Introduction to Hilbert Space

ch1 sect.

§ 8. CODA

On looking over the foregoing definitions and theorems, it will be apparent to the reader that no use has been made of any particularly characteristic property of the system of complex scalars (such as the "fundamental theorem of algebra," real and imaginary parts, complex conjugation). The definitions and theorems make sense, and are valid, for other systems of scalars; for example, real scalars, or rational scalars, or Gaussian-rational scalars (complex numbers of the form $\alpha + i\beta$, where α and β are rational). More generally, the scalars may be drawn from an algebraic system known as a *field* (see any book on abstract algebra). Thus, one speaks of *complex vector spaces*, *real vector spaces*, and so on, in conformity with the system of scalars which is employed.

In the sequel, only complex vector spaces will be considered, and the terms **vector space** *and* **complex vector space** *will be used interchangeably.*

Exercises

1. (i) Every complex vector space can be regarded also as a real vector space.

(ii) The complex numbers form a one-dimensional complex vector space (*Example 1.8*), and a two-dimensional real vector space.

(iii) If $\mathcal{V}$ is an n-dimensional complex vector space, it is a $2n$-dimensional real vector space.

*2. View the real numbers as a rational vector space in the obvious way. Then $1, \sqrt{2}$ are independent. So are $1, \pi$. So are $1, \sqrt{2}, \pi$. What is the dimension?

Chapter

II Hilbert Spaces

§ 1. PRE-HILBERT SPACES

The *conjugate* of a complex number λ will be denoted λ^*. Thus, if $\lambda = \alpha + i\beta$, where α and β are real numbers, then $\lambda^* = \alpha - i\beta$. The familiar properties of conjugation are as follows: $(\lambda^*)^* = \lambda$, $(\lambda + \mu)^* = \lambda^* + \mu^*$, $(\lambda\mu)^* = \lambda^*\mu^*$, $|\lambda| = \sqrt{\lambda^*\lambda}$, and $\lambda^* = \lambda$ if and only if λ is real.

Definition 1. *A* **pre-Hilbert space** *is a complex vector space* $\mathcal{P}$. *For each pair of vectors* x,y *of* $\mathcal{P}$, *there is determined a complex number called the* **scalar product** *of* x *and* y, *denoted* $(x|y)$. *Scalar products are assumed to obey these rules:*

(P1) $(y|x) = (x|y)^*$

(P2) $(x + y|z) = (x|z) + (y|z)$

(P3) $(\lambda x|y) = \lambda(x|y)$

(P4) $(x|x) > 0$ *when* $x \neq 0$.

A convenient verbalization of $(x|y)$ is "x scalar y." The reason for the term *pre-Hilbert space* is that one can pass, by the procedure of "completion," from a pre-Hilbert space to a Hilbert space (see *Theorem V.2.1*).

Examples

1. Let $\mathcal{P}$ be the vector space of n-ples (*Example I.1.1*). If $x = (\lambda_1, \cdots, \lambda_n)$ and $y = (\mu_1, \cdots, \mu_n)$, define

$$(x|y) = \sum_1^n \lambda_k \mu_k^*.$$

The axioms (P1)–(P4) are easily verified. This example is known as n-dimensional *unitary space*, and will be denoted $\mathbb{C}^n$.

2. Let $\mathcal{P}$ be the vector space of continuous functions on $[a,b]$ (*Example I.1.3*), where $a < b$. Define

$$(x|y) = \int_a^b x(t)y(t)^* \, dt.$$

This example will be referred to as *the pre-Hilbert space of continuous functions on* $[a,b]$.

3. Let $\mathcal{P}$ be the vector space of finitely non-zero sequences (*Example I.1.6*). If $x = (\lambda_k)$ and $y = (\mu_k)$, define

$$(x|y) = \sum_1^\infty \lambda_k \mu_k^*.$$

Since this is essentially a finite sum, convergence is not an issue here. This example will be referred to as *the pre-Hilbert space of finitely non-zero sequences.*

4. If $\mathcal{P}$ is any pre-Hilbert space, and $\mathfrak{M}$ is a linear subspace of $\mathcal{P}$, obviously $\mathfrak{M}$ is itself a pre-Hilbert space.

