An Introduction to the Theory of Hilbert Spaces

Eberhard Malkowsky

Abstract

These are the lecture notes of a mini–course of three lessons on Hilbert spaces taught by the author at the First German–Serbian Summer School of Modern Mathematical Physics in Sokobanja, Yugoslavia, 13–25 August, 2001. The main objective was to present the fundamental definitions and results and their proofs from the theory of Hilbert spaces that are needed in applications to quantum physics. In order to make this paper self–contained, Section 1 was added; it contains well–known basic results from linear algebra and functional analysis.

1 Notations, Basic Definitions and Well–Known Results

In this section, we give the notations that will be used throughout. Furthermore, we shall list the basic definitions and concepts from the theory of linear spaces, metric spaces, linear metric spaces and normed spaces, and deal with the most important results in these fields. Since all of them are well known, we shall only give the proofs of results that are directly applied in the theory of Hilbert spaces.

Throughout, let $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of positive integers, real and complex numbers, respectively. For $n \in \mathbb{N}$, let $\mathbb{R}^n$ and $\mathbb{C}^n$ be the sets of n–tuples $x = (x_1, x_2, \ldots, x_n)$ of real and complex numbers. If S is a set then $|S|$ denotes the cardinality of S. We write $\aleph_0$ for the cardinality of the set $\mathbb{N}$.

In the proof of Theorem 4.8, we need

Theorem 1.1 The Cantor–Bernstein Theorem
If each of two sets allows a one–to–one map into the other, then the sets are of equal cardinality.

We shall frequently make use of *Hölder's* and *Minkowski's inequalities.*

Theorem 1.2 Hölder's and Minkowski's Inequalities
Let $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{C}$.
(a) If $p > 1$ *and* $q = p/(p-1)$ *then*

$$\sum_{k=1}^{n} |a_k b_k| \leq \left(\sum_{k=1}^{n} |a_k|^p \right)^{1/p} \cdot \left(\sum_{k=1}^{n} |b_k|^q \right)^{1/q} \tag{1}$$

(Hölder's inequality).
(b) If $1 \leq p < \infty$ *then*

$$\left(\sum_{k=1}^{n} |a_k + b_k|^q \right)^{1/p} \leq \left(\sum_{k=1}^{n} |a_k|^p \right)^{1/p} + \left(\sum_{k=1}^{n} |b_k|^p \right)^{1/p} \tag{2}$$

(Minkowski's inequality).

1.1 Partially ordered sets and the axiom of transfinite induction

A *partially ordered set* is a set for which a *transitive* and *reflexive binary relation* is defined. We shall write $a \geq b$ to denote that the ordered pair (a, b) is in relation. Our definition requires that

if $a \geq b$ and $b \geq c$ then $a \geq c$ (*transitivity*) and

$$a \geq a \text{ for all } a \text{ (\textit{reflexivity}).}$$

Finally $a \leq b$ means $b \geq a$. The most obvious example is $\mathbb{R}$, taking $a \geq b$ with its usual meaning.

Example **1.3** *In* $\mathbb{R}^2$ *we define* $(x, y) \geq (u, v)$ *to mean* $x \geq u$ *and* $y \geq v$. *Let* $\mathcal{A}$ *be a collection of subsets of a set* S. *For* $A, B \in \mathcal{A}$, *we define* $A \geq B$ *to mean* $A \subset B$. *This ordering is called* ordering by inclusion.

We see from Example 1.3 that a partially ordered set may have incomparable elements; $(5, 2)$ and $(4, 3)$ are incomparable. We call a and b *comparable* if either $a \geq b$ or $b \geq a$; otherwise they are *incomparable*.

A partially ordered set is called *totally ordered* if the ordering is *antisymmetric*, that is

$$a \geq b \text{ and } b \geq a \text{ together imply } a = b,$$

and all elements are comparable.

A *chain* is a totally ordered subset of a partially ordered set. For example, in $\mathbb{R}$, any set is a chain. In Example 1.3, the set $\{(n, n) : n = 1, 2, \ldots\}$ is a chain, also $\{(t, t + 1) : t \in \mathbb{R}\}$. A *maximal* chain is one not properly included in any chain. The chain $\{(n, n) : n+1, 2, \ldots\}$ is not maximal, since it is included in $\{(t, t) : t \in \mathbb{R}\}$. This latter chain is maximal. The chain $\{(t, t + 1) : t \in \mathbb{R}\}$ is maximal in Example 1.3. These two maximal chains are incomparable (neither includes the other), thus neither is a *maximum* (or largest) chain.

Axiom of transfinite induction. *Every partially ordered set includes a maximal chain.*

This axiom can also be expressed in the form

Theorem 1.4 *Every chain is included in a maximal chain.*

1.2 Linear spaces and linear maps

The concepts of linear spaces and linear maps are fundamental in many fields of mathematics. *Linear spaces* are nonempty sets which are given an algebraic structure.

Let $\mathbb{F}$ denote a field throughout, in general $\mathbb{F} = \mathbb{C}$, the field of complex numbers, or $\mathbb{F} = \mathbb{R}$, the field of real numbers.

First we give the definition of a *linear* or *vector space*.

Definition 1.5 *Let $\mathbb{F}$ be a field. A* **linear space** *or* **vector space** *V (***over $\mathbb{F}$***) is a set for which are defined an* addition $+ : (V \times V) \to V$, $+(v, w) = v + w$ *for all* $v, w \in V$ *and a* multiplication by scalars $\cdot : (\mathbb{F} \times V) \to V$, $\cdot(\lambda, v) = \lambda v$ *for all* $\lambda \in \mathbb{F}$ *and for all* $v \in V$*, such that* V *is an* Abelian group *with respect to* $+$ *and the following* distributive laws *are satisfied for all* $\lambda, \mu \in \mathbb{F}$ *and all* $v, w \in V$

(D.1) $\lambda(v + w) = \lambda v + \lambda w$

(D.2) $(\lambda + \mu)v = \lambda v + \mu v$

(D.3) $(\lambda\mu)v = \lambda(\mu v)$

(D.4) $1v = v$ *for the* unit element *of* $\mathbb{F}$

The elements of a linear space V *over a field* $\mathbb{F}$ *are called* **vectors** *and the elements of* $\mathbb{F}$ *are referred to as* **scalars***.*
A subset S *of a linear space is called* **convex** *if* $\lambda s + \mu t \in S$ *for all* $s, t \in S$ *and all scalars* $\lambda, \mu \geq 0$ *with* $\lambda + \mu = 1$*.*
A subset S *of a linear space* V *is called* **linear subspace of V** *if it is a linear space with the same operations and scalars as* V*, that is, if* $s, t \in S$ *and* $\lambda \in \mathbb{F}$ *imply* $s + t \in S$ *and* $\lambda s \in S$*.*

Usually the operation of an Abelian group is denoted by $+$ in which case the neutral element is denoted by 0.

Example **1.6** *(a) On $I\!R^n$ we define*

$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \text{ for all } x, y \in I\!R^n$$
$$\lambda x = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n) \text{ for all } x \in I\!R^n \text{ and all } \lambda \in I\!R.$$

Then $I\!R^n$ is a linear space over $I\!R$. Similarly $\mathbb{C}^n$ becomes a linear space over $\mathbb{C}$. Let $p > 0$ and $C_p = \{x \in I\!R^n : \sum_{k=1}^n |x_k|^p \le 1\}$. Then C_p is a convex set if and only if $p \ge 1$. If $p \ge 1$, let $x, y \in C$ and $\lambda, \mu \ge 0$ satisfy $\lambda + \mu = 1$. Then, by Minkowski's inequality

$$\left(\sum_{k=1}^n |\lambda x_k + \mu y_k|^p\right)^{1/p} \le \lambda \left(\sum_{k=1}^n |x_k|^p\right)^{1/p} + \mu \left(\sum_{k=1}^n |y_k|^p\right)^{1/p} \le 1,$$

hence $\lambda x + \mu y \in C_p$. If $0 < p < 1$, we consider $x = e^{(1)} = (1, 0, \ldots, 0), y = e^{(2)} = (0, 1, 0, \ldots, 0) \in C_p$. Then $z = (1/2)(x+y) \notin C_p$, since $\sum_{k=1}^n |z_k|^p = 2^{-p+1} > 1$.

(b) Let M be a set, V be a linear space over $I\!F$ and V^M denote the set of all functions from M to V. We define $f + g$ and λf for all $f, g \in V^M$ and for all $\lambda \in I\!F$ by

$$(f + g)(m) = f(m) + g(m) \text{ and } (\lambda f)(m) = \lambda f(m) \text{ for all } m \in M.$$

Then V^M is a linear space over $I\!F$. In particular, let $M = I\!N$ and $V = \mathbb{C}$. Then $\omega = \mathbb{C}^{I\!N} = \{x = (x_k)_{k=1}^\infty : x_k \in \mathbb{C}$ for all $k\}$, the set of all complex sequences, is a linear space with

$$x + y = (x_k + y_k)_{k=1}^\infty \text{ and } \lambda x = (\lambda x_k)_{k=1}^\infty \text{ for all } x, y \in \omega \text{ and all } \lambda \in \mathbb{C}.$$

Let $1 \le p < \infty$. We put $\ell_p = \{x \in \omega : x \in \omega : \sum_{k=1}^\infty |x_k|^p < \infty\}$. Then ℓ_p is a linear subspace of ω; this follows from Minkowski's inequality.
Let $C[0,1]$ denote the set of all continuous real–valued functions on the

interval $[0,1]$. *Then* $[0,1]$ *is a linear subspace of* $\mathbb{R}^{[0,1]}$.
Let $1 \le p < \infty$ *and* $L_p[0,1]$ *be the set of all Lebesgue measurable functions* f *on the interval* $[0,1]$ *such that* $\int_0^1 |f|^p < \infty$. *Then* $L_p[0,1]$ *is a linear space.*

Let S be a subset of a linear space V. A *linear combination of* S is an element $\sum_{k=1}^n \lambda_k s_k$ where $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{F}$ and $s_1, s_2, \ldots, s_n \in S$ are distinct. Any linear combination with at least one nonzero scalar coefficient is called *nontrivial*; 0 is the *trivial* linear combination of every set, for the empty set, this is taken as a convention. The set S is called *linearly dependent* if 0 is a nontrivial linear combination of S; otherwise it is called *linearly independent.* The set of all linear combinations of S is called *span of* S denoted by span$S$; S is said to *span a set* T if $T \subset$ span$S$.

We observe that linear combinations are *finite* sums. By definition, a set S in a linear space is linearly independent if and only if, whenever $s_1, s_2, \ldots, s_n \in S$, $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{F}$ and $\sum_{k=1}^n \lambda_k s_k = 0$ then $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$. A set is linearly independent, if each finite subset is. Every subset of a linearly independent set is linearly independent.

Example **1.7** *(a) In* $\mathbb{C}^n$, *the set* $B = \{e^{(k)} : k = 1, 2, \ldots, n\}$ *of vectors with* $e_k^{(k)} = 1$ *and* $e_j^{(k)} = 0$ *for* $j \neq k$ *is linearly independent, and span*$B = \mathbb{C}^n$. *The set* $B \cup \{e\}$ *with* $e = (1, 1, \ldots, 1)$ *is linearly dependent, since* $e - \sum_{k=1}^n e^{(k)} = 0$ *is a nontrivial linear combination of* B.
(b) Let $M = [0,1]$. *For each* $t \in M$, *we define* $f_t \in V^M$ *by* $f_t(t) = 1$ *and* $f_t(x) = 0$ *for* $x \in M \setminus \{t\}$. *Then* $B = \{f_t : t \in M\}$ *is linearly independent, for if* $f_{t_1}, f_{t_2}, \ldots, f_{t_n} \in B$ *and* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *are scalars, then* $\sum_{k=1}^n \lambda_k f_{t_k} = 0$ *means* $\sum_{k=1}^n \lambda_k f_{t_k}(x) = 0$ *for all* $x \in M$. *If we choose* $x = t_j$ *for* $j = 1, 2, \ldots$ *then*

$$\sum_{k=1}^{n} \lambda_k f_{t_k}(t_j) = \lambda_j = 0 \textit{ for } j = 1, 2, \ldots, n.$$

Let $g \in V^M$ be defined by $g(x) = 1$ for all $x \in M$. Then $g \in V^M \setminus spanB$, and so B does not span V^M. In particular, we consider the set $B = \{e^{(n)} : n = 1, 2, \ldots\}$ in ω where, for each n, $e^{(n)}$ is the sequence with $e^{(n)}_n = 1$ and $e^{(n)}_k = 0$ for $k \neq n$. Then B is linearly independent, and $e = (1, 1, \ldots) \in \omega \setminus spanB$.

