4

Elementary Hilbert Space Theory

Inner Products and Linear Functionals

4.1 Definition A complex vector space H is called an *inner product space* (or *unitary space*) if to each ordered pair of vectors x and $y \, \varepsilon \, H$ there is associated a complex number (x,y), the so-called "inner product" (or "scalar product") of x and y, such that the following rules hold:

(a) $(y,x) = \overline{(x,y)}$. (The bar denotes complex conjugation.)

(b) $(x + y, z) = (x,z) + (y,z)$ if x, y, and $z \, \varepsilon \, H$.

(c) $(\alpha x, y) = \alpha(x,y)$ if x and $y \, \varepsilon \, H$ and α is a scalar.

(d) $(x,x) \geq 0$ for all $x \, \varepsilon \, H$.

(e) $(x,x) = 0$ only if $x = 0$.

Let us list some immediate consequences of these axioms:

(c) implies that $(0,y) = 0$ for all $y \, \varepsilon \, H$.

(b) and (c) may be combined into the statement: *For every $y \, \varepsilon \, H$, the mapping $x \to (x,y)$ is a linear functional on H.*

(a) and (c) show that $(x, \alpha y) = \bar{\alpha}(x,y)$.

(a) and (b) imply the second distributive law:

$$(z, x + y) = (z,x) + (z,y).$$

By (d), we may define $\|x\|$, the *norm* of the vector $x \, \varepsilon \, H$, to be the nonnegative square root of (x,x). Thus

(f) $\|x\|^2 = (x,x)$.

4.2 The Schwarz Inequality *The properties 4.1(a) to (d) imply that*

$$|(x,y)| \leq \|x\| \, \|y\|$$

for all x and $y \, \varepsilon \, H$.

75

PROOF   Put $A = \|x\|^2$, $B = |(x,y)|$, and $C = \|y\|^2$. There is a complex number α such that $|\alpha| = 1$ and $\alpha(y,x) = B$. For any real r, we then have

(1) $(x - r\alpha y, x - r\alpha y) = (x,x) - r\alpha(y,x) - r\bar{\alpha}(x,y) + r^2(y,y)$.

The expression on the left is real and not negative. Hence

(2) $A - 2Br + Cr^2 \geq 0$

for every real r. If $C = 0$, we must have $B = 0$, otherwise (2) is false for large positive r. If $C > 0$, take $r = B/C$ in (2), and obtain $B^2 \leq AC$.

4.3   The Triangle Inequality   For x and $y \, \varepsilon \, H$, we have

$$\|x + y\| \leq \|x\| + \|y\|.$$

PROOF   By the Schwarz inequality,

$$\|x + y\|^2 = (x + y, x + y) = (x,x) + (x,y) + (y,x) + (y,y)$$
$$\leq \|x\|^2 + 2\|x\| \, \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

4.4   Definition   It follows from the triangle inequality that

(1) $\|x - z\| \leq \|x - y\| + \|y - z\|$ $(x, y, z \, \varepsilon \, H)$.

If we define the distance between x and y to be $\|x - y\|$, all the axioms for a metric space are satisfied; here, for the first time, we use part (e) of Definition 4.1.

Thus H is now a metric space. If this metric space is *complete*, i.e., if every Cauchy sequence converges in H, then H is called a *Hilbert space*.

Throughout the rest of this chapter, the letter H will denote a Hilbert space.

4.5   Examples

(a)   For any fixed n, the set C^n of all n-tuples

$$x = (\xi_1, \ldots, \xi_n),$$

where $\xi_1, \ldots, \xi_n$ are complex numbers, is a Hilbert space if addition and scalar multiplication are defined componentwise, as usual, and if

$$(x,y) = \sum_{j=1}^{n} \xi_j \bar{\eta}_j \qquad (y = (\eta_1, \ldots, \eta_n)).$$

(b)   If μ is any positive measure, $L^2(\mu)$ is a Hilbert space, with inner product

$$(f,g) = \int_X f\bar{g} \, d\mu.$$

The integrand on the right is in $L^1(\mu)$, by Theorem 3.8, so that (f,g) is well defined. Note that

$$\|f\| = (f,f)^{\frac{1}{2}} = \left\{ \int_X |f|^2 \, d\mu \right\}^{\frac{1}{2}} = \|f\|_2.$$

The completeness of $L^2(\mu)$ (Theorem 3.11) shows that $L^2(\mu)$ is indeed a Hilbert space. [We recall that $L^2(\mu)$ should be regarded as a space of *equivalence classes* of functions; compare the discussion in Sec. 3.10.]

For $H = L^2(\mu)$, the inequalities 4.2 and 4.3 turn out to be special cases of the inequalities of Hölder and Minkowski.

Note that Example (*a*) is a special case of (*b*). What is the measure in (*a*)?

(*c*) The vector space of all continuous complex functions on [0,1] is an inner product space if

$$(f,g) = \int_0^1 f(t)\overline{g(t)} \, dt$$

but is not a Hilbert space.

4.6 Theorem *For any fixed $y \in H$, the mappings*

$$x \to (x,y), \qquad x \to (y,x), \qquad x \to \|x\|$$

are continuous functions on H.

PROOF The Schwarz inequality implies that

$$|(x_1,y) - (x_2,y)| = |(x_1 - x_2, y)| \le \|x_1 - x_2\| \, \|y\|,$$

which proves that $x \to (x,y)$ is, in fact, uniformly continuous, and the same is true for $x \to (y,x)$. The triangle inequality $\|x_1\| \le \|x_1 - x_2\| + \|x_2\|$ yields

$$\|x_1\| - \|x_2\| \le \|x_1 - x_2\|,$$

and if we interchange x_1 and x_2 we see that

$$|\,\|x_1\| - \|x_2\|\,| \le \|x_1 - x_2\|$$

for all x_1 and $x_2 \in H$. Thus $x \to \|x\|$ is also uniformly continuous.

4.7 Subspaces A subset M of a vector space V is called a *subspace* of V if M is itself a vector space, relative to the addition and scalar multiplication which are defined in V. A necessary and sufficient condition for a set $M \subset V$ to be a subspace is that $x + y \in M$ and $\alpha x \in M$ whenever x and $y \in M$ and α is a scalar.

In the vector space context, the word "subspace" will always have this meaning. Sometimes, for emphasis, we may use the term "linear subspace" in place of subspace.

For example, if V is the vector space of all complex functions on a set S, the set of all bounded complex functions on S is a subspace of V, but the set of all $f \, \varepsilon \, V$ with $|f(x)| \leq 1$ for all $x \, \varepsilon \, S$ is not. The real vector space R^3 has the following subspaces, and no others: (a) R^3, (b) all planes through the origin 0, (c) all straight lines through 0, and (d) $\{0\}$.

