \geq 1/n$, the number of elements in K_n is at most

$$\sum_{k\in K_n} |nc_k|^2 \leq n^2 \sum_{k\in K} |c_k|^2 \leq n^2 \|v\|^2,$$

which is finite, and thus, at most a countable number of the c_k may be nonzero.

Since $(|c_k|^2)_{k\in K}$ is summable with sum c, for every $\epsilon > 0$, there is some finite subset I of K such that

$$\sum_{j\in J} |c_j|^2 < \epsilon^2$$

for every finite subset J of K such that $I \cap J = \emptyset$. Since

$$\Big\|\sum_{j\in J} c_j u_j\Big\|^2 = \sum_{j\in J} |c_j|^2,$$

we get

$$\Big\|\sum_{j\in J} c_j u_j\Big\| < \epsilon.$$

This proves that $(c_k u_k)_{k\in K}$ is a Cauchy family, which, by Proposition 1.9, implies that $(c_k u_k)_{k\in K}$ is summable, since E is complete. Thus, the Fourier series $\sum_{k\in K} c_k u_k$ is summable, with its sum denoted $u \in V$.

Since $\sum_{k\in K} c_k u_k$ is summable with sum u, for every $\epsilon > 0$, there is some finite subset I_1 of K such that

$$\Big\| u - \sum_{j\in J} c_j u_j \Big\| < \epsilon$$

for every finite subset J of K such that $I_1 \subseteq J$. By the triangle inequality, for every finite subset I of K,

$$\Big\| u - v \Big\| \leq \Big\| u - \sum_{i\in I} c_i u_i \Big\| + \Big\| \sum_{i\in I} c_i u_i - v \Big\|.$$

By (2), every $w \in V$ is the sum of its Fourier series $\sum_{k\in K} \lambda_k u_k$, and for every $\epsilon > 0$, there is some finite subset I_2 of K such that

$$\Big\| w - \sum_{j\in J} \lambda_j u_j \Big\| < \epsilon$$

for every finite subset J of K such that $I_2 \subseteq J$. By the triangle inequality, for every finite subset I of K,

$$\Big\| v - \sum_{i\in I} \lambda_i u_i \Big\| \leq \|v - w\| + \Big\| w - \sum_{i\in I} \lambda_i u_i \Big\|.$$

Letting $I = I_1 \cup I_2$, since we showed in (2) that

$$\Big\| v - \sum_{i\in I} c_i u_i \Big\| \leq \Big\| v - \sum_{i\in I} \lambda_i u_i \Big\|$$

for every finite subset I of K, we get

$$\begin{aligned}
\|u - v\| &\leq \Big\| u - \sum_{i\in I} c_i u_i \Big\| + \Big\| \sum_{i\in I} c_i u_i - v \Big\| \\
&\leq \Big\| u - \sum_{i\in I} c_i u_i \Big\| + \Big\| \sum_{i\in I} \lambda_i u_i - v \Big\| \\
&\leq \Big\| u - \sum_{i\in I} c_i u_i \Big\| + \|v - w\| + \Big\| w - \sum_{i\in I} \lambda_i u_i \Big\|,
\end{aligned}$$

and thus

$$\|u - v\| \leq \|v - w\| + 2\epsilon.$$

Since this holds for every $\epsilon > 0$, we have

$$\|u - v\| \leq \|v - w\|$$

for all $w \in V$, i.e. $\|v - u\| = d(v, V)$, with $u \in V$, which proves that $u = p_V(v)$. ☐

1.3 The Hilbert Space $l^2(K)$ and the Riesz-Fischer Theorem

Proposition 1.10 suggests looking at the space of sequences $(z_k)_{k\in K}$ (where $z_k \in \mathbb{C}$) such that $(|z_k|^2)_{k\in K}$ is summable. Indeed, such spaces are Hilbert spaces, and it turns out that every Hilbert space is isomorphic to one of those. Such spaces are the infinite-dimensional version of the spaces $\mathbb{C}^n$ under the usual Euclidean norm.

Definition 1.5. Given any nonempty index set K, the space $l^2(K)$ is the set of all sequences $(z_k)_{k\in K}$, where $z_k \in \mathbb{C}$, such that $(|z_k|^2)_{k\in K}$ is summable, i.e., $\sum_{k\in K} |z_k|^2 < \infty$.