An *algebraic* or *Hamel basis* for a linear space V is a linearly independent set which spans V.

Example **1.8** *(a) The set $B = \{e^{(1)}, e^{(2)}, \ldots, e^{(n)}\}$ is an algebraic basis for $\mathbb{C}^n$.*
(b) The set $B = \{e^{(n)} : n = 1, 2, \ldots\}$ is not an algebraic basis for ω since $e \in \omega \setminus spanB$.

Theorem 1.9 *Every linear space has an algebraic basis. Any two algebraic bases of a linear space are in a one–to–one correspondence.*

In view of Theorem 1.9, we define the *algebraic* or *Hamel dimension of a linear space* as the cardinality of its algebraic basis. The *algebraic* or *Hamel dimension* of a set in a linear space is defined to be the algebraic dimension of its span.

Now we study *linear maps* which naturally arise in connection with linear spaces.

Definition 1.10 *Let V and W be linear spaces over the same field $\mathbb{F}$. A map $f : V \to W$ is said to be a* **linear map** *or a* **homomorphism***, if for all $u, v \in V$ and all $\lambda \in \mathbb{F}$,*

$$f(u + v) = f(u) + f(v) \ (\text{additivity})$$

and

$$f(\lambda v) = \lambda f(v) \ (\text{homogenity});$$

the set of all linear maps from V to W is denoted by $L(V,W)$. If V is a linear space over $\mathbb{C}$ and $W = \mathbb{C}$, we write $V^{\#} = L(V,\mathbb{C})$ and each $f \in V^{\#}$ is called a linear functional *on V. Similarly, if V is a linear space over $\mathbb{R}$ and $W = \mathbb{R}$, we write $V^{\#}_{\mathbb{R}} = L(V,\mathbb{R})$ and each $f \in V^{\#}_{\mathbb{R}}$ is called a* real linear functional *on V.*
A one–to–one linear map is called an **isomorphism into***. We say that two linear spaces are* **isomorphic** *if there is an isomorphism from each onto the other.*

Example **1.11** *(a) Let $\mathcal{P}$ be the linear space of polynomials. Then $D : \mathcal{P} \to \mathcal{P}$, the differentiation operator, is linear.*
(b) Let $\ell_\infty = \{x \in \omega : \sup |x_k| < \infty\}$ and $c_0 = \{x \in \omega : \lim_{k\to\infty} x_k = 0\}$ denote the sets of bounded *and* null sequences*. Then the map $f : \ell_\infty \to c_0$ defined by $f(x) = (x_k/k)_{k=1}^\infty$ for all $x = (x_k)_{k=1}^\infty \in \ell_\infty$ is linear. The function $g : c_0 \to c_0$ defined by $g(x) = (x_k^2)_{k=1}^\infty$ for all $x = (x_k)_{k=1}^\infty \in c_0$ is not linear.*

Theorem 1.12 *An additive map f is* rational homogeneous*, that is*

$$f(\lambda v) = \lambda f(v) \text{ for all rational } \lambda \text{ and all vectors } v.$$

Proof. Since f is additive, it follows by induction that

$$f(mx) = mf(x) \text{ for all } m \in \mathbb{N} \text{ and all } x \in X. \tag{3}$$

Furthermore $f(0) + f(0) = f(0 + 0) = f(0)$ implies

$$f(0) = 0, \tag{4}$$

and $f(x) + f(-x) = f(x + (-x)) = f(0) = 0$ implies

$$f(-x) = -f(x) \text{ for all } x \in X. \tag{5}$$

Thus, if $m \in -\mathbb{N}$, we put $n = -m \in \mathbb{N}$ and obtain from (3) and (5)

$$f(mx) = f(-nx) = -f(nx) = -nf(x) = mf(x),$$

hence

$$f(mx) = mf(x) \text{ for all } m \in \mathbb{Z} \text{ and all } \in X. \tag{6}$$

Now let $r \in \mathbb{Q}$. Then there are $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ such that $r = m/n$ and

$$\begin{aligned} f(rx) &= \frac{1}{n} nf(rx) = \frac{1}{n} f(nrx) = \frac{1}{n} f(mx) \\ &= \frac{m}{n} f(x) = rf(x) \text{ for all } x \in X. \end{aligned}$$

It turns out in Theorem 1.13 that a linear functional is uniquely determined by its real part. Specifically, let V be a (complex) linear space and $f \in V^{\#}$. We define g and h on V by $g(v) = \mathrm{Re} f(v)$ and $h(v) = \mathrm{Im} f(v)$ for all $v \in V$. Thus $f(v) = g(v) + \mathrm{i}h(v)$ for each $v \in V$, and g and h are real–valued, *real linear* functions.

Theorem 1.13 *Let g be a real–valued linear function defined on a (complex) linear space V. Then there exists a unique real–valued function h defined on V such that $f = g + ih \in V^{\#}$.*

Proof. First we prove the uniqueness of such h.
We suppose that $f = g + \mathrm{i}h$ is linear. Then for any $v \in V$,

$$g(\mathrm{i}v) + \mathrm{i}h(\mathrm{i}v) = f(\mathrm{i}v) = \mathrm{i}f(v) = \mathrm{i}g(v) - h(v).$$

Comparing real parts yields $h(v) = -g(\mathrm{i}v)$. This shows that h is uniquely determined and real linear.
Next we show that such an h exists by constructing it.
We define h by $h(v) = -g(\mathrm{i}v)$ or all $v \in V$. Let $f = g + \mathrm{i}h$, that is

$$f(v) = g(v) - \mathrm{i}g(\mathrm{i}v) \text{ for all } v \in V. \tag{7}$$

Clearly f is additive and real linear. Thus we have only to show $f(\mathrm{i}v) = \mathrm{i}f(v)$ for all $v \in V$. But

$$\begin{aligned} f(\mathrm{i}v) &= g(\mathrm{i}v) + \mathrm{i}h(\mathrm{i}v) = g(\mathrm{i}v) - \mathrm{i}g(\mathrm{i}\mathrm{i}v) = g(\mathrm{i}v) - \mathrm{i}g(-v) = g(\mathrm{i}v) + \mathrm{i}g(v) \\ &= -h(v) + \mathrm{i}g(v) = \mathrm{i}(g(v) + \mathrm{i}h(v)) = \mathrm{i}f(v). \end{aligned}$$

1.3 Metric and normed spaces

Metric spaces are sets with a topological structure that arises from the concept of *distance*. *Continuity of functions* and *convergence of sequences* can be defined in metric spaces.

Definition 1.14 *Let X be a set. A function $d : X \times X \to I\!R$ is called a* **semimetric** *if for all $x, y, z \in X$*

(D.1) $\quad d(x, y) = d(y, x) \geq 0$

(D.2) $\quad d(x, x) = 0$

(D.3) $\quad d(x, y) \leq d(x, z) + d(z, y)$ *(*triangle inequality*);*

the number $d(x, y)$ is called the **distance between x and y**. *The set X together with the semimetric d is called a* **semimetric space**, *denoted by (X, d). A* **metric** *on X is a semimetric satisfying*

(D.4) $\quad d(x, y) > 0$ *if $x \neq y$;*

the set X together with the metric d is called a **metric space**.

Example **1.15** *(a) The set $I\!R^2$ is a semimetric space with d defined by $d(x, y) = |x_1 - y_1|$ for all $x = (x_1, x_2), y = (y_1, y_2) \in I\!R^2$.*

(b) The set ω is a metric space with d defined by

$$d(x,y) = \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{|x_k - y_k|}{|1 + |x_k - y_k|} \text{ for all } x = (x_k)_{k=1}^{\infty}, y = (y_k)_{k=1}^{\infty} \in \omega.$$

(c) The set ℓ_∞ of bounded sequences is a metric space with d_∞ defined by

$$d_\infty(x,y) = \sup_k |x_k - y_k| \text{ for all } x = (x_k)_{k=1}^{\infty}, y = (y_k)_{k=1}^{\infty} \in \ell_\infty.$$

The sets ℓ_p $(1 \leq p < \infty)$ are metric spaces with d_p defined by

$$d_p(x,y) = \left(\sum_{k=1}^{\infty} |x_k - y_k|^p \right)^{1/p} \text{ for all } x = (x_k)_{k=1}^{\infty}, y = (y_k)_{k=1}^{\infty} \in \ell_p;$$

the triangle inequality follows from Minkowski's inequality.

(d) The set $C[0,1]$ is a metric space with d and ρ defined by

$$d(f,g) = \max_{x \in [0,1]} |f(x) - g(x)| \text{ and } \rho(f,g) = \int_0^1 |f(x) - g(x)|\, dx$$

for all $f, g \in C[0,1]$.

Let (X,d) be a semimetric space, $\delta > 0$ and $x_0 \in X$. By

$$U_\delta(x_0) = \{x \in X : d(x, x_0) < \delta\}$$

we denote the *open δ neighbourhood of x_0*.

We already mentioned that the concept of convergence can be introduced in semimetric spaces. A sequence $(x_n)_{n=1}^{\infty}$ in a semimetric space X is said to be *convergent with limit x*, or *converge to x*, if given any $\varepsilon > 0$ there is $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that $x_n \in U_\varepsilon(x)$ for all $n \geq n_0$; this is denoted by $x_n \to x$ $(n \to \infty)$. A *Cauchy sequence* in a semimetric space is a sequence $(x_n)_{n=1}^{\infty}$ such that to each $\varepsilon > 0$, there corresponds an $N \in \mathbb{N}$ such that $d(x_m, x_n) < \varepsilon$ for all $m, n > N$. A semimetric space is called *complete* if every Cauchy sequence in it is convergent.

Remark 1.16 *(a) It is easy to see that, in a semimetric space,* $x_n \to x$ $(n \to \infty)$ *if and only if* $d(x_n, x) \to 0$ $(n \to \infty)$.
(b) It is possible for a sequence in a semimetric space to converge to different limits. For instance, if $d(x, y) = 0$ *for distinct* x *and* y *then the sequence* $(x_n)_{n=1}^{\infty}$ *with* $x_n = x$ *for all* n *converges to both* x *and* y*. In metric spaces, however, limits are unique, for if* $x_n \to x$ *and* $x_n \to y$ $(n \to \infty)$ *then* $d(x, y) \leq d(x, x_n) + d(x_n, y) \to 0$ $(n \to \infty)$ *by Part (a), and so* $d(x, y) = 0$ *which implies* $x = y$*. If a sequence* $(x_n)_{n=1}^{\infty}$ *has a unique limit* x *we write* $\lim_{n\to\infty} x_n = x$.

Now we introduce the concept of continuity of functions in semimetric spaces. Let (X, d_X) and (Y, d_Y) be semimetric spaces. A function $f : X \to Y$ is said to be *continuous at* $x \in X$ if or every $\varepsilon > 0$ there is $\delta > 0$ such that $f(U_\delta(x)) \subset U_\varepsilon(f(x))$; here, of course, $U_\delta(x) = \{x' \in X : d_X(x', x) < \delta\} \subset X$ and $U_\varepsilon(f(x)) = \{y \in Y : d_Y(y, f(x)) < \varepsilon\} \subset Y$; f is said to be *continuous on* X if it is continuous at each point of X.

Theorem 1.17 *Let* X *and* Y *be semimetric spaces. A function* $f : X \to Y$ *is continuous at* $x \in X$ *if and only if* $f(x_n) \to f(x)$ $(n \to \infty)$ *whenever* $(x_n)_{n=1}^{\infty}$ *is a sequence in* X *converging to* x.

Let X and Y be semimetric spaces. A function $f : X \to Y$ is called an *isometry* if it is one–to–one and $d(f(x), f(y)) = d(x, y)$ for all $x, y \in X$. Two spaces are called *isometric* if there is an isometry from each into the other.

A subset S of a semimetric space X is called *dense*, if given $x \in X$ and $\varepsilon > 0$ there is $s \in S$ with $d(x, s) < \varepsilon$. A semimetric space is called *separable* if it has a countable dense set in it.

1.4 Linear metric spaces

So far we studied linear and semimetric spaces separately. To join the two concepts, some connection should be assumed between them. The natural assumption is that the algebraic operations of the linear space should be continuous with respect to the semimetric.

A useful notion is that of a *paranorm.* A *paranorm* is a real function p defined on a linear space X such that for all $x, y \in X$

(P.1) $p(0) = 0$

(P.2) $p(x) \geq 0$

(P.3) $p(-x) = p(x)$

(P.4) $p(x + y) \leq p(x) + p(y)$
(*triangle inequality; continuity of addition)*

(P.5) if $(\lambda_n)_{n=1}^{\infty}$ is a sequence of scalars with $\lambda_n \to \lambda$ $(n \to \infty)$ and $(x_n)_{n=1}^{\infty}$ is a sequence of vectors with $p(x_n - x) \to 0$ $(n \to \infty)$, then $p(\lambda_n x_n - \lambda x) \to 0$ $(n \to \infty)$ (*continuity of multiplication by scalars*).