A *closed subspace* M of H is a subspace which is a closed set relative to the topology induced by the metric of H.

4.8 Convex Sets A set E in a vector space V is said to be *convex* if it has the following geometric property: Whenever $x \, \varepsilon \, E$, $y \, \varepsilon \, E$, and $0 < t < 1$, the point

$$z_t = (1 - t)x + ty$$

also lies in E. As t runs from 0 to 1, one may visualize z_t as describing a straight line segment in V, from x to y. Convexity requires that E contain the segments between any two of its points.

It is clear that every subspace of V is convex.

Also, if E is convex, so is each of its translates

$$E + x = \{y + x : y \, \varepsilon \, E\}.$$

4.9 Orthogonality If $(x,y) = 0$ for some x and $y \, \varepsilon \, H$, we say that x is orthogonal to y, and sometimes write $x \perp y$. Since $(x,y) = 0$ implies $(y,x) = 0$, the relation $\perp$ is symmetric.

Let $x^\perp$ denote the set of all $y \, \varepsilon \, H$ which are orthogonal to x; and if M is a subspace of H, let $M^\perp$ be the set of all $y \, \varepsilon \, H$ which are orthogonal to every $x \, \varepsilon \, M$.

Note that $x^\perp$ is a subspace of H, since $x \perp y$ and $x \perp y'$ implies $x \perp (y + y')$ and $x \perp \alpha y$. Also, $x^\perp$ is precisely the set of points where the continuous function $y \to (x,y)$ is 0. Hence $x^\perp$ is a *closed* subspace of H. Since

$$M^\perp = \bigcap_{x \varepsilon M} x^\perp,$$

$M^\perp$ is an intersection of closed subspaces, and it follows that $M^\perp$ *is a closed subspace of H.*

4.10 Theorem *Every nonempty, closed, convex set E in a Hilbert space H contains a unique element of smallest norm.*

In other words, there is one and only one $x_0 \, \varepsilon \, E$ such that $\|x_0\| \leq \|x\|$ for every $x \, \varepsilon \, E$.

PROOF An easy computation, using only the properties listed in Definition 4.1, establishes the identity

(1) $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$ $(x \text{ and } y \, \varepsilon \, H)$.

This is known as the *parallelogram law:* If we interpret $\|x\|$ to be the length of the vector x, (1) says that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of its sides, a familiar proposition in plane geometry.

Let $\delta = \inf \{\|x\|: x \, \varepsilon \, E\}$. For any x and $y \, \varepsilon \, E$, we apply (1) to $\frac{1}{2}x$ and $\frac{1}{2}y$ and obtain

$$(2) \qquad \tfrac{1}{4}\|x - y\|^2 = \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|y\|^2 - \left\|\frac{x + y}{2}\right\|^2.$$

Since E is convex, $(x + y)/2 \, \varepsilon \, E$. Hence

$$(3) \qquad \|x - y\|^2 \leq 2\|x\|^2 + 2\|y\|^2 - 4\delta^2 \qquad (x \text{ and } y \, \varepsilon \, E).$$

If also $\|x\| = \|y\| = \delta$, then (3) implies $x = y$, and we have proved the uniqueness assertion of the theorem.

The definition of δ shows that there is a sequence $\{y_n\}$ in E so that $\|y_n\| \to \delta$ as $n \to \infty$. Replace x and y in (3) by y_n and y_m. Then, as $n \to \infty$ and $m \to \infty$, the right side of (3) will tend to 0. This shows that $\{y_n\}$ is a Cauchy sequence. Since H is complete, there exists an $x_0 \, \varepsilon \, H$ so that $y_n \to x_0$, i.e., $\|y_n - x_0\| \to 0$, as $n \to \infty$. Since $y_n \, \varepsilon \, E$ and E is closed, $x_0 \, \varepsilon \, E$. Since the norm is a continuous function on H (Theorem 4.6), it follows that

$$\|x_0\| = \lim_{n \to \infty} \|y_n\| = \delta.$$

4.11 Theorem *Let M be a closed subspace of H. There exists a unique pair of mappings P and Q such that P maps H into M, Q maps H into $M^{\perp}$, and*

$$(1) \qquad x = Px + Qx$$

for all $x \, \varepsilon \, H$. These mappings have the following further properties:

(2) *If $x \, \varepsilon \, M$, then $Px = x$, $Qx = 0$; if $x \, \varepsilon \, M^{\perp}$, then $Px = 0$, $Qx = x$.*

$$(3) \qquad \|x - Px\| = \inf \{\|x - y\|: y \, \varepsilon \, M\} \text{ if } x \, \varepsilon \, H.$$

$$(4) \qquad \|x\|^2 = \|Px\|^2 + \|Qx\|^2.$$

(5) *P and Q are linear mappings.*

Corollary *If $M \neq H$, there exists a $y \, \varepsilon \, H$, $y \neq 0$, such that $y \perp M$.*

P and Q are called the orthogonal projections of H onto M and $M^{\perp}$.

PROOF   For any $x \, \varepsilon \, H$, the set $x + M = \{x + y: y \, \varepsilon \, M\}$ is closed and convex. Define Qx to be the unique element of smallest norm in $x + M$; this exists, by Theorem 4.10. Define $Px = x - Qx$. Then

(1) holds.   Since $Qx \, \varepsilon \, x + M$, it is clear that $Px \, \varepsilon \, M$.   Thus P maps H into M.

We next have to show that $(Qx,y) = 0$ for all $y \, \varepsilon \, M$.   Assume $\|y\| = 1$, without loss of generality, and put $z = Qx$.   The minimal property of Qx shows that

$$(z,z) = \|z\|^2 \leq \|z - \alpha y\|^2 = (z - \alpha y, \, z - \alpha y)$$

for every scalar α.   This simplifies to

$$0 \leq -\alpha(y,z) - \bar{\alpha}(z,y) + |\alpha|^2.$$

With $\alpha = (z,y)$, this gives $0 \leq -|(z,y)|^2$, so that $(z,y) = 0$.   Thus Q maps H into $M^\perp$.

Now if $x = x_0 + x_1$, with $x_0 \, \varepsilon \, M$, $x_1 \, \varepsilon \, M^\perp$, then

$$x_0 - Px = Qx - x_1.$$

Since $x_0 - Px \, \varepsilon \, M$, $Qx - x_1 \, \varepsilon \, M^\perp$, and $M \cap M^\perp = \{0\}$ [an immediate consequence of the fact that $(x,x) = 0$ implies $x = 0$], we have $x_0 = Px$, $x_1 = Qx$, which proves the uniqueness assertion.