Remarks:

(1) When K is a finite set of cardinality n, $l^2(K)$ is isomorphic to $\mathbb{C}^n$.

(2) When $K = \mathbb{N}$, the space $l^2(\mathbb{N})$ corresponds to the space l^2 of Example 2 in Section **??**. In that example, we claimed that l^2 was a Hermitian space, and in fact, a Hilbert space. We now prove this fact for any index set K.

Proposition 1.11. *Given any nonempty index set K, the space $l^2(K)$ is a Hilbert space under the Hermitian product*

$$\langle (x_k)_{k\in K}, (y_k)_{k\in K}\rangle = \sum_{k\in K} x_k\overline{y_k}.$$

The subspace consisting of sequences $(z_k)_{k\in K}$ such that $z_k = 0$, except perhaps for finitely many k, is a dense subspace of $l^2(K)$.

Proof. First, we need to prove that $l^2(K)$ is a vector space. Assume that $(x_k)_{k\in K}$ and $(y_k)_{k\in K}$ are in $l^2(K)$. This means that $(|x_k|^2)_{k\in K}$ and $(|y_k|^2)_{k\in K}$ are summable, which, in view of Proposition 1.9, is equivalent to the existence of some positive bounds A and B such that $\sum_{i\in I} |x_i|^2 < A$ and $\sum_{i\in I} |y_i|^2 < B$, for every finite subset I of K. To prove that $(|x_k + y_k|^2)_{k\in K}$ is summable, it is sufficient to prove that there is some $C > 0$ such that $\sum_{i\in I} |x_i + y_i|^2 < C$ for every finite subset I of K. However, the parallelogram inequality implies that

$$\sum_{i\in I} |x_i + y_i|^2 \leq \sum_{i\in I} 2(|x_i|^2 + |y_i|^2) \leq 2(A + B),$$

for every finite subset I of K, and we conclude by Proposition 1.9. Similarly, for every $\lambda \in \mathbb{C}$,

$$\sum_{i\in I} |\lambda x_i|^2 \leq \sum_{i\in I} |\lambda|^2|x_i|^2 \leq |\lambda|^2 A,$$

and $(\lambda_k x_k)_{k\in K}$ is summable. Therefore, $l^2(K)$ is a vector space.

By the Cauchy-Schwarz inequality,

$$\sum_{i\in I}|x_i\overline{y_i}| \leq \sum_{i\in I}|x_i||y_i| \leq \big(\sum_{i\in I}|x_i|^2\big)^{1/2}\big(\sum_{i\in I}|y_i|^2\big)^{1/2} \leq \sum_{i\in I}(|x_i|^2+|y_i|^2)/2 \leq (A+B)/2,$$

for every finite subset I of K. Here, we used the fact that

$$4CD \leq (C+D)^2,$$

which is equivalent to

$$(C-D)^2 \geq 0.$$

By Proposition 1.9, $(|x_k\overline{y_k}|)_{k\in K}$ is summable. The customary language is that $(x_k\overline{y_k})_{k\in K}$ is absolutely summable. However, it is a standard fact that this implies that $(x_k\overline{y_k})_{k\in K}$ is summable (For every $\epsilon > 0$, there is some finite subset I of K such that

$$\sum_{j\in J}|x_j\overline{y_j}| < \epsilon$$

for every finite subset J of K such that $I \cap J = \emptyset$, and thus

$$|\sum_{j\in J}x_j\overline{y_j}| \leq \sum_{i\in J}|x_j\overline{y_j}| < \epsilon,$$

proving that $(x_k\overline{y_k})_{k\in K}$ is a Cauchy family, and thus summable). We still have to prove that $l^2(K)$ is complete.

Consider a sequence $((\lambda_k^n)_{k\in K})_{n\geq 1}$ of sequences $(\lambda_k^n)_{k\in K} \in l^2(K)$, and assume that it is a Cauchy sequence. This means that for every $\epsilon > 0$, there is some $N \geq 1$ such that

$$\sum_{k\in K}|\lambda_k^m - \lambda_k^n|^2 < \epsilon^2$$

for all $m, n \geq N$. For every fixed $k \in K$, this implies that

$$|\lambda_k^m - \lambda_k^n| < \epsilon$$

for all $m, n \geq N$, which shows that $(\lambda_k^n)_{n\geq 1}$ is a Cauchy sequence in $\mathbb{C}$. Since $\mathbb{C}$ is complete, the sequence $(\lambda_k^n)_{n\geq 1}$ has a limit $\lambda_k \in \mathbb{C}$. We claim that $(\lambda_k)_{k\in K} \in l^2(K)$ and that this is the limit of $((\lambda_k^n)_{k\in K})_{n\geq 1}$.