A paranorm p for which $p(x) = 0$ implies $x = 0$ is called *total.*

Example **1.18** *A paranorm p defines a semimetric d which is a metric if and only if p is total. We put $d(x, y) = p(x - y)$ for all x and y. Then $d(x, y) = p(x - y) = p(-(x - y)) = p(y - x) = d(y, x)$, $d(x, x) = p(x - x) = p(0) \geq 0$ and $d(x, y) = p(x - y) = p(x - z + z - y) \leq p(x - z) + p(z - y) = d(x, z) + d(z, y)$.*

A semimetric d for a linear space X is said to be *invariant* if there exists a paranorm p such that $d(x, y) = p(x - y)$ for all $x, y \in X$. We say that the paranorm *yields* the semimetric and the semimetric *comes from* the paranorm. Since $p(x) = d(x, 0)$, a given semimetric comes from at most one paranorm. Clearly an invariant semimetric d satisfies the equation

$$d(x+z, y+z) = d(x, y) \text{ for all } x,\ y \text{ and } z, \tag{8}$$

hence its name. The discrete metric, however, that is the metric d with $d(x, x) = 0$ and $d(x, y) = 1$ for all $x \neq y$, satisfies equation (8), but is not an invariant metric. For if we put $p(x) = 1$ for $x \neq 0$ and $p(0) = 0$ and choose $\lambda_n = 1/n$ and $x_n = x \neq 0$ for all n, then $\lambda_n \to \lambda = 0$ and $p(x_n - x) \to 0$ $(n \to \infty)$, but $p(\lambda_n x_n - \lambda x) = p((1/n)x) = 1$ for all n, that is $p(\lambda_n x_n - \lambda x) \not\to 0$ $(n \to \infty)$ and multiplication by scalars is not continuous.

Definition 1.19 *A semimetric space which is also a linear space is called a* **linear semimetric space***, if the semimetric comes from a paranorm; it is called a* **linear metric space***, if the semimetric comes from a total paranorm.*

The set of all continuous linear functionals on a linear semimetric space X will be denoted by X' and called the *dual space of* X.

Theorem 1.20 *Let X be a linear semimetric space, Y be a linear metric space and $f : X \to Y$ be a continuous additive function. Then f is real linear.*

Proof. By Theorem 1.12, the map f is rational homogeneous. Let $\lambda \in \mathbb{R}$ and $(\lambda_n)_{n=1}^{\infty}$ be a sequence of rationals converging to λ. Then for any $x \in X$, $\lambda_n x \to \lambda x$ by (P.5) of the definition of a paranorm. Since f is continuous this implies

$$f(\lambda_n x) \to f(\lambda x) \ (n \to \infty).$$

Also $\lambda_n f(x) \to \lambda f(x)$ $(n \to \infty)$ for each x. Since $f(\lambda_n x) = \lambda_n f(x)$ for each $x \in X$ $(n = 1, 2, \ldots)$ and the limits in metric spaces are unique by Remark 1.16, we have

$$f(\lambda x) = \lambda f(x) \text{ for all } x \in X.$$

A subset S of a linear semimetric space X is called *fundamental* if span$S$ is dense in X, that is, given any vector $x \in X$ and any $\varepsilon > 0$, there is a linear combination y of S such that $d(x, y) < \varepsilon$. A *Schauder basis* of a linear metric space X is a sequence $(b_n)_{n=1}^{\infty}$ of elements of X such that for any vector $x \in X$, there exists a unique sequence $(\lambda_n)_{n=1}^{\infty}$ of scalars such that $\sum_{n=1}^{\infty} \lambda_n b_n = x$. Every linear metric space with a Schauder basis is separable. For a finite–dimensional space the concepts of Schauder and algebraic basis coincide. Also in ϕ, the set of all sequences that terminate in zeros, with the metric of ℓ_∞, $(e^{(n)})_{n=1}^{\infty}$ is both a Schauder and an algebraic basis. In all other interesting cases, however, the concepts differ. For example, in c_0, $B = (e^{(n)})_{n=1}^{\infty}$ is a Schauder basis, but not an algebraic basis, since span$B = \phi$ and ϕ is a proper subset of c_0. On the other hand, any algebraic basis of c_0 is uncountable, and so is automatically not a Schauder basis. Whereas every linear space has an algebraic basis, ℓ_∞ has no Schauder basis, since it is not separable.

1.5 Normed spaces

A *seminorm* is a paranorm which has an additional homogenity property. Later we shall consider Hilbert spaces as special normed spaces.

Definition 1.21 *Let X be a linear space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. A map $\|\cdot\| : V \to \mathbb{R}$ is called a* **seminorm** *if it satisfies the following conditions for all vectors x and y and for all scalars λ*

(N.1) $\|x\| \geq 0$

(N.2) $\|\lambda x\| = |\lambda| \, \|x\|$ *(*homogenity*)*

(N.3) $\|x + y\| \leq \|x\| + \|y\|$ *(*triangle inequality);

a **norm** *is a seminorm which satisfies*

(N.4) $\|x\| > 0$ *if* $x \neq 0$.

A linear space X with a seminorm is called a **seminormed space***, and a linear space with a norm is called a* **normed space***.*

Remark 1.22 *A seminorm is a paranorm. The conditions (P.1) to (P.4) of a paranorm are trivial. Condition (P.5) follows from*

$$\begin{aligned}\|\lambda_n x_n - \lambda x\| &= \|\lambda_n (x_n - x) + (\lambda_n - \lambda) x\| \\ &\leq |\lambda_n| \, \|x_n - x\| + |\lambda_n - \lambda| \, \|x\| \to 0 \\ &\quad \text{if } \lambda_n \to \lambda \text{ and } \|x_n - x\| \to 0 \ (n \to \infty).\end{aligned}$$

Thus every seminormed space is a linear semimetric space, and the concepts of Cauchy sequences, convergence, continuity and completeness can be translated from semimetric to seminormed spaces.

A normed space is called *Banach space* if it is complete with respect to the metric given by its norm.

Example **1.23** *(a) The sets* $\mathbb{R}^n$ *and* $\mathbb{C}^n$ *are Banach spaces with*

$$\|x\|_\infty = \max_{1 \leq k \leq n} |x_k|.$$

Any two norms $\| \cdot \|$ *and* $\| \cdot \|'$ *are equivalent on* $\mathbb{R}^n$ *or* $\mathbb{C}^n$*, that is, there are constants* C_1 *and* C_2 *such that*

$$C_1 \|x\| \leq \|x\|' \leq C_2 \|x\| \text{ for all } x$$

(b) Let $1 \le p < \infty$*. Then* ℓ_p *is a Banach space with*

$$\|x\|_p = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} \quad \textit{for all } x \in \ell_p.$$

Also $L_p[0,1]$ *is a Banach space with*

$$\|f\|_p = \left(\int_0^1 |f|^p \right)^{1/p} \quad \textit{for all } f \in L_p[0,1].$$

The set $C[0,1]$ *is a normed space with*

$$\|f\| = \int_0^1 |f(x)|\, dx \textit{ for all } x \in C[0,1]$$

which is not a Banach space; it is a Banach space with

$$\|f\|_\infty = \sup\{|f(x)| : x \in [0,1]\} \textit{ for all } f \in C[0,1].$$

Now we turn to *bounded linear maps.* Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. Then $L \in L(X,Y)$ is called bounded if

$$\|L\| = \sup\{\|L(x)\|_Y : \|x\|_X \le 1\} < \infty; \tag{9}$$

$\|L\|$ is said to be the *norm of* L; the set of all bounded $L \in L(X,Y)$ is denoted by $B(X,Y)$. In the special case of $Y = \mathbb{C}$, we write $X^* = B(X,\mathbb{C})$ for the set of all bounded linear functionals on X. It is well known, that *if* X *and* Y *are normed spaces and* $L \in L(X,Y)$ *then* L *is bounded if and only if it is continuous. If* X *and* Y *are Banach spaces, then* $B(X,Y)$ *is a Banach space with the norm* $\|\cdot\|$ *defined in (9); in particular,* X^* *is a Banach space with* $\|\cdot\|$ *defined by*

$$\|f\| = \sup\{|f(x)| : \|x\| \le 1\}. \tag{10}$$

A normed space is called *uniformly convex* if, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|x\|, \|y\| \leq 1 \text{ and } |x + y\| > 2 - \delta \text{ together imply } \|x - y\| < \varepsilon.$$

If we write the third inequality as $\|(1/2)(x + y)\| > 1 - \delta/2$, we see that a uniformly convex space is one such that if x and y are far apart members of the unit disk, then their mid–point must be deep within it, since $\|x-y\| \geq \varepsilon$ implies $\|(1/2)(x + y)\| \leq 1 - \delta/2$. Uniform convexity has to do with the roundness of the unit sphere.

Example **1.24** *Let* $I\!R^2$ *have the norm* $\|(x, y)\| = |x| + |y|$. *We choose* $e^{(1)} = (1, 0)$ *and* $e^{(2)} = (0, 1)$. *Then* $\|e^{(1)} + e^{(2)}\| = 2 > 2 - \delta$ *for all* $\delta > 0$ *while* $\|e^{(1)} - e^{(2)}\| = 2$. *Thus the space is not uniformly convex.*

We close this section with the famous *Hahn-Banach extension theorem* which is one of the fundamental theorems in functional analysis. It has applications in embedding theory, representation theory and existence theory, as well as in classical analysis.

Theorem 1.25 The Hahn–Banach extension theorem
Let S *be a linear subspace of a linear space* V, $\|\cdot\|$ *be a seminorm defined on* V *and* $f \in S^{\#}$ *with* $|f(s)| \leq \|s\|$ *for all* $s \in S$. *Then there is an extension* $F \in V^{\#}$ *of* f *with* $|F(v)| \leq \|v\|$ *for all* $v \in V$.

The next two corollaries of Theorem 1.25 will be applied in Section 6.

Corollary 1.26 *Let* X *be a seminormed space,* S *a linear subspace of* X *and* $x \in X \setminus S$. *Then there exists* $f \in X^{\#}$ *with* $f(x) = 1$, $f(s) = 0$ *for all* $s \in S$ *and* $\|f\| = 1/d(x, S)$ *where* $d(x, S) = \inf\{d(x, s) : s \in S\}$; *we intend* $\|f\| = \infty$ *if* $d(x, S) = 0$.

Corollary 1.27 *Let* X *be a seminormed space and* $x \in X$ *with* $\|x\| \neq 0$. *Then there is* $f \in X^{\#}$ *with* $f(x) = \|x\|$ *and* $\|f\| = 1$.

2 Hilbert Spaces

In this section, we introduce the concept of *Hilbert spaces*. We shall restrict ourselves to a study of Hilbert spaces as a special kind of normed spaces, namely as *Banach spaces* with their norms given by *inner products*. Throughout this section, a linear space will always be understood as a linear space over $\mathbb{C}$, unless explicitly stated otherwise.

Definition 2.1 *Let V be a linear space. A map $< \cdot, \cdot >: V \times V \to \mathbb{C}$ is called an* **inner** *or* **scalar product** *if it satisfies the following conditions for all vectors u, v and w and all scalars λ*

(SP.1) $< u + v, w >=< u, w > + < v, w >$

(SP.2) $< \lambda v, w >= \lambda < v, w >$

(SP.3) $< w, v >= \overline{< v, w >}$ *(*the complex conjugate*)*

(SP.4) $< v, v > > 0$ *if* $v \neq 0$

A linear space with an inner product is said to be an **inner product space***.*

Conditions (SP.1) and (SP.2) together mean that an inner product is linear in its first variable. A *real inner product* is defined similarly except that it is real valued, real linear in its first variable, and satisfies $< v, w >=< w, v >$ for all vectors v and w. The real case is always similar and we shall assume it covered.

Example **2.2** *As examples of inner products we have, in*
(a) the space $\mathbb{C}^n$,

$$< x, y >= \sum_{k=1}^{n} x_k \overline{y_k}$$

for all n–tuples $x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n) \in \mathbb{C}^n$*;*

(b) the spaces ℓ_p *for* $1 \le p \le 2$*,*

$$< x, y >= \sum_{k=1}^{\infty} x_k \overline{y_k} \text{ for all sequences } x = (x_k)_{k=1}^{\infty}, y = (y_k)_{k=1}^{\infty} \in \ell_p;$$

this series always converges, since

$$\sum_{k=1}^{\infty} |x_k \overline{y_k}| \le \left(\sum_{k=1}^{\infty} |x_k|^2 \right)^{1/2} \left(\sum_{k=1}^{\infty} |y_k|^2 \right)^{1/2}$$

by Hölder's inequality;

(c) the spaces $L_2[0,1]$ *and* $C[0,1]$*,* $< f, g >= \int_0^1 f(t)\overline{g(t)}\, dt$*.*

(d) Any linear space can be given an inner product. If H *is an algebraic basis and* $x = \sum \lambda h,\ y = \sum \mu h,$ *set* $< x, y >= \sum \lambda \bar{\mu}$*. (We observe that only a finite number of scalars are unequal to zero in each sum.)*

The next result is very important and can be applied to introduce a norm on inner product spaces.