The linearity of P and Q is proved similarly: applying (1) to x, to y, and to $\alpha x + \beta y$, we obtain

$$P(\alpha x + \beta y) - \alpha Px - \beta Py = \alpha Qx + \beta Qy - Q(\alpha x + \beta y).$$

The left side is in M, the right side in $M^\perp$; hence both are 0, so P and Q are linear.

Property (2) follows from (1); (3) was used to define P; and (4) follows from (1), since $(Px,Qx) = 0$.   To prove the corollary, take $x \, \varepsilon \, H$, $x \, \notin \, M$, and put $y = Qx$; since $x \neq Px$, $y \neq 0$.

We have already observed that $x \to (x,y)$ is, for each $y \, \varepsilon \, H$, a continuous linear functional on H.   It is a very important fact that *all* continuous linear functionals on H are of this type.

4.12 Theorem   *If L is a continuous linear functional on H, then there is a unique $y \, \varepsilon \, H$ such that*

(1) $Lx = (x,y)$ $(x \, \varepsilon \, H)$.

PROOF   If $Lx = 0$ for all x, take $y = 0$.   Otherwise, define

(2) $M = \{x : Lx = 0\}$.

The linearity of L shows that M is a subspace.   The continuity of L shows that M is closed.   Since $Lx \neq 0$ for some $x \, \varepsilon \, H$, Theorem 4.11 shows that $M^\perp$ does not consist of 0 alone.

It is clear that we must look for our desired y in $M^\perp$, and that we must have $Ly = (y,y)$.

Choose $z \, \varepsilon \, M^\perp$, $z \neq 0$. Then $z \notin M$, hence $Lz \neq 0$. Put $y = \alpha z$, where $\bar{\alpha} = (Lz)/(z,z)$. Then $y \, \varepsilon \, M^\perp$, $Ly = (y,y)$, and $y \neq 0$. For any $x \, \varepsilon \, H$, put

$$(3) \qquad x' = x - \frac{Lx}{(y,y)} y \qquad \text{and} \qquad x'' = \frac{Lx}{(y,y)} y.$$

Then $Lx' = 0$, hence $x' \, \varepsilon \, M$, hence $(x',y) = 0$, hence

$$(4) \qquad (x,y) = (x'',y) = Lx,$$

which gives the desired representation of Lx.

The uniqueness of y is easily proved, for if $(x,y) = (x,y')$ for all $x \, \varepsilon \, H$, set $z = y - y'$; then $(x,z) = 0$ for all $x \, \varepsilon \, H$; in particular, $(z,z) = 0$, hence $z = 0$.

Orthonormal Sets

4.13 Definitions If V is a vector space, if $x_1, \ldots, x_k \, \varepsilon \, V$, and if $c_1, \ldots, c_k$ are scalars, then $c_1 x_1 + \cdots + c_k x_k$ is called a *linear combination* of $x_1, \ldots, x_k$. The set $\{x_1, \ldots, x_k\}$ is called *independent* if $c_1 x_1 + \cdots + c_k x_k = 0$ implies that $c_1 = \cdots = c_k = 0$. A set $S \subset V$ is independent if every finite subset of S is independent. The set $[S]$ of all linear combinations of all finite subsets of S (also called the set of all *finite linear combinations* of members of S) is clearly a vector space; $[S]$ is the smallest subspace of V which contains S; $[S]$ is called the *span* of S, or the space spanned by S.

A set of vectors u_α in a Hilbert space H, where α runs through some index set A, is called *orthonormal* if it satisfies the orthogonality relations $(u_\alpha, u_\beta) = 0$ for all $\alpha \neq \beta$, $\alpha \, \varepsilon \, A$, and $\beta \, \varepsilon \, A$, and if it is normalized so that $\|u_\alpha\| = 1$ for each $\alpha \, \varepsilon \, A$. In other words, $\{u_\alpha\}$ is orthonormal provided that

$$(1) \qquad (u_\alpha, u_\beta) = \begin{cases} 1 & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

If $\{u_\alpha : \alpha \, \varepsilon \, A\}$ is orthonormal, we associate with each $x \, \varepsilon \, H$ a complex function $\hat{x}$ on the index set A, defined by

$$(2) \qquad \hat{x}(\alpha) = (x, u_\alpha) \qquad (\alpha \, \varepsilon \, A).$$

One sometimes calls the numbers $\hat{x}(\alpha)$ the *Fourier coefficients* of x, relative to the set $\{u_\alpha\}$.

4.14 Theorem *If $u_1, \ldots, u_k$ is an orthonormal set, and if $x = \sum\limits_1^k c_n u_n$,*
then

(1) $c_n = (x, u_n), \quad \text{for } 1 \leq n \leq k,$

(2) $\|x\|^2 = \sum\limits_1^k |c_n|^2.$

PROOF   Apply the relations 4.13(1).

Corollary   *Every orthonormal set is independent.*

PROOF   This follows from (2).

4.15 An Approximation Problem   Let $v_1, \ldots, v_k$ be a set of independent vectors in H, and suppose $x \, \varepsilon \, H$. The problem is to *find a method of computing the minimum value of*

(1) $\left\| x - \sum\limits_{j=1}^k c_j v_j \right\|,$

where $c_1, \ldots, c_k$ range over all scalars, and to find the corresponding values of $c_1, \ldots, c_k$.

Let M be the span of $v_1, \ldots, v_k$. If we knew that M is closed, we could apply Theorem 4.11 and deduce the existence of a unique minimizing element $x_0 := Px$, where

(2) $x_0 = \sum\limits_{j=1}^k \tilde{c}_j v_j,$

which also has the property that $x - x_0 \, \varepsilon \, M^\perp$. These facts could then be used to obtain information about the coefficients $\tilde{c}_1, \ldots, \tilde{c}_k$ in (2).

Since M is the span of a finite set of vectors, it may seem obvious that M is closed.   One may prove it by induction, observing that $\{0\}$ is certainly closed and proceeding with the aid of the following lemma:

If V is a closed subspace of H, if $y \, \varepsilon \, H, y \notin V$, and V^* is the space spanned by V and y, then V^* is closed.