Given any $\epsilon > 0$, the fact that $((\lambda_k^n)_{k\in K})_{n\geq 1}$ is a Cauchy sequence implies that there is some $N \geq 1$ such that for every finite subset I of K, we have

$$\sum_{i\in I}|\lambda_i^m - \lambda_i^n|^2 < \epsilon/4$$

for all $m, n \geq N$. Let $p = |I|$. Then,

$$|\lambda_i^m - \lambda_i^n| < \frac{\sqrt{\epsilon}}{2\sqrt{p}}$$

for every $i \in I$. Since λ_i is the limit of $(\lambda_i^n)_{n\geq 1}$, we can find some n large enough so that

$$|\lambda_i^n - \lambda_i| < \frac{\sqrt{\epsilon}}{2\sqrt{p}}$$

for every $i \in I$. Since

$$|\lambda_i^m - \lambda_i| \leq |\lambda_i^m - \lambda_i^n| + |\lambda_i^n - \lambda_i|,$$

we get

$$|\lambda_i^m - \lambda_i| < \frac{\sqrt{\epsilon}}{\sqrt{p}},$$

and thus,

$$\sum_{i\in I} |\lambda_i^m - \lambda_i|^2 < \epsilon,$$

for all $m \geq N$. Since the above holds for every finite subset I of K, by Proposition 1.9, we get

$$\sum_{k\in K} |\lambda_k^m - \lambda_k|^2 < \epsilon,$$

for all $m \geq N$. This proves that $(\lambda_k^m - \lambda_k)_{k\in K} \in l^2(K)$ for all $m \geq N$, and since $l^2(K)$ is a vector space and $(\lambda_k^m)_{k\in K} \in l^2(K)$ for all $m \geq 1$, we get $(\lambda_k)_{k\in K} \in l^2(K)$. However,

$$\sum_{k\in K} |\lambda_k^m - \lambda_k|^2 < \epsilon$$

for all $m \geq N$, means that the sequence $(\lambda_k^m)_{k\in K}$ converges to $(\lambda_k)_{k\in K} \in l^2(K)$. The fact that the subspace consisting of sequences $(z_k)_{k\in K}$ such that $z_k = 0$ except perhaps for finitely many k is a dense suspace of $l^2(K)$ is left as an easy exercise. □

Remark: The subspace consisting of all sequences $(z_k)_{k\in K}$ such that $z_k = 0$, except perhaps for finitely many k, provides an example of a subspace which is not closed in $l^2(K)$. Indeed, this space is strictly contained in $l^2(K)$, since there are countable sequences of nonnull elements in $l^2(K)$ (why?).

We just need two more propositions before being able to prove that every Hilbert space is isomorphic to some $l^2(K)$.

Proposition 1.12. *Let E be a Hilbert space, and $(u_k)_{k\in K}$ an orthogonal family in E. The following properties hold:*

(1) *For every family $(\lambda_k)_{k\in K} \in l^2(K)$, the family $(\lambda_k u_k)_{k\in K}$ is summable. Furthermore, $v = \sum_{k\in K} \lambda_k u_k$ is the only vector such that $c_k = \lambda_k$ for all $k \in K$, where the c_k are the Fourier coefficients of v.*

(2) *For any two families $(\lambda_k)_{k\in K} \in l^2(K)$ and $(\mu_k)_{k\in K} \in l^2(K)$, if $v = \sum_{k\in K} \lambda_k u_k$ and $w = \sum_{k\in K} \mu_k u_k$, we have the following equation, also called Parseval identity:*