Theorem 2.3 Cauchy–Schwarz–Bunyakowski Inequality

Let V *be an inner product space. Then*

$$| < v, w > |^2 \le < v, v >< w, w > \text{ for all } v, w \in V; \tag{11}$$

equality holds in (11) if and only if v *and* w *are linearly dependent.*

Proof. Since the result is trivial if $w = 0$, we assume $w \neq 0$. Let $\lambda =< w, w >$ and $\mu =< v, w >$. Then $\lambda \in \mathbb{R}$, $\lambda > 0$ by (SP.4), $\bar{\mu} =< w, v >$ and

$$\begin{aligned} 0 &\le < \lambda v - \mu w, \lambda v - \mu w > \\ &=< \lambda v, \lambda v > - < \lambda v, \mu w > - < \mu w, \lambda v > + < \mu w, \mu w > \\ &= \lambda^2 < v, v > - \lambda \bar{\mu} \mu - \mu \lambda \bar{\mu} + \mu \bar{\mu} \lambda = \lambda^2 < v, v > - \lambda |\mu|^2. \end{aligned}$$

Thus $| < v, w > |^2 = |\mu|^2 \leq \lambda < v, v > = < v, v > < w, w >$, since $\lambda > 0$.
If $| < v, w > |^2 = < v, v > < w, w >$ then both ends in the above string of inequalities are zero and so $< \lambda v - \mu w, \lambda v - \mu w > = 0$, hence $\lambda v - \mu w = 0$ and $\lambda \neq 0$. Thus v and w are linearly dependent.
Conversely if v and w are linearly dependent, we may assume that $v, w \neq 0$ and then there is $\mu \neq 0$ such that $v - \mu w = 0$. This implies

$$< v, v > = \mu < w, v > \text{ and } < v, w > = \mu < w, w >, \text{ hence}$$

$$\mu | < v, w > |^2 = \mu < v, v > < w, w > .$$

Since $\mu \neq 0$, this implies $| < v, w > |^2 = < v, v > < w, w >$.

We now construct a norm from a given inner product.

Theorem 2.4 *Let V be a linear space with inner product $< \cdot, \cdot >$. Then*

$$\|v\| = \sqrt{< v, v >} \text{ for all } v \in V \tag{12}$$

defines a norm on V.

Proof. Obviously $\|v\| \geq 0$, and $\|v\| > 0$ if $v \neq 0$. Furthermore $\|\lambda v\|^2 = < \lambda v, \lambda v > = \lambda \bar{\lambda} < v, v > = |\lambda|^2 \|v\|^2$, so that $\|\lambda v\| = |\lambda| \cdot \|v\|$ for every scalar λ and every vector v. Finally, by (11)

$$\begin{aligned}
\|v + w\|^2 &= < v + w, v + w > \\
&= < v, v > + < v, w > + < w, v > + < w, w > \\
&= \|v\|^2 + 2\mathrm{Re}(< v, w >) + \|w\|^2 \\
&\leq \|v\|^2 + 2| < v, w > | + \|w\|^2 \\
&\leq \|v\|^2 + 2\|v\| \cdot \|w\| + \|v\|^2 = (\|v\| + \|w\|)^2.
\end{aligned}$$

This proves the triangle inequality.

Now we give the definition of a *Hilbert space*.

Definition 2.5 *An inner product space is called* **Hilbert space** *if it is a Banach space with the norm constructed by its inner product as in (12).*

Example **2.6** *Every finite dimensional inner product space is a Hilbert space. In particular, the space* $\mathbb{C}^n$ *is a Hilbert space with the* Euclidean norm *defined by*

$$\|x\| = (\sum_{k=1}^{n} |x_k|^2)^{1/2} \text{ for all } x = (x_1, x_2, \ldots, x_n) \in \mathbb{C}^n.$$

The space ℓ_2 *is a Hilbert space with its natural norm*

$$\|x\|_2 = \left(\sum_{k=1}^{\infty} |x_k|^2 \right)^{1/2} \text{ for all sequences } x = (x_k)_{k=1}^{\infty} \in \ell_2.$$

The space $L_2[0,1]$ *is a Hilbert space with* $\|f\|_2 = (\int_0^1 |f|^2)^{1/2}$ *for all* $f \in L_2[0,1]$.

The fundamental property of an inner product, real or complex, is the parallelogram law.

Theorem 2.7 Parallelogram Law
Let V *be an inner product space and* $\|\cdot\|$ *the norm constructed by its inner product as in (12). Then*

$$\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2 \text{ for all } v, w \in V. \tag{13}$$

Proof. The proof follows from

$$\|v \pm w\|^2 = < v \pm w, v \pm w > = \|v\|^2 \pm 2\mathrm{Re}(< v, w >) + \|w\|^2.$$

The parallelogram law is expressed in terms of the norm, the inner product not occurring explicitly. Any norm not satisfying the parallelogram law is not derived from an inner product.

Example **2.8** *The spaces* ℓ_p $(1 \leq p < \infty, p \neq 2)$ *are not Hilbert spaces. We put* $v = e^{(1)} = (1, 0, 0, \ldots)$ *and* $w = e^{(2)} = (0, 1, 0, \ldots)$. *Then, for* $p \neq 2$,

$$\|v+w\|_p^2 + \|v-w\|_p^2 = 2^{2/p} + 2^{2/p} \neq 2 + 2 = 2\|v\|_p^2 + 2\|w\|_p^2.$$

Theorem 2.9 *Let V be an inner product space, $w \in V$ be fixed and $f : V \to \mathbb{C}$ defined by $f(v) =< v, w >$ for all $v \in V$. Then $f \in V^*$ and $\|f\| = \|w\|$.*

Proof. We mean of course that f is continuous as a function on the inner product space considered as a normed space.
The linearity of f is an immediate consequence of (SP.1) and (SP.2), and the continuity follows from (11), since $|f(v)| \leq \|v\| \cdot \|w\|$, and so $\|f\| \leq \|w\|$. Furthermore $f(w) = \|w\|^2$ also implies $\|f\| \geq \|w\|$.

Since $< v, w >= \overline{< w, v >}$, it follows that for fixed $v \in V$, the function $g : V \to \mathbb{C}$ defined by $g(w) =< v, w >$ for all $w \in V$ is continuous.

Corollary 2.10 *Let V be an inner product space. Then*

$$v^{\perp} = \{w \in V :< v, w >= 0\}$$

is a closed linear subspace of V.

Proof. It is obvious that $v^{\perp}$ is a linear subspace of V. That it is closed follows from the continuity of the function f in Theorem 2.9.

Corollary 2.11 *Let V be an inner product space. If $\sum w_n$ is a convergent series of vectors $w_n \in V$, then $< \sum w_n, v >= \sum < w_n, v >$ for all $v \in V$, that is $< \cdot, \cdot >$ is infinitely additive.*

Proof. This is an immediate consequence of Theorem 2.9

Some norms are not derived from inner products. A norm derived from an inner product must satisfy the parallelogram law. The converse of this is also true.

Theorem 2.12 Jordan–von Neumann
A norm satisfying the parallelogram law is derived from an inner product.

Proof. (i) *First assume that V is a real normed space.*

We put

$$< v, w >= \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right) \text{ for all } v, w \in V, \tag{14}$$

and prove show $< \cdot, \cdot >$ is a real inner product.

It is clear that $< v, w >=< w, v >$ for all $v, w \in V$, and $< v, v >= \|v\|^2 > 0$ for $v \neq 0$, that is, (SP.3) and (SP.4) are satisfied.

Now we prove (SP.1). Applying the parallelogram law to $u+w$ and $v+w$, $u-w$ and $v-w$, we obtain

$$\begin{aligned}
< u, w > + < v, w > &= \frac{1}{4}(\|u+w\|^2 + \|v+w\|^2) \\
&\quad -\frac{1}{4}(\|u-w\|^2 + \|v-w\|^2) \\
&= \frac{1}{8}(\|u+w+v+w\|^2 + \|u+w-v-w\|^2) \\
&\quad -\frac{1}{8}(\|u-w+v-w\|^2 + \|u-w-v+w\|^2) \\
&= \frac{1}{8}(\|u+v+2w\|^2 - \|u+v-2w\|^2) = \frac{1}{2} < u+v, 2w > .
\end{aligned}$$

This proves

$$< u, w > + < v, w >= \frac{1}{2} < u+v, 2w > \text{ for all } u, v, w \in V. \tag{15}$$

If we put $v = 0$ in (15), then $< u, w >= (1/2) < u, 2w >$ for all $u, w \in V$. Replacing u by $u+v$ in this relation and applying (15), we obtain

$$< u+v, w >= \frac{1}{2} < u+v, 2w >=< u, w > + < v, w > \text{ for all } u, v, w \in V.$$

Thus we have shown (SP.1).

Since $< v, w >$ is an additive function of v for each fixed w which is continuous by Theorem 2.9, it follows from Theorem 1.20 that it is linear. So (SP.2) is satisfied. Thus we have shown that $< \cdot, \cdot >$ is a real inner product.

(ii) *Now let V be a complex linear space.*

We use relation (14) to define a product which will turn out to be the real part of an inner product. We put

$$< v, w >_{\mathbb{R}}= \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right) \text{ for all } v, w \in V.$$

For fixed w, this defines a real–valued, real linear function. By Theorem 1.13, there exists, for each w, a linear function $< v, w >$ of v with $< v, w >_{\rm I\!R} = \mathrm{Re} < v, w >$, indeed we have

$$< v, w > = < v, w >_{\rm I\!R} - \mathrm{i} < \mathrm{i}v, w >_{\rm I\!R} . \qquad (16)$$

Using relation (14), we observe

$$\begin{aligned} < \mathrm{i}v, v >_{\rm I\!R} &= \frac{1}{4}\left(\|\mathrm{i}v + v\|^2 - \|\mathrm{i}v - v\|^2\right) = \frac{1}{4}\left(\|(1+\mathrm{i})v\|^2 - \|\mathrm{i}(1+\mathrm{i})v\|^2\right) \\ &= \frac{1}{4}\left(\|(1+i)v\|^2 - |\mathrm{i}| \cdot \|(1+\mathrm{i})v\|^2\right) = 0, \end{aligned}$$

that is $< \mathrm{i}v, v >_{\rm I\!R} = 0$ for all $v \in V$. Therefore (16) implies $< v, v > = < v, v >_{\rm I\!R} = \|v\|^2$ for all $v \in V$. It also follows from (14) that $< \mathrm{i}v, \mathrm{i}w >_{\rm I\!R} = < v, w >_{\rm I\!R}$ for all $v, w \in V$. Finally, making use of this identity, we conclude

$$\begin{aligned} < w, v > &= < w, v >_{\rm I\!R} - \mathrm{i} < \mathrm{i}w, v >_{\rm I\!R} = < v, w >_{\rm I\!R} - \mathrm{i} < w, -\mathrm{i}v >_{\rm I\!R} \\ &= < v, w >_{\rm I\!R} + \mathrm{i} < \mathrm{i}v, w >_{\rm I\!R} = \overline{< v, w >} \text{ for all } v, w \in V. \end{aligned}$$

3 The Conjugate Space

We shall deal with a fixed Hilbert space H and investigate the continuous linear functionals on H. We already know some of them. In Theorem 2.9 we saw that each $a \in H$ leads to an $f \in H^*$ by means of the formula $f(x) = < x, a >$ for all $x \in H$; indeed, $\|f\| = \|a\|$. We shall see in this section that all elements of H^* are of this form; the map from H to H^* just described is actually onto. This is the statement of the famous *F. Riesz representation theorem* 3.7.

In the proof of Theorem 3.7, we need some results on uniformly convex spaces.

Theorem 3.1 *An inner product space is uniformly convex.*

Proof. Let $\varepsilon > 0$ be given. We choose $\delta = \varepsilon^2/4$. If $\|v\|, \|w\| = 1$ and $\|v + w\| > 2 - \delta$, then it follows from the parallelogram law that

$$\begin{aligned}\|v-w\|^2 &= 2\|v\|^2 + 2\|w\|^2 - \|v+w\|^2 \le 4 - \|v+w\|^2 \le 4 - (2-\delta)^2 \\ &= 4\delta - \delta^2 < 4\delta = \varepsilon^2.\end{aligned}$$

Example **3.2** *The space $I\!R^2$ with $\|(x,y)\| = \sqrt{x^2+y^2}$ for all $(x,y) \in I\!R^2$ is uniformly convex.*
The space l_2 with its natural norm is uniformly convex.