To prove this, suppose z is a limit point of V^*.   Then

$$z = \lim_{n \to \infty} (x_n + \lambda_n y),$$

where $x_n \, \varepsilon \, V$, and λ_n are scalars.   Since convergent sequences in metric spaces are bounded, there exists an $\eta < \infty$ such that $\|x_n + \lambda_n y\| < \eta$ for $n = 1, 2, 3, \ldots$.   If it were true that $|\lambda_n| \to \infty$, we should have

$$\|\lambda_n^{-1} x_n + y\| < \frac{\eta}{|\lambda_n|} \to 0$$

so that $-y \, \varepsilon \, V$, since V is closed. But $y \notin V$. Hence $\{\lambda_n\}$ has a Cauchy subsequence $\{\lambda_{n_i}\}$ converging to some λ, and so $\{x_{n_i}\}$, being a difference of two Cauchy sequences, is itself a Cauchy sequence in H and converges to some $x \, \varepsilon \, V$. Then $z = x + \lambda y$. This proves that V^* contains all its limit points.

We now return to our problem. Put

$$(3) \qquad\qquad a_{ij} = (v_j, v_i), \qquad b_i = (x, v_i).$$

Then if x_0, given by (2), is the minimizing element, we must have

$$(x - x_0, \, v_i) = 0$$

for $i = 1, \ldots, k$, which leads to a set of k linear equations in the unknowns $c_1, \ldots, c_k$:

$$(4) \qquad\qquad \sum_{j=1}^{k} a_{ij} c_j = b_i \qquad (1 \leq i \leq k).$$

We know from Theorem 4.11 that x_0 exists and is unique. Hence the determinant of the a_{ij} is not 0, and the c_j can be computed from (4).

Next, let δ be the minimum value of (1). Since $(x - x_0, \, v_i) = 0$, we have $(x - x_0, \, x_0) = 0$; hence

$$\delta^2 = (x - x_0, \, x - x_0) = (x, \, x - x_0) = \left(x, \, x - \sum_{j=1}^{k} c_j v_j \right),$$

so that

$$(5) \qquad\qquad \delta^2 = \|x\|^2 - \sum_{j=1}^{k} \bar{c}_j b_j.$$

This solves our problem, in terms of the quantities (3).

Let us now turn to a special case: Replace $v_1, \ldots, v_k$ by an orthonormal set $u_1, \ldots, u_k$. Then $a_{ij} = 1$ if $i = j$, $a_{ij} = 0$ if $i \neq j$, hence (4) gives $c_i = b_i$, and (5) becomes

$$(6) \qquad\qquad \delta^2 = \|x\|^2 - \sum_{j=1}^{k} |b_j|^2.$$

We may summarize as follows:

4.16 Theorem *Let $u_1, \ldots, u_k$ be an orthonormal set in H, and let $x \, \varepsilon \, H$. Then*

$$(1) \qquad\qquad \left\| x - \sum_{j=1}^{k} (x, u_j) u_j \right\| \leq \left\| x - \sum_{j=1}^{k} \lambda_j u_j \right\|$$

for all scalars $\lambda_1, \ldots, \lambda_k$. *Equality holds in* (1) *if and only if* $\lambda_j = (x,u_j)$ *for* $1 \leq j \leq k$. *The vector*

$$(2) \qquad \sum_{j=1}^{k} (x,u_j)u_j$$

is the orthogonal projection of x *into the subspace* $[u_1, \ldots, u_k]$, *and if* δ *is the distance from* x *to this subspace, then*

$$(3) \qquad \sum_{j=1}^{k} |(x,u_j)|^2 = \|x\|^2 - \delta^2.$$

Corollary (Bessel's Inequality) *If* $\{u_\alpha : \alpha \varepsilon A\}$ *is any orthonormal set in* H, *if* $x \varepsilon H$, *and if* $\hat{x}(\alpha) = (x,u_\alpha)$, *then*

$$(4) \qquad \sum_{\alpha\varepsilon A} |\hat{x}(\alpha)|^2 \leq \|x\|^2.$$

This corollary calls for some explanation and comment. The set A is any index set, possibly even uncountable, and not ordered in any way. Under those conditions, what does the sum on the left side of (4) signify? We define it as follows: If $0 \leq \varphi(\alpha) \leq \infty$ for each $\alpha \varepsilon A$, the symbol

$$(5) \qquad \sum_{\alpha\varepsilon A} \varphi(\alpha)$$

denotes the supremum of the set of all finite sums $\varphi(\alpha_1) + \varphi(\alpha_2) + \cdots + \varphi(\alpha_k)$, where $\alpha_1, \ldots, \alpha_k$ are distinct members of A. With this agreement, it is clear that (4) follows from (3).

A moment's consideration will show that *the sum* (5) *is precisely the Lebesgue integral of* φ *relative to the counting measure on* A. Let $\ell^2(A)$ be the L^2-space relative to this counting measure. Then (4) asserts that $\hat{x} \varepsilon \ell^2(A)$ and that $\|\hat{x}\|_2 \leq \|x\|$.

One immediate consequence of (4) should be mentioned explicitly:

For any $x \varepsilon H$ *and any orthonormal set* $\{u_\alpha\}$ *in* H, *the set of all* α *such that* $\hat{x}(\alpha) \neq 0$ *is at most countable.*

Let F be the mapping which assigns to each $x \varepsilon H$ the function $\hat{x}$ on A. For each $\alpha \varepsilon A$, $x \to (x,u_\alpha)$ is a linear functional. Hence F is a linear transformation of H into $\ell^2(A)$ (see Definition 2.1). Also, F does not increase distances, since $\|\hat{x} - \hat{y}\|_2 \leq \|x - y\|$. In particular, F is continuous.

We shall now see that the completeness of H implies that F maps H onto $\ell^2(A)$ and that under certain conditions on $\{u_\alpha\}$, F is actually an isometry, i.e., that $\|\hat{x}\|_2 = \|x\|$ for all $x \varepsilon H$. Then, of course, F will be one-to-one.