$$\langle v, w\rangle = \sum_{k\in K} \lambda_k \overline{\mu_k}.$$

Proof. (1) The fact that $(\lambda_k)_{k\in K} \in l^2(K)$ means that $(|\lambda_k|^2)_{k\in K}$ is summable. The proof given in Proposition 1.10 (3) applies to the family $(|\lambda_k|^2)_{k\in K}$ (instead of $(|c_k|^2)_{k\in K}$), and yields the fact that $(\lambda_k u_k)_{k\in K}$ is summable. Letting $v = \sum_{k\in K} \lambda_k u_k$, recall that $c_k = \langle v, u_k\rangle / \|u_k\|^2$. Pick some $k \in K$. Since $\langle -, -\rangle$ is continuous, for every $\epsilon > 0$, there is some $\eta > 0$ such that

$$|\langle v, u_k\rangle - \langle w, u_k\rangle| < \epsilon \|u_k\|^2$$

whenever

$$\|v - w\| < \eta.$$

However, since for every $\eta > 0$, there is some finite subset I of K such that

$$\left\| v - \sum_{j\in J} \lambda_j u_j \right\| < \eta$$

for every finite subset J of K such that $I \subseteq J$, we can pick $J = I \cup \{k\}$, and letting $w = \sum_{j\in J} \lambda_j u_j$, we get

$$\left| \langle v, u_k\rangle - \left\langle \sum_{j\in J} \lambda_j u_j, u_k \right\rangle \right| < \epsilon \|u_k\|^2.$$

However,

$$\langle v, u_k\rangle = c_k \|u_k\|^2 \quad \text{and} \quad \left\langle \sum_{j\in J} \lambda_j u_j, u_k \right\rangle = \lambda_k \|u_k\|^2,$$

and thus, the above proves that $|c_k - \lambda_k| < \epsilon$ for every $\epsilon > 0$, and thus, that $c_k = \lambda_k$.

(2) Since $\langle -, -\rangle$ is continuous, for every $\epsilon > 0$, there are some $\eta_1 > 0$ and $\eta_2 > 0$, such that

$$|\langle x, y\rangle| < \epsilon$$

whenever $\|x\| < \eta_1$ and $\|y\| < \eta_2$. Since $v = \sum_{k\in K} \lambda_k u_k$ and $w = \sum_{k\in K} \mu_k u_k$, there is some finite subset I_1 of K such that

$$\left\| v - \sum_{j\in J} \lambda_j u_j \right\| < \eta_1$$

for every finite subset J of K such that $I_1 \subseteq J$, and there is some finite subset I_2 of K such that

$$\left\| w - \sum_{j \in J} \mu_j u_j \right\| < \eta_2$$

for every finite subset J of K such that $I_2 \subseteq J$. Letting $I = I_1 \cup I_2$, we get

$$\left| \left\langle v - \sum_{i \in I} \lambda_i u_i, w - \sum_{i \in I} \mu_i u_i \right\rangle \right| < \epsilon.$$

Furthermore,

$$\begin{aligned} \langle v, w \rangle &= \left\langle v - \sum_{i \in I} \lambda_i u_i + \sum_{i \in I} \lambda_i u_i, w - \sum_{i \in I} \mu_i u_i + \sum_{i \in I} \mu_i u_i \right\rangle \\ &= \left\langle v - \sum_{i \in I} \lambda_i u_i, w - \sum_{i \in I} \mu_i u_i \right\rangle + \sum_{i \in I} \lambda_i \overline{\mu_i}, \end{aligned}$$

since the u_i are orthogonal to $v - \sum_{i \in I} \lambda_i u_i$ and $w - \sum_{i \in I} \mu_i u_i$ for all $i \in I$. This proves that for every $\epsilon > 0$, there is some finite subset I of K such that

$$\left| \langle v, w \rangle - \sum_{i \in I} \lambda_i \overline{\mu_i} \right| < \epsilon.$$

We already know from Proposition 1.11 that $(\lambda_k \overline{\mu_k})_{k \in K}$ is summable, and since $\epsilon > 0$ is arbitrary, we get

$$\langle v, w \rangle = \sum_{k \in K} \lambda_k \overline{\mu_k}.$$

□

The next proposition states properties characterizing Hilbert bases (total orthogonal families).

Proposition 1.13. *Let E be a Hilbert space, and let $(u_k)_{k \in K}$ be an orthogonal family in E. The following properties are equivalent:*

(1) The family $(u_k)_{k \in K}$ is a total orthogonal family.

(2) For every vector $v \in E$, if $(c_k)_{k \in K}$ are the Fourier coefficients of v, then the family $(c_k u_k)_{k \in K}$ is summable and $v = \sum_{k \in K} c_k u_k$.