Lemma 3.3 *A convex set in an absolutely convex normed space has at most one point of minimum norm. In other words, let C be a convex subset of an absolutely convex normed space X and $d = \inf\{\|c\| : c \in C\}$. Then C meets the ball $B(0,d) = \{x \in X : \|x\| \le d\}$ at most once.*

Proof. The result is obvious for $d = 0$. Therefore let $d > 0$. We assume that $c_1, c_2 \in C$ are two distinct points with $\|c_1\| = \|c_2\| = d$. Then $c = (1/2)(c_1 + c_2) \in C$, since C is convex. Furthermore, X absolutely convex implies $\|c\| < d$. For if we assume $\|c\| \ge d$ and choose $a = (1/d)c_1$, $b = (1/d)c_2$ and $\varepsilon = \|a - b\|$, then $\varepsilon > 0$ since $c_1 \ne c_2$; moreover $\|a\| = \|b\| = 1$ and

$$\|a+b\| = \left\| \frac{1}{d}(c_1 + c_2) \right\| = \frac{1}{d}\|c_1 + c_2\| \ge 2 > 2 - \delta$$

for every $\delta > 0$, hence X is not absolutely convex. The existence of $c \in C$ with $\|c\| < d$ is a contradiction to the definition of d. Therefore C can at most contain one point with norm d.

Lemma 3.4 *If a sequence $(a_n)_{n=1}^\infty$ in a uniformly convex normed space satisfies the conditions $\|a_n\| \to 1$ and $\|a_m + a_n\| \to 2$, then it is a Cauchy sequence.*

Proof. The condition $\|a_m + a_n\| \to 2$ means, given $\delta > 0$, there exists N such that

$$| \, \|a_m + a_n\| - 2 \, | < \delta \text{ for all } m, n > N. \tag{17}$$

The proof will be given in two cases.

(i) *First we assume* $\|a_n\| \leq 1$ *for all* n.

Let $\varepsilon > 0$ be given and $\delta > 0$ be chosen such that the conditions $\|a\|, \|b\| \leq 1$, $\|a + b\| > 2 - \delta$ imply $\|a - b\| < \varepsilon$. Such a choice is possible, since the space is uniformly convex. There exists N such that condition (17) holds. This implies $\|a_m + a_n\| > 2 - \delta$ for all $m, n > N$ and so $\|a_m - a_n\| < \varepsilon$ for all $m, n > N$ by the choice of δ.

(ii) Since $\|a_n\| \to 1$, we may assume $\|a_n\| \neq 0$ for all n. Then

$$\begin{aligned} 2 &\geq \left\| \frac{a_n}{\|a_n\|} + \frac{a_m}{\|a_m\|} \right\| = \left\| \frac{a_n + a_m}{\|a_n\|} + \frac{\|a_n\| - \|a_m\|}{\|a_n\| \cdot \|a_m\|} \cdot a_m \right\| \\ &\geq \frac{\|a_n + a_m\| - | \, \|a_n\| - \|a_m\| \, |}{\|a_n\|} \to 2 \text{ as } m, n \to \infty. \end{aligned}$$

Thus by case (i), $(a_n/\|a_n\|)_{n=1}^{\infty}$ is a Cauchy sequence. Hence $(a_n)_{n=1}^{\infty}$ is also a Cauchy sequence. To see this, let $\lambda_n = \|a_n\|$ for all n, so that $\lambda_n \to 1$ $(n \to \infty)$. Furthermore, let $c_n = a_n/\|a_n\|$ for all n, so that $\|c_n\| = 1$ and $(c_n)_{n=1}^{\infty}$ is a Cauchy sequence. Then

$$\begin{aligned} &\|a_m - a_n\| = \|\lambda_m c_m - \lambda_n c_n\| \\ &\leq |\lambda_m - 1| \cdot \|c_m\| + \|c_m - c_n\| + |1 - \lambda_n| \cdot \|c_n\| \to 0 \text{ as } m, n \to \infty. \end{aligned}$$

Theorem 3.5 *In a uniformly convex Banach space X, every closed convex set C has a unique element of smallest norm.*

Proof. If $0 \in C$ then there is nothing to prove. If $0 \notin C$ then $d = \inf\{\|c\| : c \in C\} > 0$, since C is closed. Let $(c_n)_{n=1}^{\infty}$ be a sequence of points of C with $\|c_n\| \to d$ $(n \to \infty)$. Then for any m and n, $\|c_m + c_n\| = 2\|(1/2)(c_m + c_n)\| \geq 2d$ since $(1/2)(c_m + c_n) \in C$ by the convexity of C, and so, since also $\|c_n\| \to d\ (n \to \infty)$, we have $\|c_m + c_n\| \to 2d\ (m, n \to \infty)$. Applying Lemma 3.4 to the sequence $(c_n/d)_{n=1}^{\infty}$, we conclude that $(c_n/d)_{n=1}^{\infty}$ is a Cauchy sequence, and hence so is $(c_n)_{n=1}^{\infty}$. Since X is complete, the sequence $(c_n)_{n=1}^{\infty}$ is convergent, $c_n \to c\ (n \to \infty)$, say. Then $c \in C$, since C is closed, and $\|c\| = d$. That the element of smallest norm is unique follows from Lemma 3.3.

Corollary 3.6 *A nonzero continuous linear functional f on a uniformly convex Banach space X assumes its maximum exactly once on the unit disk.*

Proof. We apply Theorem 3.5 to the set $C = \{x \in X : f(x) = \|f\|\}$.

Now we are able to prove the famous *F. Riesz representation theorem* which is one of the most important results in the theory of Hilbert spaces.

Theorem 3.7 The F. Riesz representation theorem
Let $f \in H^*$. Then there exists a unique vector $a \in H$ with

$$f(x) =< x, a > \text{ for all } x \in H; \text{ moreover, } \|a\| = \|f\|.$$

Proof. We may assume $f \neq 0$.
(i) First we show the uniqueness of the representation
If such an $a \in H$ exists then it will be unique. For $< x, a >=< x, b >$ for all $x \in H$ implies $< x, a - b >= 0$ for all $x \in H$. In particular, we have $< a - b, a - b >= 0$, hence $a = b$.
(ii) Now we show the existence of the representation and $\|f\| = \|a\|$
Since H is uniformly convex by Theorem 3.1, f assumes its maximum at some point b on the unit disk by Corollary 3.6. Then $\|b\| = 1$ and $f(b) = \|f\|$, by the definition of $\|f\|$. We put $a = \|f\|b$. Then

$$f(a) = \|f\|f(b) = \|f\|^2 \text{ and } \|a\| = \|f\|. \tag{18}$$

(iii) Finally we show $f(x) = < x, a >$ for all $x \in H$.

Let $f^{\perp} = \{x \in H : f(x) = 0\}$. First we observe that

$$\|a + \lambda y\| \geq \|a\| \text{ for all } y \in f^{\perp} \text{ and all scalars } \lambda. \tag{19}$$

This follows from Lemma 3.3, since $f^{\perp}$ is a linear subspace, $S = a + f^{\perp}$ is a convex set, and so S has at most one point of minimum norm, namely a. Therefore

$$< a, y > = 0 \text{ for all } y \in f^{\perp}.$$

To see this, we may assume $\|y\| = 1$. Then by (19)

$$0 \leq \|a - < a, y > y\|^2 - \|a\|^2 = -| < a, y > |^2, \text{ and so } < a, y > = 0.$$

Hence we have shown

$$f(y) = 0 \text{ implies } < y, a > = 0. \tag{20}$$

It follows from (20) that there exists a scalar λ such that

$$< x, a > = \lambda f(x) \text{ for all } x \in H. \tag{21}$$

To see this, let $c \notin f^{\perp}$ and put

$$\lambda = \frac{< c, a >}{f(c)}.$$

Let $x \in H$ be given and put

$$y = x - \frac{f(x)}{f(c)} \cdot c.$$

Then $f(y) = 0$ and (20) implies

$$< x, a > = < y, a > + \frac{f(x)}{f(c)} < c, a > = \lambda f(x).$$

This shows (21). Putting $x = a$ in (21) and using (19), we obtain

$$\|a\|^2 = \lambda f(a) = \lambda \|f\|^2, \text{ hence } \lambda = 1.$$

As an immediate consequence of the F. Riesz representation theorem we obtain

Corollary 3.8 *Let H be a Hilbert space and $T : H \to H^*$ defined by*

$$Ta = f_a \text{ where } f_a : H \to \mathbb{C} \text{ is given by } f_a(x) =< x, a > \quad (x \in H).$$

Then T is a conjugate–linear isometry onto.

Proof. By Theorem 3.7, T is an isometry onto. Let $a, b \in H$ and λ be a scalar. Then, for all $x \in H$,

$$\begin{aligned} T(a+b)(x) &= f_{a+b}(x) =< x, a+b >=< x, a > + < x, b > \\ &= f_a(x) + f_b(x) = (Ta)(x) + (Tb)(x) = (Ta + Tb)(x), \end{aligned}$$

hence $T(a+b) = Ta + Tb$, and

$$\begin{aligned} T(\lambda a)(x) &= f_{\lambda a}(x) =< x, \lambda a >= \overline{\lambda} < x, a >= \overline{\lambda} f_a(x) \\ &= \overline{\lambda}(Ta)(x) = (\overline{\lambda} Ta)(x), \end{aligned}$$

hence $T(\lambda a) = \overline{\lambda} Ta$.

4 Orthonormal Sets in Inner Product Spaces

The material of the next two sections is a generalization of parts of Fourier analysis, and, at the same time, of the study of Euclidean spaces. The underlying concept is that of *orthogonality*.

The following will illustrate the geometric meaning of the scalar product in $\mathbb{R}^2$. Let v and w be nonzero vectors in $\mathbb{R}^2$. Then

$$\|w - v\|^2 = \|v\|^2 + \|w\|^2 - 2 < v, w > .$$

On the other hand, by the *cosine law*, the angle ϕ between v and w satisfies

$$\|w - v\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\| \, \|w\| \cos \phi.$$

Thus $< v, w >= \|v\| \, \|w\| \cos \phi$. If $\|v\| = 1$, then $< v, w >= \|w\| \cos \phi$, that is $< v, w >$ is the length of the projection of w on the straight line in the direction of the vector v. So, if V is a real inner product space and $v, w \in V \setminus \{0\}$ then the *angle* ϕ *between* v *and* w is defined by

$$\cos\phi = \frac{< v, w >}{\|v\|\,\|w\|} \text{ with } \phi \in [0, \pi).$$

This leads to

Definition 4.1 *Let* V *be an inner product space, and* $v, w \in V \setminus \{0\}$. *If* $< v, w >= 0$, *then* v *and* w *are said to be* **orthogonal**; *we denote this by* $v \perp w$. *If* $v \perp w$ *and* $\|v\| = \|w\| = 1$ *then* v *and* w *are said to be* **orthonormal**. *A subset* M *of* V *is said to be* **orthogonal**, *if* $x \perp y$ *for all distinct members* x *and* y *of* M; *it is said to be* **orthonormal**, *if it is orthogonal and* $\|x\| = 1$ *for all* $x \in M$.

Example **4.2** *(a) In* $\mathbb{C}^n$, *the set* $\{e^{(1)}, e^{(2)}, \ldots, e^{(n)}\}$ *is an orthonormal set.*
(b) In ℓ_2, *the set* $\{e^{(n)} : n \in \mathbb{N}\}$ *is an orthonormal set.*
(c) Let $C[-\pi, \pi]$ *be given the real inner product*

$$< f, g >= \int_{-\pi}^{\pi} f(t)g(t)\,dt \text{ for all } f, g \in C[-\pi, \pi]$$

and the functions s_n $(n = 1, 2, \ldots)$ *be defined by* $s_n(t) = (1/\sqrt{\pi}) \sin nt$. *Then the set* $\{s_n : n \in \mathbb{N}\}$ *is an orthonormal set.*

If the class of orthonormal sets is *ordered by inclusion*, we may take a *maximal chain* by the *axiom of transfinite induction.* Thus we have, taking the union of the maximal chain, that *there exists a maximal orthonormal set in any inner product space. An orthonormal set is maximal if there exists no vector except* 0 *which is orthogonal to all its members.* If an orthonormal set is *fundamental* then it is maximal. The sets in Example 4.2 (a) and (b) are maximal, the set in Example 4.2 (c) is not, since $\cos \perp s_n$ for all n.

The next result shows that a linearly independent set can always be orthonormalized.