4.17 The Riesz-Fischer Theorem *Let $\{u_\alpha: \alpha \,\varepsilon\, A\}$ be an orthonormal set in H. Suppose $\varphi \,\varepsilon\, \ell^2(A)$. Then $\varphi = \hat{x}$ for some $x \,\varepsilon\, H$.*

PROOF For $n = 1, 2, 3, \ldots$, let $A_n = \{\alpha: |\varphi(\alpha)| > 1/n\}$. Each A_n is a finite set. (In fact, one checks easily that A_n has at most $n^2\|\varphi\|_2^2$ elements.) Put

$$(1) \qquad x_n = \sum_{\alpha \varepsilon A_n} \varphi(\alpha)u_\alpha \qquad (n = 1, 2, 3, \ldots).$$

Then $\hat{x}_n = \varphi \cdot \chi_{A_n}$, so that $\hat{x}_n(\alpha) \to \varphi(\alpha)$ for every $\alpha \,\varepsilon\, A$, and $|\varphi - \hat{x}_n|^2 \leq |\varphi|^2$. Hence, by an elementary case of the dominated convergence theorem, $\|\varphi - \hat{x}_n\|_2 \to 0$. It follows that $\{\hat{x}_n\}$ is a Cauchy sequence in $\ell^2(A)$. Since the sets A_n are finite, Theorem 4.14 shows that $\|x_n - x_m\| = \|\hat{x}_n - \hat{x}_m\|_2$. Thus $\{x_n\}$ is a Cauchy sequence in H, and since H is complete, there exists an $x = \lim_{n \to \infty} x_n$ in H. For any $\alpha \,\varepsilon\, A$ we then have

$$\hat{x}(\alpha) = (x, u_\alpha) = \lim_{n \to \infty} (x_n, u_\alpha) = \lim_{n \to \infty} \hat{x}_n(\alpha) = \varphi(\alpha),$$

which completes the proof.

4.18 Theorem *Let $\{u_\alpha: \alpha \,\varepsilon\, A\}$ be an orthonormal set in H. Each of the following four conditions on $\{u_\alpha\}$ implies the other three:*

(a) *$\{u_\alpha\}$ is a maximal orthonormal set in H.*

(b) *The set S of all finite linear combinations of members of $\{u_\alpha\}$ is dense in H.*

(c) *For every $x \,\varepsilon\, H$, we have $\|x\|^2 = \sum_{\alpha \varepsilon A} |\hat{x}(\alpha)|^2$.*

(d) *If $x \,\varepsilon\, H$ and $y \,\varepsilon\, H$, then $(x,y) = \sum_{\alpha \varepsilon A} \hat{x}(\alpha)\overline{\hat{y}(\alpha)}$.*

This last formula is known as *Parseval's identity*. Observe that $\hat{x} \,\varepsilon\, \ell^2(A)$ and $\hat{y} \,\varepsilon\, \ell^2(A)$, hence $\hat{x}\hat{y} \,\varepsilon\, \ell^1(A)$, so that the summation in (d) is well defined. Of course, (c) is the special case $x = y$ of (d). Maximal orthonormal sets are frequently called *complete orthonormal sets* or *orthonormal bases*.

PROOF To say that $\{u_\alpha\}$ is maximal means simply that no vector of H can be adjoined to $\{u_\alpha\}$ in such a way that the resulting set is still orthonormal. This happens precisely when there is no $x \neq 0$ in H which is orthogonal to every u_α.

We shall prove that $(a) \to (b) \to (c) \to (d) \to (a)$.

Let M be the closure of S. Since S is a subspace, so is M ($x_n \to x$ and $y_n \to y$ implies $x_n + y_n \to x + y$, $\lambda x_n \to \lambda x$); and if S is not dense in H, then $M \neq H$, so that $M^\perp$ contains a nonzero vector, by Theorem 4.11. Thus $\{u_\alpha\}$ is not maximal if S is not dense, and (a) implies (b).

Suppose (b) holds. Fix $x \in H$, $\epsilon > 0$. Since S is dense, there is a finite set $u_{\alpha_1}, \ldots, u_{\alpha_n}$ such that some linear combination of these vectors has distance less than ϵ from x. By Theorem 4.16, this approximation can only be improved if we take $\hat{x}(\alpha_j)$ for the coefficient of u_{α_j}. Thus if

$$(1) \qquad z = \hat{x}(\alpha_1)u_{\alpha_1} + \cdots + \hat{x}(\alpha_k)u_{\alpha_k},$$

we have $\|x - z\| < \epsilon$, hence $\|x\| < \|z\| + \epsilon$, and Theorem 4.14 gives

$$(2) \qquad (\|x\| - \epsilon)^2 < \|z\|^2 = |\hat{x}(\alpha_1)|^2 + \cdots + |\hat{x}(\alpha_k)|^2 \leq \sum_{\alpha \in A} |\hat{x}(\alpha)|^2.$$

Since ϵ was arbitrary, (c) follows from (2) and the Bessel inequality.

The equation in (c) can also be written in the form

$$(3) \qquad (x,x) = (\hat{x},\hat{x}),$$

the inner product on the right being the one in the Hilbert space $\ell^2(A)$, as in Example 4.5(b). Fix $x \in H$, $y \in H$. If (c) holds, then

$$(4) \qquad (x + \lambda y, x + \lambda y) = (\hat{x} + \lambda\hat{y}, \hat{x} + \lambda\hat{y})$$

for every scalar λ; hence

$$(5) \qquad \bar{\lambda}(x,y) + \lambda(y,x) = \bar{\lambda}(\hat{x},\hat{y}) + \lambda(\hat{y},\hat{x}).$$

Take $\lambda = 1$ and $\lambda = i$. Then (5) shows that (x,y) and $(\hat{x},\hat{y})$ have the same real and imaginary parts, hence are equal. Thus (c) implies (d).

Finally, if (a) is false, there exists a $u \neq 0$ in H so that $(u,u_\alpha) = 0$ for all $\alpha \in A$. If $x = y = u$, then $(x,y) = \|u\|^2 \neq 0$, but $\hat{x}(\alpha) = 0$ for all $\alpha \in A$, hence (d) fails. Thus (d) implies (a), and the proof is complete.

4.19 Isomorphisms Speaking informally, two algebraic systems of the same nature are said to be isomorphic if there is a one-to-one mapping of one onto the other which preserves all relevant properties. For instance, we may ask whether two groups are isomorphic or whether two fields are isomorphic. Two vector spaces are isomorphic if there is a one-to-one linear mapping of one onto the other. The linear mappings are the ones which preserve the relevant concepts in a vector space, namely, addition and scalar multiplication.

In the same way, two Hilbert spaces H_1 and H_2 are isomorphic if there is a one-to-one linear mapping Λ of H_1 onto H_2 which also preserves inner products: $(\Lambda x, \Lambda y) = (x,y)$ for all x and $y \in H_1$. Such a Λ is an isomorphism (or, more specifically, a *Hilbert space isomorphism*) of H_1 onto H_2. Using this terminology, Theorems 4.17 and 4.18 yield the following statement:

If $\{u_\alpha : \alpha \, \varepsilon \, A\}$ is a maximal orthonormal set in a Hilbert space H, and if $\hat{x}(\alpha) = (x, u_\alpha)$, then the mapping $x \rightarrow \hat{x}$ is a Hilbert space isomorphism of H onto $\ell^2(A)$.