(3) For every vector $v \in E$, we have the Parseval identity:

$$\|v\|^2 = \sum_{k \in K} |c_k|^2.$$

(4) For every vector $u \in E$, if $\langle u, u_k \rangle = 0$ for all $k \in K$, then $u = 0$.

Proof. The equivalence of (1), (2), and (3), is an immediate consequence of Proposition 1.10 and Proposition 1.12.

(4) If $(u_k)_{k\in K}$ is a total orthogonal family and $\langle u, u_k \rangle = 0$ for all $k \in K$, since $u = \sum_{k\in K} c_k u_k$ where $c_k = \langle u, u_k \rangle / \|u_k\|^2$, we have $c_k = 0$ for all $k \in K$, and $u = 0$.

Conversely, assume that the closure V of $(u_k)_{k\in K}$ is different from E. Then, by Proposition 1.6, we have $E = V \oplus V^\perp$, where $V^\perp$ is the orthogonal complement of V, and $V^\perp$ is nontrivial since $V \neq E$. As a consequence, there is some nonnull vector $u \in V^\perp$. But then, u is orthogonal to every vector in V, and in particular,

$$\langle u, u_k \rangle = 0$$

for all $k \in K$, which, by assumption, implies that $u = 0$, contradicting the fact that $u \neq 0$. $\square$

Remarks:

(1) If E is a Hilbert space and $(u_k)_{k\in K}$ is a total orthogonal family in E, there is a simpler argument to prove that $u = 0$ if $\langle u, u_k \rangle = 0$ for all $k \in K$, based on the continuity of $\langle -, - \rangle$. The argument is to prove that the assumption implies that $\langle v, u \rangle = 0$ for all $v \in E$. Since $\langle -, - \rangle$ is positive definite, this implies that $u = 0$. By continuity of $\langle -, - \rangle$, for every $\epsilon > 0$, there is some $\eta > 0$ such that for every finite subset I of K, for every family $(\lambda_i)_{i\in I}$, for every $v \in E$,

$$\left| \langle v, u \rangle - \left\langle \sum_{i\in I} \lambda_i u_i, u \right\rangle \right| < \epsilon$$

whenever

$$\left\| v - \sum_{i\in I} \lambda_i u_i \right\| < \eta.$$

Since $(u_k)_{k\in K}$ is dense in E, for every $v \in E$, there is some finite subset I of K and some family $(\lambda_i)_{i\in I}$ such that

$$\left\| v - \sum_{i\in I} \lambda_i u_i \right\| < \eta,$$

and since by assumption, $\left\langle \sum_{i\in I} \lambda_i u_i, u \right\rangle = 0$, we get

$$|\langle v, u \rangle| < \epsilon.$$

Since this holds for every $\epsilon > 0$, we must have $\langle v, u \rangle = 0$

(2) If V is any nonempty subset of E, the kind of argument used in the previous remark can be used to prove that $V^{\perp}$ is closed (even if V is not), and that $V^{\perp\perp}$ is the closure of V.

We will now prove that every Hilbert space has some Hilbert basis. This requires using a fundamental theorem from set theory known as *Zorn's Lemma*, which we quickly review.

Given any set X with a partial ordering $\leq$, recall that a nonempty subset C of X is a *chain* if it is totally ordered (i.e., for all $x, y \in C$, either $x \leq y$ or $y \leq x$). A nonempty subset Y of X is *bounded* iff there is some $b \in X$ such that $y \leq b$ for all $y \in Y$. Some $m \in X$ is *maximal* iff for every $x \in X$, $m \leq x$ implies that $x = m$. We can now state Zorn's Lemma. For more details, see Rudin [7], Lang [4], or Artin [1].

Proposition 1.14. *Given any nonempty partially ordered set X, if every (nonempty) chain in X is bounded, then X has some maximal element.*

We can now prove the existence of Hilbert bases. We define a partial order on families $(u_k)_{k\in K}$ as follows: For any two families $(u_k)_{k\in K_1}$ and $(v_k)_{k\in K_2}$, we say that

$$(u_k)_{k\in K_1} \leq (v_k)_{k\in K_2}$$

iff $K_1 \subseteq K_2$ and $u_k = v_k$ for all $k \in K_1$. This is clearly a partial order.