Theorem 4.3 The Gram–Schmidt orthogonalization process
Let $S = \{s_k : 1 \le k \le n\}$ $(n \in I\!N)$ be a linearly independent set in an inner product space. Then there exists an orthonormal set $B = \{b_k : 1 \le k \le n\}$ such that, for each k with $1 \le k \le n$, the vector b_k is a linear combination of the set $\{s_1, s_2, \dots, s_k\}$, that is, for each k there are scalars λ_{kj} $(1 \le j \le k)$ such that

$$b_k = \sum_{j=1}^{k} \lambda_{kj} s_j. \tag{22}$$

Proof. Since S is a linearly independent set, $s_k \neq 0$ for all k. We put

$$b_1 = \frac{s_1}{\|s_1\|}, \text{ hence } \lambda_{11} = \frac{1}{\|s_1\|}.$$

Now we assume that, for some $k \ge 1$, the orthonormal vectors b_j $(1 \le j \le k)$ have been determined to satisfy (22). Then we put

$$\tilde{b}_{k+1} = \sum_{j=1}^{k} \tilde{\lambda}_{k+1,j} b_j - s_{k+1}. \tag{23}$$

Since each vector b_j is a linear combination of $\{s_1, s_2, \dots, s_j\}$, the sum is a linear combination of $\{s_1, s_2, \dots, s_k\}$, and by the linear independence of the set $\{s_1, s_2, \dots, s_{k+1}\}$, we have $\tilde{b}_{k+1} \neq 0$. Now the orthogonality condition yields

$$\begin{aligned} 0 &=< \tilde{b}_{k+1}, b_m >=< \sum_{j=1}^{k} \tilde{\lambda}_{k+1,j} b_j, b_m > - < s_{k+1}, b_m > \\ &= \sum_{j=1}^{k} \tilde{\lambda}_{k+1,j} < b_j, b_m > - < s_{k+1}, b_m >= \tilde{\lambda}_{k+1,m} - < s_{k+1}, b_m > \end{aligned}$$

for $1 \le m \le k$, that is

$$\tilde{\lambda}_{k+1,j} =< s_{k+1}, b_j > \text{ for } 1 \le j \le k. \tag{24}$$

If we choose the scalars $\tilde{\lambda}_{k+1,j}$ $(1 \le j \le k)$ as in (24) then $\tilde{b}_{k+1} \perp \{b_1, b_2, \dots, b_n\}$. Now we put

$$b_{k+1} = \frac{\tilde{b}_{k+1}}{\|\tilde{b}_{k+1}\|} = \frac{1}{\|\tilde{b}_{k+1}\|} \left(\sum_{j+1}^{k} < s_{n+1}, b_j > b_j - s_{k+1} \right).$$

If we substitute the linear combinations (22) for $b_1, b_2, \ldots, b_k$ in this representation, then b_{k+1} also has a representation (22) and

$$\lambda_{k+1,k+1} = -\frac{1}{\|\tilde{b}_{k+1}\|}.$$

Example **4.4** *We consider $C[-1,1]$ with the inner product*

$$< f, g >= \int_{-1}^{1} f(t)g(t)\,dt \text{ for all } f, g \in C[-1,1]$$

and the set $S = \{p_n : n \in I\!N_0\}$ of powers $p_n(t) = t^n$ $(t \in [-1,1])$. From Theorem 4.3, we obtain

$$\tilde{L}_0(t) = \sqrt{\tfrac{1}{2}},\ \tilde{L}_1(t) = \sqrt{\tfrac{3}{2}}\,t,\ \tilde{L}_2(t) = \sqrt{\tfrac{5}{3}}\left(\tfrac{3}{2}t^2 - \tfrac{1}{2}\right),$$
$$\tilde{L}_3(t) = \sqrt{\tfrac{7}{2}}\left(\tfrac{5}{2}t^3 - \tfrac{3}{2}t\right), \ldots.$$

Each $\tilde{L}_n$ is a polynomial of degree n; the polynomial L_n with

$$L_n(t) = \sqrt{\frac{2}{2n+1}}\tilde{L}_n(t)$$

are called Legendre polynomials*.*
Similarly, the Gram–Schmidt process applied to S on $C[-1,1]$ with the inner product

$$< f, g >= \int_{-1}^{1} f(t)g(t)w(t)\,dt \text{ where } w(t) > 0 \text{ on } [-1,1]$$

yields the Chebyshev polynomials of first *and* second kind *for*

$$w(t) = \sqrt{1-t^2} \text{ and } w(t) = \frac{1}{\sqrt{1-t^2}}.$$

The Hermite *and* Laguerre polynomials H_n *and* L_n *are obtained by applying the Gram–Schmidt orthogonalization process to the sets* $\{t^n \exp(-t^2/2) : n = 0, 1, \ldots\}$ *on* $L_2(-\infty, \infty)$ *and* $\{t^n \exp(-t/2) : n = 0, 1, \ldots\}$ *on* $L_2(0, \infty)$*, respectively*

Now we consider a generalization of the classical Fourier coefficients.

Definition 4.5 *Let S be an orthonormal set in an inner product space V. For each $v \in V$, the set $\{< v, s >: s \in S\}$ is called the* **set of orthogonal coefficients of v (relative to S)**.

The familiar Fourier coefficients are a special case of orthogonal coefficients. In Example 4.2 (c), we have

$$< f, s_n >= \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(t) \sin nt \, dt \text{ for } f \in C[-\pi, \pi].$$

Theorem 4.6 *Let S be an orthonormal set in an inner product space V and $v \in V$. Then $< v, s >= 0$ for all but a countable set of $s \in S$, moreover, arranging those s with $< v, s > \neq 0$ in a sequence $(s_n)_{n=1}^{\infty}$, then*

$$\sum_{n=1}^{\infty} | < v, s_n > |^2 \leq \|v\|^2 \qquad \text{(Bessel's inequality).} \tag{25}$$

Furthermore, we have $v = \sum_{n=1}^{\infty} < v, s_n > s_n$ if and only if v satisfies

$$\sum_{n=1}^{\infty} | < v, s_n > |^2 = \|v\|^2 \qquad \text{(Parseval's relation).} \tag{26}$$

Proof. (i) *First let $S = \{s_1, s_2, \ldots, s_n\}$ be a finite orthonormal set.* Given $v \in V$, we put $\lambda_k =< v, s_k >$ for $k = 1, 2, \ldots, n$ and obtain

$$\begin{aligned}
&< v - \sum_{k=1}^{n} \lambda_k s_k, v - \sum_{k=1}^{n} \lambda_k s_k > \\
&= \|v\|^2 - < v, \sum_{k=1}^{n} \lambda_k s_k > - < \sum_{k=1}^{n} \lambda_k s_k, v > + < \sum_{k=1}^{n} \lambda_k s_k, \sum_{j=1}^{n} \lambda_j s_j > \\
&= \|v\|^2 - \sum_{k=1}^{n} \overline{\lambda}_k < v, s_k > - \sum_{k=1}^{n} \lambda_k < s_k, v > + \sum_{k=1}^{n} \sum_{j=1}^{n} \lambda_k \overline{\lambda}_j < s_k, s_j > \\
&= \|v\|^2 - \sum_{k=1}^{n} \overline{\lambda}_k \lambda_k - \sum_{k=1}^{n} \lambda_k \overline{\lambda}_k + \sum_{k=1}^{n} \lambda_k \overline{\lambda}_k = \|v\|^2 - \sum_{k=1}^{n} |\lambda_k|^2,
\end{aligned}$$

hence

$$0 \leq \left\| v - \sum_{k=1}^{n} \lambda_k s_k \right\|^2 = \|v\|^2 - \sum_{k=1}^{n} |\lambda_k|^2, \tag{27}$$

and so

$$\sum_{k=1}^{n} | < v, s_k > |^2 \leq \|v\|^2. \tag{28}$$

(ii) Now let $S = \{s_n : n \in I\!N\}$ be a countable orthonormal set.

Then inequality (28) holds for any n, so $(\lambda_k)_{k=1}^{\infty} \in \ell_2$ where $\lambda_k = < v, s_k >$ for $k = 1, 2, \ldots$, and $\sum_{k=1}^{\infty} |\lambda_k|^2 \leq \|v\|^2$.

(iii) Finally let S be an uncountable set.

Let $v \in V$ and $\varepsilon > 0$ be given. We put $T_\varepsilon = \{s \in S :< v, s > > \varepsilon\}$. Let $s_1, s_2, \ldots, s_n$ be distinct members of the set T_ε. Then

$$\|v\| \geq \sum_{k=1}^{n} | < v, s_k > |^2 \geq n\varepsilon^2,$$

hence $n \leq \|v\|^2/\varepsilon^2$. Therefore, the set T_ε is finite, and consequently the set

$$\{s \in S :< v, s >\neq 0\} = \bigcap_{n=1}^{\infty} T_{1/n}$$

is countable. Thus we have proved Bessel's inequality.

(iv) Finally we show Parseval's relation.

First we assume $v = \sum_{k=1}^{\infty} < v, s_k > s_k$. Letting $n \to \infty$ in (27) and taking into account that $(\lambda_k)_{k=1}^{\infty} \in \ell_2$, we conclude

$$0 = \left\| v - \sum_{k=1}^{\infty} \lambda_k s_k \right\|^2 = \|v\|^2 - \sum_{k=1}^{\infty} |\lambda_k|^2,$$

and (26) holds. Conversely, if (26) holds, then letting $n \to \infty$ again in (27), we conclude $\|v - \sum_{k=1}^{\infty} \lambda_k s_k\|^2 = 0$, hence $v = \sum_{k=1}^{\infty} \lambda_k s_k$.

Remark 4.7 *If $S = \{s_n : n \in I\!N\}$ is an orthonormal set then*

$$\lim_{n\to\infty} < v, s_n >= 0 \textit{ for all } v.$$

If specialized to Example 4.2 (c), this yields a special case of the Riemann–Lebesgue lemma.

Theorem 4.8 *Any two maximal orthonormal sets in an inner product space V are in a one–to–one correspondence.*

Proof. *(i) First we assume that there exists a finite maximal orthonormal set $S = \{s_1, s_2, \dots, s_n\}$.*
Given any $v \in V$, let $w = v - \sum_{k=1}^{n} < v, s_k > s_k$. Then, for each j,

$$< w, s_j > = < v, s_j > - < \sum_{k=1}^{n} < v, s_k > s_k, s_j > = < v, s_j > - < v, s_j > = 0.$$

This implies $w = 0$ by the maximality of S. Thus $V = \text{span} S$. Also S is linearly independent. Therefore S is an algebraic basis of V. Any other maximal orthonormal set, being linearly independent, has no more than n elements nor can it have fewer, since we can apply the above argument to it.
(ii) Now we assume that there are two maximal orthonormal sets B and H in V.
For each $b \in B$, let $H_b = \{h \in H : < b, h > \neq 0\}$. Every member of H occurs in at least one set H_b. For otherwise, if there were $h \in H$ with $h \notin H_b$ for all b, then $< b, h > = 0$ for all $b \in B$, a contradiction to the maximality of B. By Theorem 4.6, each set H_b is finite or countable. Hence, since $H = \bigcup_{b \in B} H_b$ as just proved, $|H| \leq |B| \aleph_0 = |B|$. By symmetry, also $|B| \leq |H|$. The result now follows from the *Cantor–Bernstein theorem.*

In view of Theorem 4.8, we may define the *dimension of an inner product space.*

Definition 4.9 *The* **dimension of an inner product space** *is the cardinality of one of its maximal orthonormal sets.*

Finite dimensional and *infinite dimensional* are unambiguous, since a finite maximal orthonormal set is an *algebraic basis* and if there exists an infinite orthonormal set, the space has infinite *algebraic dimension.*

Example **4.10** *The dimension of ℓ_2 is $\aleph_0$, since $\{e^{(n)} : n \in I\!N\}$ is a maximal orthonormal set.*

5 Orthonormal Sets in Hilbert Spaces

The assumption of completeness allows the results of the preceding section to be put in a more satisfactory form.

Theorem 5.1 *Let B be an orthonormal set in a Hilbert space H and $a \in H$ be given. Then the* orthogonal series $\sum_{b\in B} < a, b > b$ *is convergent.*
If $a' = \sum_{b\in B} < a, b > b$, then $(a - a') \perp B$, that means $(a - a') \perp b$ for all $b \in B$.

It is understood that the countable set of those $b \in B$ for which $< a, b > \neq 0$ (cf. Theorem 4.6) is arranged in a sequence. The value of the sum does not depend on this arrangement.