Exercises

1. Fill in the details in the above examples (especially *Example 2*, for which the verification of (P4) is non-trivial).

2. The vector space $\mathbb{C}$ of complex numbers (*Example I.1.8*) is a pre-Hilbert space, via $(\lambda|\mu) = \lambda\mu^*$. [This is essentially the unitary space $\mathbb{C}^1$ of *Example 1*.]

3. Let $\mathcal{P}$ be the vector space of finitely non-zero functions defined on a set $\mathfrak{T}$ (*Example I.1.7*). Define

$$(x|y) = \sum_t x(t)y(t)^*,$$

the sum being extended over all t in $\mathfrak{T}$. Then, $\mathcal{P}$ is a pre-Hilbert space.

4. Let $\mathcal{P}$ be any vector space of finite dimension n, and $x_1, \cdots, x_n$ any basis of $\mathcal{P}$. If $x = \sum_1^n \lambda_k x_k$ and $y = \sum_1^n \mu_k x_k$, define $(x|y) = \sum_1^n \lambda_k \mu_k^*$. Then, $\mathcal{P}$ is a pre-Hilbert space.

§2. FIRST PROPERTIES OF PRE-HILBERT SPACES

Axioms (P2) and (P3) for a pre-Hilbert space can be expressed as follows: the scalar product $(x|y)$ is "additive" and "homogeneous" in the first factor. The first two statements of *Theorem 1* assert that $(x|y)$ is "additive" and "conjugate-homogeneous" in the second factor:

Theorem 1. *In any pre-Hilbert space:*

(1) $(x|y+z) = (x|y) + (x|z)$

(2) $(x|\lambda y) = \lambda^*(x|y)$

(3) $(0|y) = (x|0) = 0$

(4) $(x-y|z) = (x|z) - (y|z)$

$(x|y-z) = (x|y) - (x|z)$

(5) *If* $(x|z) = (y|z)$ *for all* z, *necessarily* $x = y$.

Proof.

(1): Using axioms (P1) and (P2), $(x|y+z) = (y+z|x)^* = [(y|x) + (z|x)]^* = (y|x)^* + (z|x)^* = (x|y) + (x|z)$.

(2): Using axioms (P1) and (P3), $(x|\lambda y) = (\lambda y|x)^* = [\lambda(y|x)]^* = \lambda^*(y|x)^* = \lambda^*(x|y)$.

(3): $(0|y) = (0+0|y) = (0|y) + (0|y)$, hence $(0|y) = 0$. Similarly, $(x|0) = 0$.

(4): $(x-y|z) = (x+(-y)|z) = (x|z) + (-y|z) = (x|z) + ((-1)y|z) = (x|z) + (-1)(y|z) = (x|z) - (y|z)$. Similarly for the second relation.

(5): Suppose $(x|z) = (y|z)$ for all z. Then, $(x - y|z) = (x|z) - (y|z) = 0$ for all z; in particular, $(x - y|x - y) = 0$, hence $x - y = \theta$ by axiom (P4). ∎

Exercises

1. In any pre-Hilbert space,

$$\left(\sum_1^n \lambda_k x_k \middle| y\right) = \sum_1^n \lambda_k (x_k|y)$$

$$\left(x \middle| \sum_1^m \mu_j y_j\right) = \sum_1^m \mu_j^* (x|y_j)$$

$$\left(\sum_1^n \lambda_k x_k \middle| \sum_1^m \mu_j y_j\right) = \sum_{k,j} \lambda_k \mu_j^* (x_k|y_j).$$

§3. THE NORM OF A VECTOR

Definition 1. *In a pre-Hilbert space, the* **norm** *(or "length") of a vector x, denoted $\| x \|$, is the non-negative real number defined by the formula* $\| x \| = \sqrt{(x|x)}$.

Suggested verbalization of $\| x \|$: "norm x."

Theorem 1. *In a pre-Hilbert space:*

(1) $\| \lambda x \| = |\lambda| \, \| x \|$.

(2) $\| x \| > 0$ *when $x \neq \theta$; $\| x \| = 0$ if and only if $x = \theta$.*

Proof.

(1): $\| \lambda x \|^2 = (\lambda x|\lambda x) = \lambda\lambda^*(x|x) = |\lambda|^2 \| x \|^2$.