Proposition 1.15. *Let E be a Hilbert space. Given any orthogonal family $(u_k)_{k\in K}$ in E, there is a total orthogonal family $(u_l)_{l\in L}$ containing $(u_k)_{k\in K}$.*

Proof. Consider the set $\mathcal{S}$ of all orthogonal families greater than or equal to the family $B = (u_k)_{k\in K}$. We claim that every chain in $\mathcal{S}$ is bounded. Indeed, if $C = (C_l)_{l\in L}$ is a chain in $\mathcal{S}$, where $C_l = (u_{k,l})_{k\in K_l}$, the union family

$$(u_k)_{k\in\bigcup_{l\in L} K_l}, \text{ where } u_k = u_{k,l} \text{ whenever } k \in K_l,$$

is clearly an upper bound for C, and it is immediately verified that it is an orthogonal family. By Zorn's Lemma 1.14, there is a maximal family $(u_l)_{l\in L}$ containing $(u_k)_{k\in K}$. If $(u_l)_{l\in L}$ is not dense in E, then its closure V is strictly contained in E, and by Proposition 1.6, the orthogonal complement $V^{\perp}$ of V is nontrivial since $V \neq E$. As a consequence, there is some nonnull vector $u \in V^{\perp}$. But then, u is orthogonal to every vector in $(u_l)_{l\in L}$, and we can form an orthogonal family strictly greater than $(u_l)_{l\in L}$ by adding u to this family, contradicting the maximality of $(u_l)_{l\in L}$. Therefore, $(u_l)_{l\in L}$ is dense in E, and thus, it is a Hilbert basis. $\square$

Remark: It is possible to prove that all Hilbert bases for a Hilbert space E have index sets K of the same cardinality. For a proof, see Bourbaki [2].

At last, we can prove that every Hilbert space is isomorphic to some Hilbert space $l^2(K)$ for some suitable K.

Theorem 1.16. *(Riesz-Fischer) For every Hilbert space E, there is some nonempty set K such that E is isomorphic to the Hilbert space $l^2(K)$. More specifically, for any Hilbert basis $(u_k)_{k\in K}$ of E, the maps $f\colon l^2(K) \to E$ and $g\colon E \to l^2(K)$ defined such that*

$$f\left((\lambda_k)_{k\in K}\right) = \sum_{k\in K} \lambda_k u_k \quad \text{and} \quad g(u) = \left(\langle u, u_k\rangle / \|u_k\|^2\right)_{k\in K} = (c_k)_{k\in K},$$

are bijective linear isometries such that $g \circ f = \mathrm{id}$ and $f \circ g = \mathrm{id}$.

Proof. By Proposition 1.12 (1), the map f is well defined, and it it clearly linear. By Proposition 1.10 (3), the map g is well defined, and it is also clearly linear. By Proposition 1.10 (2b), we have

$$f(g(u)) = u = \sum_{k\in K} c_k u_k,$$

and by Proposition 1.12 (1), we have

$$g(f\left((\lambda_k)_{k\in K}\right)) = (\lambda_k)_{k\in K},$$

and thus $g \circ f = \mathrm{id}$ and $f \circ g = \mathrm{id}$. By Proposition 1.12 (2), the linear map g is an isometry. Therefore, f is a linear bijection and an isometry between $l^2(K)$ and E, with inverse g. □

Remark: The surjectivity of the map $g\colon E \to l^2(K)$ is known as the *Riesz-Fischer* theorem.

Having done all this hard work, we sketch how these results apply to Fourier series. Again, we refer the readers to Rudin [7] or Lang [5, 6] for a comprehensive exposition.

Let $\mathcal{C}(T)$ denote the set of all periodic continuous functions $f\colon [-\pi, \pi] \to \mathbb{C}$ with period 2π. There is a Hilbert space $\mathrm{L}^2(T)$ containing $\mathcal{C}(T)$ and such that $\mathcal{C}(T)$ is dense in $\mathrm{L}^2(T)$, whose inner product is given by

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

The Hilbert space $\mathrm{L}^2(T)$ is the space of *Lebesgue square-integrable periodic functions* (of period 2π).