One can prove (we shall omit this) that $\ell^2(A)$ and $\ell^2(B)$ are isomorphic if and only if the sets A and B have the same cardinal number. But we shall prove that every nontrivial Hilbert space (this means that the space does not consist of 0 alone) is isomorphic to some $\ell^2(A)$, by proving that every such space contains a maximal orthonormal set (Theorem 4.22). The proof will depend on a property of partially ordered sets which is equivalent to the axiom of choice.

4.20 Partially Ordered Sets A set $\mathcal{P}$ is said to be *partially ordered* by a binary relation $\leq$ if

(a) $a \leq b$ and $b \leq c$ implies $a \leq c$.

(b) $a \leq a$ for every $a \, \varepsilon \, \mathcal{P}$.

(c) $a \leq b$ and $b \leq a$ implies $a = b$.

A subset $\mathcal{Q}$ of a partially ordered set $\mathcal{P}$ is said to be *totally ordered* (or *linearly ordered*) if every pair a, $b \, \varepsilon \, \mathcal{Q}$ satisfies either $a \leq b$ or $b \leq a$.

For example, every collection of subsets of a given set is partially ordered by the inclusion relation $\subset$.

To give a more specific example, let $\mathcal{P}$ be the collection of all open subsets of the plane, partially ordered by set inclusion, and let $\mathcal{Q}$ be the collection of all open circular discs with center at the origin. Then $\mathcal{Q} \subset \mathcal{P}$, $\mathcal{Q}$ is totally ordered by $\subset$, and $\mathcal{Q}$ is a *maximal* totally ordered subset of $\mathcal{P}$. This means that if any member of $\mathcal{P}$ not in $\mathcal{Q}$ is adjoined to $\mathcal{Q}$, the resulting collection of sets is no longer totally ordered by $\subset$.

4.21 The Hausdorff Maximality Theorem *Every nonempty partially ordered set contains a maximal totally ordered subset.*

This is a consequence of the axiom of choice and is, in fact, equivalent to it; another (very similar) form of it is known as Zorn's lemma. We give the proof in the Appendix.

If now H is a nontrivial Hilbert space, then there exists a $u \, \varepsilon \, H$ with $\|u\| = 1$, so that there is a nonempty orthonormal set in H. The existence of a maximal orthonormal set is therefore a consequence of the following theorem:

4.22 Theorem *Every orthonormal set B in a Hilbert space H is contained in a maximal orthonormal set in H.*

PROOF Let $\mathcal{P}$ be the class of all orthonormal sets in H which contain the given set B. Partially order $\mathcal{P}$ by set inclusion. Since $B \, \varepsilon \, \mathcal{P}$,

$\mathcal{P} \neq \varnothing$. Hence $\mathcal{P}$ contains a maximal totally ordered class Ω. Let S be the union of all members of Ω. It is clear that $B \subset S$. We claim that S is a maximal orthonormal set:

If u_1 and $u_2 \varepsilon S$, then $u_1 \varepsilon A_1$ and $u_2 \varepsilon A_2$ for some A_1 and $A_2 \varepsilon \Omega$. Since Ω is totally ordered, $A_1 \subset A_2$ (or $A_2 \subset A_1$), so that $u_1 \varepsilon A_2$ and $u_2 \varepsilon A_2$. Since A_2 is orthonormal, $(u_1, u_2) = 0$ if $u_1 \neq u_2$, $(u_1, u_2) = 1$ if $u_1 = u_2$. Thus S is an orthonormal set.

Suppose S is not maximal. Then S is a proper subset of an orthonormal set S^*. Clearly, $S^* \notin \Omega$, and S^* contains every member of Ω. Hence we may adjoin S^* to Ω and still have a total order. This contradicts the maximality of Ω.

Trigonometric Series

4.23 Definitions Let T be the unit circle in the complex plane, i.e., the set of all complex numbers of absolute value 1. If F is any function on T and if f is defined on R^1 by

$$(1) \qquad\qquad f(t) = F(e^{it}),$$

then f is a periodic function of period 2π. This means that $f(t + 2\pi) = f(t)$ for all real t. Conversely, if f is a function on R^1, with period 2π, then there is a function F on T such that (1) holds. Thus we may identify functions on T with 2π-periodic functions on R^1; and, for simplicity of notation, we shall sometimes write $f(t)$ rather than $f(e^{it})$, even if we think of f as being defined on T.

With these conventions in mind, we define $L^p(T)$, for $1 \leq p < \infty$, to be the class of all complex, Lebesgue measurable, 2π-periodic functions on R^1 for which the norm

$$(2) \qquad\qquad \|f\|_p = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t)|^p \, dt \right\}^{1/p}$$

is finite.

In other words, we are looking at $L^p(\mu)$, where μ is Lebesgue measure on $[0, 2\pi)$ (or on T), divided by 2π. $L^\infty(T)$ will be the class of all 2π-periodic members of $L^\infty(R^1)$, with the essential supremum norm, and $C(T)$ consists of all continuous complex functions on T (or, equivalently, of all continuous, complex, 2π-periodic functions on R^1), with norm

$$(3) \qquad\qquad \|f\|_\infty = \sup_t |f(t)|.$$

The factor $1/(2\pi)$ in (2) simplifies the formalism we are about to develop. For instance, the L^p-norm of the constant function 1 is 1.

A *trigonometric polynomial* is a finite sum of the form

$$(4) \qquad f(t) = a_0 + \sum_{n=1}^{N} (a_n \cos nt + b_n \sin nt) \qquad (t \varepsilon R^1)$$

where a_0, a_1, . . . , a_N and b_1, . . . , b_N are complex numbers. On account of the Euler identities, (4) can also be written in the form

$$(5) \qquad\qquad f(t) = \sum_{n=-N}^{N} c_n e^{int}$$

which is more convenient for most purposes. It is clear that every trigonometric polynomial has period 2π.

We shall denote the set of all integers (positive, zero, and negative) by Z, and put

$$(6) \qquad\qquad u_n(t) = e^{int} \qquad (n \varepsilon Z).$$

If we define the inner product in $L^2(T)$ by

$$(7) \qquad\qquad (f,g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\overline{g(t)}\, dt$$

[note that this is in agreement with (2)], an easy computation shows that

$$(8) \qquad (u_n, u_m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)t}\, dt = \begin{cases} 1 & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases}$$

Thus $\{u_n : n \varepsilon Z\}$ is an orthonormal set in $L^2(T)$, usually called the *trigono-metric system*. We shall now prove that this system is maximal, and shall then derive concrete versions of the abstract theorems previously obtained in the Hilbert space context.