Proof. We fix $a \in H$. Let $S = \{b_n \in B :< a, b_n > \neq 0\}$. We put $\lambda_n =< a, b_n >$ for $n = 1, 2, \ldots$. If $S = \emptyset$, the result is trivial. Otherwise, for any $p, q \in I\!N$ with $q > p$, we have

$$\left\| \sum_{k=p}^{q} \lambda_k b_k \right\|^2 = \sum_{k=p}^{q} |\lambda_k|^2. \tag{29}$$

Since $(\lambda_k)_{k=1}^{\infty} \in \ell_2$ by Bessel's inequality, we see that $(\sum_{k=1}^{n} \lambda_k b_k)_{n=1}^{\infty}$ is a Cauchy sequence, hence convergent by the completeness of H. Let $a' = \sum_{k=1}^{\infty} \lambda_k b_k$. Then a' has the same orthogonal coefficients as a, for, if $b \notin S$, then $< a', b >= 0 =< a, b >$, and for each n

$$< a', b_n >= \sum_{k=1}^{\infty} \lambda_k < b_k, b_n >= \lambda_n =< a, b_n > .$$

In particular $< a - a', b >= 0$ for all $b \in B$, that is $a - a' \perp B$. Now we assume that a" is the sum obtained in the above construction by some other ordering of the elements of B which are not orthogonal to a. Both a' and a" are *infinite linear combinations* of B. Moreover, a" has the same orthogonal coefficients as a, hence as a'. This implies $(a' - a") \perp B$ and so $(a' - a") \perp (a' - a")$, since $a' - a"$ is a linear combination of B. Therefore $a' - a" = 0$.

Corollary 5.2 *Any maximal orthogonal subset of a Hilbert space is a basis for the space in the sense that* every vector is a unique infinite linear combination of the subset.

Theorem 5.3 *A Hilbert space has countable (or finite) dimension if and only if it is* separable.

Proof. *(i) First we assume that the dimension is uncountable.*
Let B be an uncountable orthonormal set. For each $b \in B$, the set

$$U_{1/2}(b) = \{x \in H : \|x - b\| < 1/2\}$$

satisfies $U_{1/2}(b) \cap B = \{b\}$, for if $b' \in B$ and $b \perp b'$, then $\|b - b'\|^2 = \|b\|^2 + \|b\|^2 = 2$. Thus there is an uncountable set of sets $U_{1/2}(b)$ in the Hilbert space, and so the space is not separable.
(ii) Now we assume that the dimension is countable or finite.
Then the space has a *Schauder basis* and so is separable.

Remark 5.4 *We have encountered three different types of bases,* algebraic, Schauder bases *and* bases in a Hilbert space. *These concepts should not be mixed.*
An algebraic basis B_V of a linear space V is a linearly independent set which spans the space that is, every element v of V is a unique finite *linear combination of the set B_V.*
A Schauder basis B_X for a linear metric space X is a sequence $(b_n)_{n=1}^{\infty}$ such that for every $x \in X$ there exists a uniquely defined sequence $(\lambda)_{n=1}^{\infty}$

of scalars such that $x = \sum_{n=1}^{\infty} \lambda_n b_n$.
Finally a basis B_H *of a Hilbert space* H *is a maximal orthonormal set, such that every* $x \in H$ *is a unique* infinite *linear combination of* B_H.
For finite dimensional Hilbert spaces the three concepts are the same. Every linear space has an algebraic basis, the space ℓ_∞ *of all bounded sequences has no Schauder basis. An orthonormal basis of a Hilbert space with countable dimension is a Schauder basis.*

Theorem 5.5 *Let* $\{b_n : n \in \mathbb{N}\}$ *be an orthonormal set in a Hilbert space* H *and* $(\lambda_n)_{n=1}^{\infty} \in \ell_2$. *Then the series* $\sum_{n=1}^{\infty} \lambda_n b_n$ *is convergent, and its orthogonal coefficients are* λ_n.
Conversely, if λ_n *are the orthogonal coefficients of some vector then* $(\lambda_n)_{n=1}^{\infty} \in \ell_2$.

Proof. *(i) First we assume* $(\lambda_n)_{n=1}^{\infty} \in \ell_2$.
For $p, q \in \mathbb{N}$ with $q > p$, we have (29). Thus $(\sum_{k=1}^{n} \lambda_k b_k)_{n=1}^{\infty}$ is a Cauchy sequence, hence convergent.
(ii) The converse part follows from Bessel's inequality (25).

Example **5.6** *The specialization of Theorem 5.5 to* $L_2[-\pi, \pi]$ *and the* trigonometric system *is called the* Riesz-Fischer theorem.

Theorem 5.7 *Two Hilbert spaces* H_1 *and* H_2 *with the same dimension are* congruent*, that is there is an isomorphism* $T : H_1 \to H_2$ *such that*

$$\|T(a)\| = \|a\| \text{ for all } a \in H_1.$$

Proof. Let B_1 and B_2 be bases with the same cardinality of H_1 and H_2. We assume that $f : B_1 \to B_2$ is a one–to–one correspondence from B_1 onto B_2. Given $a \in H_1$, we have $a = \sum_{b \in B_1} < a, b > b$, where at most a countable number of coefficients $< a, b > \neq 0$ in the sum by Theorem 4.6.

We put $T(a) = \sum_{b\in B_1} < a, b > f(b)$. This series converges by Theorem 5.5. We have $T|_{B_1} = f$, for the restriction of T on B, and it follows that

$$\|T(a)\|^2 = \sum_{b\in B_1} | < a, b > |^2 = \|a\|^2$$

by Parseval's relation (26). Therefore T is norm preserving. Obviously it is linear, thus it is an isometry. Finally, let $a_2 \in H_2$ be given. Then as before

$$\sum_{b\in B_2} | < a, b > |^2 < \infty.$$

This implies $a_1 = \sum_{b\in B_2} < a, b > f^{-1}(b) \in H_1$, and

$$\begin{aligned} T(a_1) &= T\left(\sum_{b\in B_2} < a, b > f^{-1}(b)\right) \\ &= \sum_{b\in B_2} < a, b > f(f^{-1}(b)) = \sum_{b\in B_2} < a, b > b = a. \end{aligned}$$

This shows that T is onto.

As an immediate consequence of Theorem 5.7 we obtain

Corollary 5.8 *Every n–dimensional Hilbert space is congruent with $\mathbb{C}^n$ (or $\mathbb{R}^n$). Every infinite dimensional separable Hilbert space is congruent with ℓ_2.*

Corollary 5.9 *Every Hilbert space is congruent with its dual space.*

Proof. Let H be a Hilbert space and H^* be its dual space. By Corollary 3.8, the map $T : H \to H^*$ with

$$T(a) = f_a \text{ where } f_a : H \to \mathbb{C} \text{ is given by } f_a(x) =< x, a > \ \ (x \in H)$$

is a conjugate–linear isometry onto. If $f, g \in H^*$ correspond to $a, b \in H$ then $< f, g >=< b, a >$. Let F be a maximal orthonormal set in H^*, and B be the set in H corresponding to F under the given correspondence between H^* and H. Then obviously B is orthonormal. Moreover it is maximal, for if $a \perp B$, then we consider $f \in H^*$ corresponding to a. Clearly $f \perp F$ and so $f = 0$ by the maximality of F. Thus $a = 0$. Therefore H and H^* have the same dimension and consequently they are congruent by Theorem 5.7.

6 Operators on Hilbert Spaces

In this section we study some properties of bounded operators and their adjoint operators.

Let H be a Hilbert space. A continuous linear map from H into itself is called a *bounded operator*; we write $B[H] = B(H, H)$; $B[H]$ is a Banach space. The members of $B[H]$ may be *multiplied by composition*. If $S, T \in B[H]$, then ST is defined by $(ST)a = S(Ta)$ for all $a \in H$. Then $ST \in B[H]$, indeed

$$\|ST\| \leq \|S\| \, \|T\|, \tag{30}$$

since

$$\|(ST)a\| = \|S(Ta)\| \leq \|S\| \, \|Ta\| \leq \|S\| \, \|T\| \, \|a\| \text{ for all } a \in H.$$

Relation (30) is referred to as the *multiplicative property of the norm.*

Now we introduce *adjoint operators*. First we consider the more general case of adjoint operators of linear operators between semimetric spaces.

Let X and Y be linear semimetric spaces and $T \in B(X, Y)$. The map $T^* : Y' \to X'$ defined by $T^* f = f \circ T$ for all $f \in Y'$ is called the *adjoint of* T.

We have to show that $T^* f \in X'$ for all $f \in Y'$. First we observe that

$$(T^* f)(x) = f(Tx) \text{ for all } f \in Y' \text{ and for all } x \in X,$$

$T^* f$ is a functional on X and its value at $x \in X$ is $f(Tx)$ by definition.

Since, for all $x, x' \in X$ and all scalars λ,

$$\begin{aligned}(T^*f)(x+x') &= (f(T(x+x')) = f(Tx+Tx') \\ &= f(Tx)+f(Tx') = (T^*f)(x)+(T^*f)(x')\end{aligned}$$

and

$$(T^*f)(\lambda x) = f(T(\lambda x)) = f(\lambda Tx) = \lambda f(Tx) = \lambda(T^*f)(x),$$

it follows that $T^*f \in X^\#$. Finally $T^*f = f \circ T$ is the composition of two continuous functions, hence $T^*f \in X'$.

Example **6.1** *Let $c_0 = \{x \in \omega : \lim_{k\to\infty} x_k = 0\}$ be the space of all null sequences, a normed space with $\|x\|_\infty = \sup_k |x_k|$ for all $x \in c_0$. We define the map $T : c_0 \to c_0$ by*

$$Tx = (x_k/k)_{k=1}^\infty \text{ for all } x = (x_k)_{k=1}^\infty \in c_0.$$

Then $\|T\| = 1$, since

$$\|Tx\|_\infty = \sup_k \left|\frac{1}{k}x_k\right| \leq \|x\|_\infty \text{ for all } x \in c_0$$

and $\|Te^{(1)}\|_\infty = \|e^{(1)}\|_\infty = 1$. Furthermore, T is one–to–one, but not onto, since $Tx \neq (\sqrt{k})_{k=0}^\infty$ for all $x \in c_0$. The range of T is dense, since it includes ϕ, the set of all finite sequences. We may consider $T^* : \ell_1 \to \ell_1$, since $c_0^* = \ell_1$, that means, every sequence $b \in \ell_1$ defines an $f \in c_0^*$ if we put

$$f(x) = \sum_{k=1}^\infty b_k x_k \text{ for all } x \in c_0, \tag{31}$$

and conversely, for each $f \in c_0^*$, there is a sequence $b \in \ell_1$ such that (31) holds. If $b \in \ell_1$ corresponds to $f \in c_0^*$, then, for all $x \in c_0$,

$$(T^*f)(x) = f(Tx) = \sum_{k=1}^\infty b_k \frac{x_k}{k} = \sum_{k=0}^\infty \frac{b_k}{k} x_k,$$

hence $T^*(x)$ corresponds to the sequence $(b_k/k)_{k=1}^\infty$.

A map T between seminormed spaces is said to be *norm increasing* if $\|Tx\| \geq \|x\|$ for all x.

Theorem 6.2 *Let X, Y and Z be seminormed spaces, $T, T_1 \in B(X, Y)$ and $T_2 \in B(Y, Z)$.*
(a) Then $T^* \in B(Y^*, X^*)$ and

$$\|T^*\| = \|T\|.$$

(b) The adjoint operator T^* is one–to–one if and only if the range of T is dense.
(c) If X is complete, Y is a normed space and T is norm increasing, then T^* is one–to–one if and only if T is onto.
(d) We have $(T_2T_1)^* = T_1^*T_2^*$.

Proof. (a) *(i) First we show that $T^* \in L(Y^*, X^*)$.*
Let $f, g \in Y^*$ and λ be a scalar. Then for all $x \in X$

$$\begin{aligned}(T^*(f+g))(x) &= (f+g)(Tx) = f(Tx) + g(Tx) \\ &= (T^*f)(x) + (T^*g)(x) = (T^*f + T^*g)(x)\end{aligned}$$

and

$$(T^*(\lambda f))(x) = (\lambda f)(Tx) = \lambda f(Tx) = \lambda (T^*f)(x),$$

hence $T^*(f+g) = T^*f + T^*g$ and $T^*(\lambda f) = \lambda T^*f$.
(ii) Now we show $\|T^*\| = \|T\|$.
For all $x \in X$ and all $f \in Y^*$, we have

$$|(T^*f)(x)| = |f(Tx)| \leq \|f\| \cdot \|Tx\| \leq \|f\| \cdot \|T\| \cdot \|x\|.$$

hence

$$\|T^*f\| \leq \|f\| \cdot \|T\| \text{ and so } \|T^*\| \leq \|T\|.$$

To prove the converse inequality, we observe that, by the definition of $\|T\|$, given $\varepsilon > 0$, there is $x \in X$ with $\|x\| \leq 1$ and

$$\|Tx\| > \|T\| - \varepsilon.$$

Furthermore, by Corollary 1.27 there is $f \in Y^*$ with $\|f\| = 1$ and $f(Tx) = \|Tx\|$. Then

$$(T^*f)(x) = f(Tx) = \|Tx\| > \|T\| - \varepsilon \text{ and } \|x\| \leq 1,$$

hence $\|T^*f\| \geq \|T\| - \varepsilon$, and since $\|f\| = 1$, we have

$$\|T^*\| \geq \|T^*f\| \geq \|T\| - \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\|T^*\| \geq \|T\|$.