(2): This is immediate from axiom (P4), and the relation $(\theta|\theta) = 0$. ∎

In particular, $\| -x \| = \| x \|$ and $\| ix \| = \| x \|$. If $x \neq \theta$, the vector $\| x \|^{-1} x$ has norm 1.

Examples

1. In the unitary space $\mathbb{C}^n$ (*Example 1.1*), if $x = (\lambda_k)$,

$$\| x \| = \left(\sum_1^n |\lambda_k|^2\right)^{1/2}.$$

2. In the pre-Hilbert space of continuous functions on $[a,b]$ (*Example 1.2*),

$$\| x \| = \left(\int_a^b |x(t)|^2 \, dt\right)^{1/2}.$$

3. In the pre-Hilbert space of finitely non-zero sequences (*Example 1.3*), if $x = (\lambda_k)$,

$$\| x \| = \left(\sum_1^\infty |\lambda_k|^2\right)^{1/2}. \quad ∎$$

The additivity of the scalar product yields an identity expressible simply in terms of norms:

Theorem 2. (Parallelogram law) *In a pre-Hilbert space,*

$$\| x + y \|^2 + \| x - y \|^2 = 2 \| x \|^2 + 2 \| y \|^2.$$

Proof.

One has $\| x + y \|^2 = (x + y|x + y) = (x|x) + (x|y) + (y|x) + (y|y) = \| x \|^2 + \| y \|^2 + (x|y) + (y|x)$. Replacing y by $-y$, $\| x - y \|^2 = \| x \|^2 + \| y \|^2 - (x|y) - (y|x)$. ∎

The norm of a vector is expressed, by definition, in terms of the scalar product. There is a useful formula which expresses the scalar product in terms of norms:

Theorem 3. (Polarization identity) *In a pre-Hilbert space,* $(x|y) = \frac{1}{4}\{ \| x + y \|^2 - \| x - y \|^2 + i \| x + iy \|^2 - i \| x - iy \|^2 \}$.

Proof.

In the identity

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

replace y by $-y$, iy, and $-iy$:

$$\| x - y \|^2 = \| x \|^2 + \| y \|^2 - (x|y) - (y|x)$$

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

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

It follows that

(b) $-\| x - y \|^2 = -\| x \|^2 - \| y \|^2 + (x|y) + (y|x)$

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

(d) $-i \| x - iy \|^2 = -i \| x \|^2 - i \| y \|^2 + (x|y) - (y|x)$.

Adding (a)–(d), the right hand side reduces to $4(x|y)$. ∎

From the definition of norm, $(x|x) = \| x \| \, \| x \|$. In general, $|(x|y)|$ is dominated by the product of $\| x \|$ and $\| y \|$:

Theorem 4. (Cauchy-Schwarz inequality) *In a pre-Hilbert space,* $|(x|y)| \leq \| x \| \, \| y \|$.

Proof.
If $x = 0$ or $y = 0$, then $(x|y) = 0$, and the conclusion is clear. Suppose, for instance, that $y \neq 0$. Dividing through the desired inequality by $\| y \|$, the problem is to show that $|(x|z)| \leq \| x \|$ when $\| z \| = 1$. For every complex number λ,

$$\begin{aligned}\| x - \lambda z \|^2 &= \| x \|^2 - \lambda^*(x|z) - \lambda(z|x) + \lambda\lambda^* \| z \|^2 \\ &= \| x \|^2 - (x|z)\lambda^* - \lambda(x|z)^* + \lambda\lambda^* \\ &= \| x \|^2 - (x|z)(x|z)^* \\ &\qquad + (x|z)(x|z)^* - (x|z)\lambda^* - \lambda(x|z)^* + \lambda\lambda^* \\ &= \| x \|^2 - |(x|z)|^2 + [(x|z) - \lambda][(x|z) - \lambda]^* \\ &= \| x \|^2 - |(x|z)|^2 + |(x|z) - \lambda|^2.\end{aligned}$$

In particular, for $\lambda_0 = (x|z)$, $0 \leq \| x - \lambda_0 z \|^2 = \| x \|^2 - |(x|z)|^2$. ∎

Applying *Theorem 4* in the unitary space $\mathbb{C}^n$ (see *Example 1*),

Corollary. *If $\lambda_1, \cdots, \lambda_n$ and $\mu_1, \cdots, \mu_n$ are complex numbers,*