4.24 The Completeness of the Trigonometric System   Theorem 4.18 shows that the maximality (or completeness) of the trigonometric system will be proved as soon as we can show that the set of all trigonometric polynomials is dense in $L^2(T)$. Since $C(T)$ is dense in $L^2(T)$, by Theorem 3.14 (note that T is compact), it suffices to show that to every $f \varepsilon C(T)$ and to every $\epsilon > 0$ there is a trigonometric polynomial P such that $\|f - P\|_2 < \epsilon$. Since $\|g\|_2 \leq \|g\|_\infty$ for every $g \varepsilon C(T)$, the estimate $\|f - P\|_2 < \epsilon$ will follow from $\|f - P\|_\infty < \epsilon$, and it is this estimate which we shall prove.

Suppose we had trigonometric polynomials Q_1, Q_2, Q_3, . . . , with the following properties:

$$(a) \qquad\qquad Q_k(t) \geq 0 \text{ for } t \varepsilon R^1.$$

$$(b) \qquad\qquad \frac{1}{2\pi} \int_{-\pi}^{\pi} Q_k(t)\, dt = 1.$$

$$(c) \quad \textit{If } \eta_k(\delta) = \sup \{Q_k(t) : \delta \leq |t| \leq \pi\}, \textit{ then}$$

$$\lim_{k \to \infty} \eta_k(\delta) = 0$$

for every $\delta > 0$.

Another way of stating (c) is to say that $Q_k(t) \to 0$ uniformly on $[-\pi, -\delta] \cup [\delta, \pi]$, for every $\delta > 0$.

To each $f \in C(T)$ we associate the functions P_k defined by

$$(1) \qquad P_k(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t - s)Q_k(s)\, ds \qquad (k = 1, 2, 3, \ldots).$$

If we replace s by $-s$ and then by $s - t$, the periodicity of f and Q_k shows that the value of the integral is not affected. Hence

$$(2) \qquad P_k(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s)Q_k(t - s)\, ds \qquad (k = 1, 2, 3, \ldots).$$

Since each Q_k is a trigonometric polynomial, Q_k is of the form

$$(3) \qquad\qquad\qquad Q_k(t) = \sum_{n = -N_k}^{N_k} a_{n,k}e^{int},$$

and if we replace t by $t - s$ in (3) and substitute the result in (2), we see that each P_k is a trigonometric polynomial.

Let $\epsilon > 0$ be given. Since f is uniformly continuous on T, there exists a $\delta > 0$ such that $|f(t) - f(s)| < \epsilon$ whenever $|t - s| < \delta$. By (b), we have

$$P_k(t) - f(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{f(t - s) - f(t)\}Q_k(s)\, ds,$$

and (a) implies, for all t, that

$$|P_k(t) - f(t)| \le \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t - s) - f(t)|Q_k(s)\, ds = A_1 + A_2,$$

where A_1 is the integral over $[-\delta, \delta]$ and A_2 is the integral over

$$[-\pi, -\delta] \cup [\delta, \pi].$$

In A_1, the integrand is less than $\epsilon Q_k(s)$, so $A_1 < \epsilon$, by (b). In A_2, we have $Q_k(s) \le \eta_k(\delta)$, hence

$$(4) \qquad\qquad\qquad A_2 \le 2\|f\|_\infty \cdot \eta_k(\delta) < \epsilon$$

for sufficiently large k, by (c). Since these estimates are independent of t, we have proved that

$$(5) \qquad\qquad\qquad \lim_{k \to \infty} \|f - P_k\|_\infty = 0.$$

It remains to construct the Q_k. This can be done in many ways. Here is a simple one. Put

$$(6) \qquad\qquad\qquad Q_k(t) = c_k \left(\frac{1 + \cos t}{2}\right)^k,$$

where c_k is chosen so that (b) holds. Since (a) is clear, we only need to show (c). Since Q_k is even, (b) shows that

$$1 = \frac{c_k}{\pi} \int_0^\pi \left(\frac{1 + \cos t}{2}\right)^k dt > \frac{c_k}{\pi} \int_0^\pi \left(\frac{1 + \cos t}{2}\right)^k \sin t \, dt = \frac{2c_k}{\pi(k+1)}.$$

Since Q_k is decreasing on $[0,\pi]$, it follows that

$$(7) \qquad Q_k(t) \leq Q_k(\delta) \leq \frac{\pi(k+1)}{2}\left(\frac{1 + \cos \delta}{2}\right)^k \qquad (0 < \delta \leq |t| \leq \pi).$$

This implies (c), since $1 + \cos \delta < 2$ if $0 < \delta \leq \pi$.

We have proved the following important result:

4.25 Theorem *If $f \in C(T)$ and $\epsilon > 0$, there is a trigonometric polynomial P such that*

$$|f(t) - P(t)| < \epsilon$$

for every real t.

A more precise result was proved by Fejér (1904): *The arithmetic means of the partial sums of the Fourier series of any $f \in C(T)$ converge uniformly to f.* For a proof (quite similar to the above) see Theorem 8.15 of [26].

4.26 Fourier Series For any $f \in L^1(T)$, we define the *Fourier coefficients* of f by the formula

$$(1) \qquad \hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^\pi f(t)e^{-int} \, dt \qquad (n \in Z),$$

where, we recall, Z is the set of all integers. We thus associate with each $f \in L^1(T)$ a function $\hat{f}$ on Z. The *Fourier series* of f is

$$(2) \qquad \sum_{-\infty}^{\infty} \hat{f}(n)e^{int},$$

and its *partial sums* are

$$(3) \qquad s_N(t) = \sum_{-N}^{N} \hat{f}(n)e^{int} \qquad (N = 0, 1, 2, \ldots).$$

Since $L^2(T) \subset L^1(T)$, (1) can be applied to every $f \in L^2(T)$. Comparing the definitions made in Secs. 4.23 and 4.13, we can now restate Theorems 4.17 and 4.18 in concrete terms:

The *Riesz-Fischer theorem* asserts that if $\{c_n\}$ is a sequence of complex numbers such that

$$(4) \qquad \sum_{n=-\infty}^{\infty} |c_n|^2 < \infty,$$

then there exists an $f \varepsilon L^2(T)$ such that

$$(5) \qquad c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)e^{-int} \, dt \qquad (n \varepsilon Z).$$

The *Parseval theorem* asserts that

$$(6) \qquad \sum_{n=-\infty}^{\infty} \hat{f}(n)\overline{\hat{g}(n)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\overline{g(t)} \, dt$$

whenever $f \varepsilon L^2(T)$ and $g \varepsilon L^2(T)$; the series on the left of (6) converges absolutely; and if s_N is as in (3), then

$$(7) \qquad \lim_{N \to \infty} \|f - s_N\|_2 = 0,$$

since a special case of (6) yields

$$(8) \qquad \|f - s_N\|_2^2 = \sum_{|n|>N} |\hat{f}(n)|^2.$$

Note that (7) says that every $f \varepsilon L^2(T)$ is the L^2-limit of the partial sums of its Fourier series; i.e., the Fourier series of f converges to f, in the L^2-sense. Pointwise convergence presents a more delicate problem, as we shall see in Chap. 5.