(b) *First we assume that T^* is one–to–one.*

It is sufficient to show that the range $T(X)$ of T is fundamental, since it is obviously a linear subspace of Y. For this, by Corollary 1.26, it is sufficient to prove that any $f \in Y^*$ with $f(y) = 0$ for all $y \in T(X)$ is identically zero. Let $f \in Y^*$ and $f(y) = 0$ for all $y \in T(X)$. Then, for all $x \in X$, $(T^*f)(x) = f(Tx) = 0$, since $T(x) \in T(X)$. Thus $T^*f = 0$, and so $f \equiv 0$, since T^* is one–to–one.

Now we assume that $T(X)$ is dense.

If $T^*f = 0$ then $f(Tx) = (T^*f)(x) = 0$ for all $x \in X$. Thus $f(y) = 0$ for all $y \in T(X)$, and so $f \equiv 0$, since $T(X)$ is dense.

(c) *First we assume that T is onto.*

Then the range of T is dense and T^* is one–to–one by Part (b).

Now we assume that T^* is one–to–one.

Let $y \in Y$ be given. Since the range of T is dense by Part (b), we can find a sequence $(y_n)_{n=1}^{\infty}$ in $T(X)$ with $y_n \to y$ $(n \to \infty)$. Furthermore, for each n, there is $x_n \in X$ with $y_n = T(x_n)$. Since T is norm increasing,

$$\|x_p - x_q\| \leq \|T(x_p - x_q)\| = \|y_p - y_q\| \text{ for all } p, q \in \mathbb{N},$$

and so $(x_n)_{n=1}^{\infty}$ is a Cauchy sequence in X which converges by the completeness of X, say $x_n \to x$ $(n \to \infty)$. Then $Tx_n \to Tx$ $(n \to \infty)$, by the continuity of T, and $Tx_n = y_n \to y$ $(n \to \infty)$ together imply $T(x) = y$, since Y is a normed space. Thus we have shown that T is onto.

(d) First we observe that $T_2T_1 : X \to Z$ and $(T_2T_1)^* : Z^* \to X^*$ is given by

$$\begin{aligned} ((T_2T_1)^*h)(x) &= h(T_2T_1x) = h(T_2(T_1x)) = (T_2^*h)(T_1x) \\ &= (T_1^*(T_2^*(h))(x) = (T_1^*T_2^*(h))(x) \\ &\text{for all } h \in Z^* \text{ and for all } x \in X. \end{aligned}$$

Since this is true for all $x \in X$, we have $(T_2T_1)^*h = (T_1^*T_2^*)h$, and from this we conclude $(T_2T_1)^* = T_1^*T_2^*$.

Now we consider Hilbert spaces.

Let $T \in B[H]$. Then the adjoint map $T^* : H^* \to H^*$ is given by

$$(T^*g)x = g(Tx) \text{ for all } g \in H^* \text{ and for all } x \in H.$$

If $b \in H$ is the representative of $g \in H^*$ (cf. Theorem 3.7), that is

$$g(x) =< x, b > \text{ for all } x \in H,$$

then we define

$$T^*b \text{ to be } T^*g,$$

and we have

$$< x, T^*b >=< Tx, b > \text{ for all } x \in H.$$

This relation defines $T^* \in B[H]$, moreover, it defines T^* uniquely, for

$$< x, T^{\#}b >=< Tx, b > \text{ for all } x, b \in H \text{ implies}$$

$$< x, T^{\#}b >=< x, T^*b > \text{ for all } x, b \in H,$$

$$\text{hence } T^{\#}b = T^*b \text{ for all } b \in H, \text{ and so } T^{\#} = T^*.$$

Let X and Y be seminormed spaces. The map $* : B(X,Y) \to B(Y^*, X^*)$ is linear. For if $T_1, T_2, T \in B(X,Y)$ and $\lambda \in \mathbb{C}$ then, for all $g \in Y^*$ and for all $x \in X$

$$\begin{aligned} ((T_1 + T_2)^* g)x &= g((T_1 + T_2)x) = g(T_1 x + T_2 x) \\ &= g(T_1 x) + g(T_2 x) = (T_1^* g)x + (T_2^* g)x = (T_1^* g + T_2^* g)x, \end{aligned}$$

hence

$$(T_1 + T_2)^* g = T_1^* g + T_2^* g = (T_1^* + T_2^*)g \text{ for all } g \in Y^*,$$

and so

$$(T_1 + T_2)^* = T_1^* + T_2^*,$$

and similarly

$$(\lambda T)^* = \lambda T^*.$$

Now we consider bounded operators on Hilbert spaces.

Since the identification from H^* to H is conjugate linear, the map $*$ is conjugate linear, that is for bounded operators on H, we have

$$(T_1 + \lambda T_2)^* = T_1^* + \overline{\lambda} T_2^*.$$

We also have

$$T^{**} = T \text{ for all } T \in B[H],$$

since, for all $a, b \in H$,

$$< a, T^{**} b > = < T^* a, b > = \overline{< b, T^* a >} = \overline{< Tb, a >} = < a, Tb >$$

and the adjoint is unique, this implies that the adjoint of T^* is T. Because of the property that $T^{**} = T$, the map $* : B[H] \to B[H]$ is called *involution*. Furthermore, for all $a \in H$ and all $g \in H^*$,

$$|(T^* g)a| = |g(Ta)| \leq \|g\| \, \|Ta\| \leq \|g\| \, \|T\| \, \|a\|$$

implies

$$\|T^* g\| \leq \|T\| \, \|g\| \text{ for all } g \in H^*,$$

and so

$$\|T^*\| \leq \|T\|.$$

Replacing T by T^* and using $T^{**} = T$, we also have

$$\|T\|\|T^{**}\| \leq \|T^*\|.$$

Thus we have shown

$$\|T^*\| = \|T\|. \tag{32}$$

(We already know this from Theorem 6.2 (a), but now, for Hilbert spaces, the proof is much simpler.)

Theorem 6.3 *Let H be a Hilbert space and $T \in B[H]$. Then*

$$\|T^*T\| = \|T\|^2.$$

Proof. By the multiplicative property and (32)

$$\|T^*T\| \leq \|T^*\| \, \|T\| = \|T\|^2.$$

On the other hand, for any $a \in H$,

$$\|Ta\|^2 =< Ta, Ta >=< a, T^*Ta >\leq \|a\| \, \|T^*Ta\| \leq \|a\|^2 \, \|T^*T\|,$$

and this implies

$$\|Ta\| \leq \sqrt{\|T^*T\|} \, \|a\|, \text{ hence } \|T\| \leq \sqrt{\|T^*T\|}.$$

Definition 6.4 *An operator $T \in B[H]$ is said to have an* **inverse T^{-1} in $B[H]$** *if there exists $T^{-1} \in B[H]$ such that*

$$T^{-1}T = TT^{-1} = I, \text{ the identity operator.}$$

A bounded operator T is said to be **self–adjoint** *or* **Hermitian** *if $T^* = T$.*

Theorem 6.5 *For every $T \in B[H]$, the operator $I + T^*T$ has an inverse in $B[H]$.*

Proof. By Theorem 6.2 (c), it suffices to show that $I + T^*T$ is norm increasing, since this implies that it is one–to–one. Let $a \in H$ be given. Then

$$\begin{aligned} \|(I + T^*T)a\| \, \|a\| &\geq < (I + T^*T)a, a > = < a, a > + < T^*Ta, a > \\ &= \|a\|^2 + < Ta, Ta > \geq \|a\|^2. \end{aligned}$$

This implies $\|(I + T^*T)a\| \geq \|a\|$.

Definition 6.6 *The* **spectrum $\boldsymbol{\sigma(T)}$ of a bounded operator $\boldsymbol{T}$** *is the set of all scalars λ such that the operator $T - \lambda I$ has no inverse. A scalar λ is called a* **characteristic value of $\boldsymbol{T}$** *if there exists a vector $a \neq 0$ such that $Ta = \lambda a$; then a is called* **characteristic vector corresponding to $\boldsymbol{\lambda}$**.

Each characteristic value lies in the spectrum, since $(\lambda I - T)a = 0$, so $\lambda I - T$ is not one–to–one.

Example **6.7** *We define $T \in B[\ell_2]$ by*

$$Ta = (0, a_1, a_2, \ldots) \text{ for all } a = (a_k)_{k=1}^{\infty} \in \ell_2.$$

Then T is not onto, since $e^{(1)} \notin T(\ell_2)$. Thus T has no inverse, and so $0 \in \sigma(T)$. But 0 is not a characteristic value, since T is one–to–one.

It turns out that the characteristic values of self–adjoint operators are always reals, and that characteristic vectors of different characteristic values of a self–adjoint operator are orthogonal.

Theorem 6.8 *Let V be an inner product space, $S \subset V$ and $T : S \to V$ a function satisfying*

$$< Ta, b >=< a, Tb > \text{ for all } a, b \in S.$$

Then the characteristic values of T are reals and two characteristic vectors in S corresponding to two different characteristic values of T are orthogonal.

Proof. If λ is a characteristic value, and a a characteristic vector of T, then

$$\lambda \|a\|^2 =< \lambda a, a >=< Ta, a >=< a, Ta >=< a, \lambda a >= \overline{\lambda} \|a\|^2,$$

and so $\lambda \in \mathbb{R}$, since $a \neq 0$.
Now let $Ta = \lambda a$, $Tb = \mu b$, $\lambda \neq \mu$ and $a, b \neq 0$. Then $\lambda, \mu \in \mathbb{R}$ and

$$\begin{aligned} \lambda < a, b > &=< \lambda a, b >=< Ta, b >=< a, Tb >=< a, \mu b > \\ &= \mu < a, b > \end{aligned}$$

implies $< a, b >= 0$.

Example **6.9** *Let $V = C[0, 1]$. We fix $f, g, h \in V$ with $h(x) > 0$ on $[0, 1]$ and define*

$$< a, b >= \int_0^1 a(t)b(t)h(t)\, dt \text{ for all } a, b \in V.$$

Let the subset S of V consist of those $a \in V$ such that

$$D(f \cdot a') \in V;$$

(here D means the derivative and a' is the derivative of a) and such that

$$a(0)f(0) = a(1)f(1) = 0.$$

Finally let

$$Ta = (D(f \cdot a') + g \cdot a)/h.$$

Then $< Ta, b >=< a, Tb >$ *for all* $a, b \in S$, *for*

$$\begin{aligned}
< a, Tb > &= \int_0^1 a(t)Tb(t)h(t)\,dt \\
&= \int_0^1 a(t)\frac{d}{dt}(f(t)b'(t))\,dt + \int_0^1 a(t)g(t)b(t)\,dt \\
&= a(t)f(t)b'(t)|_0^1 - \int_0^1 a'(t)f(t)b'(t)\,dt + \int_0^1 a(t)g(t)b(t)\,dt \\
&= \int_0^1 (a(t)g(t)b(t) - a'(t)f(t)b'(t))\,dt
\end{aligned}$$

and the result follows, since this is symmetric in a *and* b. *We now apply Theorem 6.8 to the differential equation*

$$D[f(x)y'] + (g(x) - \lambda h(x))y = 0$$

with boundary conditions

$$f(0)y(0) = f(1)y(1) = 0.$$

The result is that if a *and* b *are solutions corresponding to distinct values of* λ *then*

$$\int_0^1 a(x)b(x)h(x)\,dx = 0.$$

As a special case, consider $y" - \lambda y = 0$ *with* $y(0) = y(1) = 0$. *If* $\lambda \neq -n^2\pi^2$ $(n \in I\!N)$ *then the only solution is* $y = 0$. *For* $\lambda = n^2\pi^2$, *we have* $y = \sin nx$. *The conclusion is*

$$\int_0^1 \sin mx \sin nx\,dx = 0 \text{ if } m \neq n.$$

Mathematisches Institut, Universität Giessen, Arndtsrasse 2, D–35392 Giessen, Germany

Department of Mathematics, Faculty of Science and Mathematics, University of Niš, Višegradska 33, 18000 Niš, Yugoslavia

E-mail address: `eberhard.malkowsky@math.uni-giessen.de`
`ema@BankerInter.net`