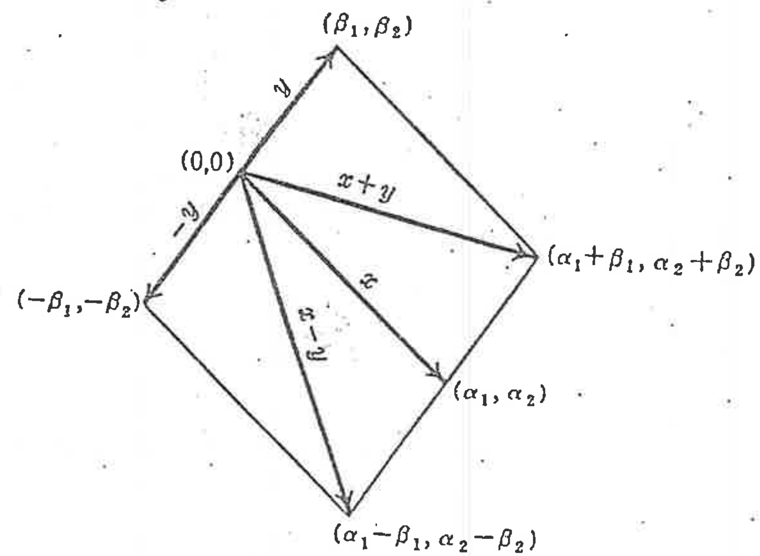
$$\left| \sum_1^n \lambda_k \mu_k^* \right| \leq \left(\sum_1^n |\lambda_k|^2 \right)^{1/2} \left(\sum_1^n |\mu_k|^2 \right)^{1/2}.$$

The Cauchy-Schwarz inequality leads to an inequality involving only norms:

Theorem 5. (Triangle inequality) *In a pre-Hilbert space, one has* $\| x + y \| \leq \| x \| + \| y \|$.

Proof.
Denote by $\rho\{\lambda\}$ the real part of the complex number λ; obviously $|\rho\{\lambda\}| \leq |\lambda|$. Applying the Cauchy-Schwarz inequality at the appropriate step, $\| x + y \|^2 = \| x \|^2 + \| y \|^2 + (x|y) + (x|y)^* = \| x \|^2 + \| y \|^2 + 2\rho\{(x|y)\} \leq \| x \|^2 + \| y \|^2 + 2|(x|y)| \leq \| x \|^2 + \| y \|^2 + 2 \| x \| \, \| y \| = (\| x \| + \| y \|)^2$. ∎

Consider, for example, the unitary space $\mathbb{C}^2$. A vector $x = (\alpha_1, \alpha_2)$ with *real* components may be interpreted, in the Cartesian plane, as the "arrow" from the origin (0,0) to the point (α_1, α_2). Then, $\| x \|$ is the familiar formula for the distance from (0,0) to (α_1, α_2). If also $y = (\beta_1, \beta_2)$, with real components, $x + y$ represents the *resultant* of the arrows x and y:



The triangle inequality and parallelogram law can now be given the obvious geometrical interpretations from which their name is drawn.
It is customary to carry over this geometrical language to pre-Hilbert spaces of arbitrary dimension:

Definition 2. *In a pre-Hilbert space, $\| x - y \|$ is called the* **distance** *from x to y.*

Certain properties of the norm have natural formulations in terms of distance:

Theorem 6. *In a pre-Hilbert space,*

(1) $\| x - y \| \geq 0$; $\| x - y \| = 0$ *if and only if* $x = y$.

(2) $\| x - y \| = \| y - x \|$

(3) $\| x - z \| \leq \| x - y \| + \| y - z \|$.