The Riesz-Fischer theorem and the Parseval theorem may be summarized by saying that the mapping $f \to \hat{f}$ is a Hilbert space isomorphism of $L^2(T)$ onto $\ell^2(Z)$.

The theory of Fourier series in other function spaces, for instance in $L^1(T)$, is much more difficult than in $L^2(T)$, and we shall touch only a few aspects of it.

Observe that the crucial ingredient in the proof of the Riesz-Fischer theorem is the fact that L^2 is complete. This is so well recognized that the name "Riesz-Fischer theorem" is sometimes given to the theorem which asserts the completeness of L^2, or even of any L^p.

Exercises

In this set of exercises, H always denotes a Hilbert space.

1 If M is a closed subspace of H, prove that $M = (M^\perp)^\perp$. Is there a similar true statement for subspaces M which are not necessarily closed?

2 For $n = 1, 2, 3, \ldots$, let $\{v_n\}$ be an independent set of vectors in H. Develop a *constructive* process which generates an ortho-

normal set $\{u_n\}$, such that u_n is a linear combination of $v_1, \ldots ,$ v_n. Note that this leads to a proof of the existence of a maximal orthonormal set in separable Hilbert spaces which makes no appeal to the Hausdorff maximality principle. (A space is *separable* if it contains a countable dense subset.)

3 Show that $L^p(T)$ is separable if $1 \leq p < \infty$, but that $L^\infty(T)$ is not separable.

4 Show that H is separable if and only if H contains a maximal orthonormal system which is at most countable.

5 If $M = \{x: Lx = 0\}$, where L is a continuous linear functional on H, prove that $M^\perp$ is a vector space of dimension 1 (unless $M = H$).

6 Let $\{u_n\}$ $(n = 1, 2, 3, \ldots)$ be an orthonormal set in H. Show that this gives an example of a closed and bounded set which is not compact. Let Q be the set of all $x \, \varepsilon \, H$ of the form

$$x = \sum_1^\infty c_n u_n, \text{ where } |c_n| \leq \frac{1}{n}.$$

Prove that Q is compact. (Q is called the Hilbert cube.)

More generally, let $\{\delta_n\}$ be a sequence of positive numbers, and let S be the set of all $x \, \varepsilon \, H$ of the form

$$x = \sum_1^\infty c_n u_n, \text{ where } |c_n| \leq \delta_n.$$

Prove that S is compact if and only if $\sum_1^\infty \delta_n{}^2 < \infty$.

Prove that H is not locally compact.

7 Suppose $\{a_n\}$ is a sequence of positive numbers such that $\Sigma a_n b_n < \infty$ whenever $b_n \geq 0$ and $\Sigma b_n{}^2 < \infty$. Prove that $\Sigma a_n{}^2 < \infty$.

8 If H_1 and H_2 are two Hilbert spaces, prove that one of them is isomorphic to a subspace of the other. (Note that every closed subspace of a Hilbert space is a Hilbert space.)

9 If $A \subset [0,2\pi]$ and A is measurable, prove that

$$\lim_{n \to \infty} \int_A \cos nx \, dx = \lim_{n \to \infty} \int_A \sin nx \, dx = 0.$$

10 Let $n_1 < n_2 < n_3 < \cdots$ be positive integers, and let E be the set of all $x \, \varepsilon \, [0,2\pi]$ at which $\{\sin n_k x\}$ converges. Prove that $m(E) = 0$. *Hint:* $2 \sin^2 \alpha = 1 - \cos 2\alpha$, so $\sin n_k x \to \pm 1/\sqrt{2}$ a.e. on E, by Exercise 9.

11 Prove that the identity

$$4(x,y) = \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2$$

is valid in every inner product space, and show that it proves the implication $(c) \to (d)$ of Theorem 4.18.

12 The constants c_k in Sec. 4.24 were shown to be such that $k^{-1}c_k$ is bounded. Estimate the relevant integral more precisely and show that

$$0 < \lim_{k \to \infty} k^{-\frac{1}{2}} c_k < \infty.$$

13 Suppose f is a continuous function on R^1, with period 1. Prove that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(n\alpha) = \int_0^1 f(t) \, dt$$

for every irrational real number α.   *Hint:* Do it first for

$$f(t) = \exp(2\pi i k t),$$

$k = 0, \pm 1, \pm 2, \ldots$

14 Compute

$$\min_{a,b,c} \int_{-1}^{1} |x^3 - a - bx - cx^2|^2 \, dx$$

and find

$$\max \int_{-1}^{1} x^3 g(x) \, dx,$$

where g is subject to the restrictions

$$\int_{-1}^{1} g(x) \, dx = \int_{-1}^{1} xg(x) \, dx = \int_{-1}^{1} x^2 g(x) \, dx = 0; \qquad \int_{-1}^{1} |g(x)|^2 \, dx = 1.$$

15 Compute

$$\min_{a,b,c} \int_0^{\infty} |x^3 - a - bx - cx^2|^2 e^{-x} \, dx.$$

State and solve the corresponding maximum problem, as in Exercise 14.

16 If $x_0 \in H$ and M is a closed linear subspace of H, prove that

$$\min \{ \|x - x_0\| : x \in M \} = \max \{ |(x_0, y)| : y \in M^{\perp}, \|y\| = 1 \}.$$

17 Show that there is a continuous one-to-one mapping γ of $[0,1]$ into H such that $\gamma(b) - \gamma(a)$ is orthogonal to $\gamma(d) - \gamma(c)$ whenever $0 \le a \le b \le c \le d \le 1$. ($\gamma$ may be called a "curve with orthogonal increments.")   *Hint:* Take $H = L^2$, and consider characteristic functions of certain subsets of $[0,1]$.

18 Give a direct proof of Theorem 4.16, i.e., one which does not depend on the more general considerations of Sec. 4.15.