It turns out that the family $(e^{ikx})_{k\in\mathbb{Z}}$ is a total orthogonal family in $\mathrm{L}^2(T)$, because it is already dense in $\mathcal{C}(T)$ (for instance, see Rudin [7]). Then, the Riesz-Fischer theorem says that for every family $(c_k)_{k\in\mathbb{Z}}$ of complex numbers such that

$$\sum_{k\in\mathbb{Z}} |c_k|^2 < \infty,$$

there is a unique function $f \in \mathrm{L}^2(T)$ such that f is equal to its Fourier series

$$f(x) = \sum_{k\in\mathbb{Z}} c_k e^{ikx},$$

where the Fourier coefficients c_k of f are given by the formula

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-ikt} dt.$$

The Parseval theorem says that

$$\sum_{k=-\infty}^{+\infty} c_k \overline{d_k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \overline{g(t)} dt$$

for all $f, g \in \mathrm{L}^2(T)$, where c_k and d_k are the Fourier coefficients of f and g.

Thus, there is an isomorphism between the two Hilbert spaces $\mathrm{L}^2(T)$ and $l^2(\mathbb{Z})$, which is the deep reason why the Fourier coefficients "work". Theorem 1.16 implies that the Fourier series $\sum_{k \in \mathbb{Z}} c_k e^{ikx}$ of a function $f \in \mathrm{L}^2(T)$ converges to f in the L^2-sense, i.e., in the mean-square sense. This does not necessarily imply that the Fourier series converges to f pointwise! This is a subtle issue, and for more on this subject, the reader is referred to Lang [5, 6] or Schwartz [10, 11].

We can also consider the set $\mathcal{C}([-1, 1])$ of continuous functions $f \colon [-1, 1] \to \mathbb{C}$. There is a Hilbert space $\mathrm{L}^2([-1, 1])$ containing $\mathcal{C}([-1, 1])$ and such that $\mathcal{C}([-1, 1])$ is dense in $\mathrm{L}^2([-1, 1])$, whose inner product is given by

$$\langle f, g \rangle = \int_{-1}^{1} f(x) \overline{g(x)} dx.$$

The Hilbert space $\mathrm{L}^2([-1, 1])$ is the space of *Lebesgue square-integrable functions* over $[-1, 1]$. The Legendre polynomials $P_n(x)$ defined in Example 5 of Section **??** (Chapter **??**) form a Hilbert basis of $\mathrm{L}^2([-1, 1])$. Recall that if we let f_n be the function

$$f_n(x) = (x^2 - 1)^n,$$

$P_n(x)$ is defined as follows:

$$P_0(x) = 1, \quad \text{and} \quad P_n(x) = \frac{1}{2^n n!} f_n^{(n)}(x),$$

where $f_n^{(n)}$ is the nth derivative of f_n. The reason for the leading coefficient is to get $P_n(1) = 1$. It can be shown with much efforts that

$$P_n(x) = \sum_{0 \le k \le n/2} (-1)^k \frac{(2(n-k))!}{2^n (n-k)! k! (n-2k)!} x^{n-2k}.$$

Bibliography

[1] Michael Artin. *Algebra*. Prentice Hall, first edition, 1991.

[2] Nicolas Bourbaki. *Espaces Vectoriels Topologiques*. Eléments de Mathématiques. Masson, 1981.

[3] Jacques Dixmier. *General Topology*. UTM. Springer Verlag, first edition, 1984.

[4] Serge Lang. *Algebra*. Addison Wesley, third edition, 1993.

[5] Serge Lang. *Real and Functional Analysis*. GTM 142. Springer Verlag, third edition, 1996.

[6] Serge Lang. *Undergraduate Analysis*. UTM. Springer Verlag, second edition, 1997.

[7] Walter Rudin. *Real and Complex Analysis*. McGraw Hill, third edition, 1987.

[8] Laurent Schwartz. *Topologie Générale et Analyse Fonctionnelle*. Collection Enseignement des Sciences. Hermann, 1980.

[9] Laurent Schwartz. *Analyse I. Théorie des Ensembles et Topologie*. Collection Enseignement des Sciences. Hermann, 1991.

[10] Laurent Schwartz. *Analyse III. Calcul Intégral*. Collection Enseignement des Sciences. Hermann, 1993.

[11] Laurent Schwartz. *Analyse IV. Applications à la Théorie de la Mesure*. Collection Enseignement des Sciences. Hermann, 1993.