Proof.
(1) and (2) are clear from *Theorem 1*, and the relation $y - x = -(x - y)$.
(3) results from the relation $x - z = (x - y) + (y - z)$, and *Theorem 5*. ∎

Exercises

1. If x and y are continuous complex-valued functions on $[a,b]$, then

$$\left| \int_a^b x(t)y(t)^* \, dt \right|^2 \leq \int_a^b |x(t)|^2 \, dt \cdot \int_a^b |y(t)|^2 \, dt,$$

and

$$\left(\int_a^b |x(t) + y(t)|^2 \, dt \right)^{1/2} \leq \left(\int_a^b |x(t)|^2 \, dt \right)^{1/2} + \left(\int_a^b |y(t)|^2 \, dt \right)^{1/2}.$$

2. Let x and y be vectors in a pre-Hilbert space. If $x = \theta$ or $y = \theta$, the Cauchy-Schwarz inequality reduces to $0 = 0$. Assuming $x \neq \theta$ and $y \neq \theta$, show that $|(x|y)| = \| x \| \| y \|$ if and only if x and y are "proportional" (i.e. $y = \lambda x$ for suitable λ).

3. In a pre-Hilbert space, $| \| x \| - \| y \| | \leq \| x - y \|$.

4. Let x and y be non-zero vectors in a pre-Hilbert space. The relation $\| x + y \| = \| x \| + \| y \|$ holds if and only if $y = \alpha x$ for some real number $\alpha > 0$.

5. Let x,y,z be vectors in a pre-Hilbert space. The relation $\| x - z \| = \| x - y \| + \| y - z \|$ holds if and only if there exists a real number α, $0 \leq \alpha \leq 1$, such that $y = \alpha x + (1 - \alpha)z$.

6. Let x and y be non-zero vectors in a pre-Hilbert space. The relation $\| x - y \| = | \| x \| - \| y \| |$ holds if and only if $y = \alpha x$ for some real number $\alpha > 0$.

تجمع

7. In view of *Exercises 4* and *6*, one has $\| x + y \| = \| x \| + \| y \|$ if and only if $\| x - y \| = | \| x \| - \| y \| |$. This is also an immediate consequence of the parallelogram law.

8. Deduce another proof of the Cauchy-Schwarz inequality, using the polarization identity and the parallelogram law.

§4. METRIC SPACES

At the close of the preceding section, a notion of *distance* between vectors in a pre-Hilbert space was introduced. The role of this concept is clarified by abstracting certain essentials:

Definition 1. *A* **metric space** *is a set $\mathfrak{X}$, composed of objects called the* **points** *of the space. It is assumed that $\mathfrak{X}$ is non-empty (that is, contains at least one point). For each pair of points x and y of the space, there is determined a non-negative real number $d(x,y)$, called the* **distance** *from x to y, subject to the following axioms:*

(M1) $d(x,y) > 0$ *when* $x \neq y$;

$d(x,y) = 0$ *if and only if* $x = y$

(M2) $d(x,y) = d(y,x)$

(M3) $d(x,z) \leq d(x,y) + d(y,z)$.

In words, distance is *strictly positive*, *symmetric*, and satisfies the *triangle inequality*.

Examples

1. Every pre-Hilbert space $\mathcal{P}$ is a metric space, with $d(x,y) = \| x - y \|$ (see *Theorem 3.6*).

2. A non-empty set $\mathfrak{X}$ of complex numbers is a metric space, with $d(\lambda,\mu) = |\lambda - \mu|$.

3. Let $\mathfrak{X}$ be any non-empty set. Define $d(x,y) = 1$ when $x \neq y$, and $d(x,x) = 0$ for all x. It is easy to see that $\mathfrak{X}$ is a metric space; such metric spaces are called *discrete*.

4. If $\mathfrak{X}$ is a metric space, and $\mathcal{S}$ is a non-empty set of points of $\mathfrak{X}$, then $\mathcal{S}$ is itself a metric space (distances in $\mathcal{S}$ being measured as they already are in $\mathfrak{X}$).

5. Let $\mathfrak{X}$ be the vector space of continuous functions defined on $[a,b]$ (*Example I.1.3*), where $a < b$. Define $d(x,y) = \text{LUB}\{|x(t) - y(t)| : a \leq t \leq b\}$; that is, $d(x,y)$ is the least upper bound of the numbers $|x(t) - y(t)|$ as t varies over the interval $[a,b]$. Then, $\mathfrak{X}$ is a metric space. [Note: a continuous function on a closed interval *is* bounded.] ∎

Certain standard notions from the calculus of real numbers have natural generalizations to metric spaces:

1. A sequence of points x_n in a metric space is said to *converge* to the point x in case $d(x_n,x) \to 0$ as $n \to \infty$. This means: given any number $\epsilon > 0$, there is an index N such that $d(x_n,x) \leq \epsilon$ whenever $n \geq N$. The point x is then unique; for, if also $d(x_n,y) \to 0$, then $x = y$ results from $0 \leq d(x,y) \leq d(x,x_n) + d(x_n,y) \to 0 + 0$. The point x is called the *limit* of the sequence x_n. Notations: $x_n \to x$, or $x_n \to x$ as $n \to \infty$, or $x = \lim x_n$, etc. . . .

2. A sequence x_n is said to be *convergent* if there exists a point x such that $x_n \to x$. Otherwise, the sequence is said to be *divergent*.

3. A sequence x_n is said to be *Cauchy* in case $d(x_m,x_n) \to 0$ as $m,n \to \infty$. This means: given any number $\epsilon > 0$, there is an index N such that $d(x_m,x_n) \leq \epsilon$ whenever $m,n \geq N$. Every convergent sequence is Cauchy; for, if $x_n \to x$, then $d(x_m,x_n) \leq d(x_m,x) + d(x,x_n) \to 0 + 0$ as $m,n \to \infty$. Not every Cauchy sequence is convergent, as is shown in the following example (more sophisticated examples are given in *Examples 7* and *8*):

Example 6. Let $\mathfrak{X}$ be the open interval $(0,1)$ with $d(\alpha,\beta) = |\alpha - \beta|$. Let $\alpha_n = 1/n$ $(n = 1,2,3,\cdots)$. Then, α_n is Cauchy (since it is convergent in the metric space of all real numbers), but the only possible limit (namely 0) lies outside of $\mathfrak{X}$. ∎

In the metric space of all real numbers, with $d(\alpha,\beta) = |\alpha - \beta|$, every Cauchy sequence is convergent (this is the well-known *Cauchy criterion* for convergence). This is an example of an important type of metric space:

Definition 2. *A metric space $\mathfrak{X}$ is said to be* **complete** *in case every Cauchy sequence is convergent. Otherwise, $\mathfrak{X}$ is said to be* **incomplete**.

Examples of complete metric spaces are given in the exercises, and elsewhere in the sequel. The property of completeness owes its importance, to a certain extent, to the existence of prominent metric

spaces which do not possess it; two examples of such spaces, both pre-Hilbert spaces, will now be given.

Example 7. Let $\mathcal{P}$ be the pre-Hilbert space of finitely non-zero sequences (*Example 1.3*), with $d(x,y) = \| x - y \|$. It will be shown that $\mathcal{P}$ is an incomplete metric space, by exhibiting a Cauchy sequence that has no limit in $\mathcal{P}$. The proposed sequence x_n is

seq of seq

$$x_1 = (1,0,0,\cdots)$$
$$x_2 = (1,1/2,0,\cdots)$$
$$x_3 = (1,1/2,1/3,0,\cdots)$$
$$\cdots$$
$$x_n = (1,1/2,1/3,\cdots,1/n,0,\cdots).$$

For all $n,p = 1,2,3,\cdots$,

$$\| x_{n+p} - x_n \|^2 = \left\| \left(0,\cdots,0,\frac{1}{n+1},\frac{1}{n+2},\cdots,\frac{1}{n+p},0,\cdots\right) \right\|^2 = \sum_{n+1}^{n+p} \frac{1}{k^2};$$

$d(x_{n+p}, x_n) \to 0$ as $n \to \infty$

since the series $\sum_1^\infty 1/k^2$ is convergent, $d(x_{n+p},x_n) = \| x_{n+p} - x_n \| \to 0$ as $n \to \infty$. It follows that x_n is a Cauchy sequence of vectors.

Suppose (to the contrary) that $\mathcal{P}$ contained a vector $x = (\lambda_1,\lambda_2,\lambda_3,\cdots,\lambda_N,0,0,\cdots)$ such that $x_n \to x$. If $n \geq N$,

$$\| x_n - x \|^2 = \sum_1^n \left| \frac{1}{k} - \lambda_k \right|^2 + \sum_{n+1}^\infty |\lambda_k|^2 = \sum_1^n \left| \frac{1}{k} - \lambda_k \right|^2 + 0;$$

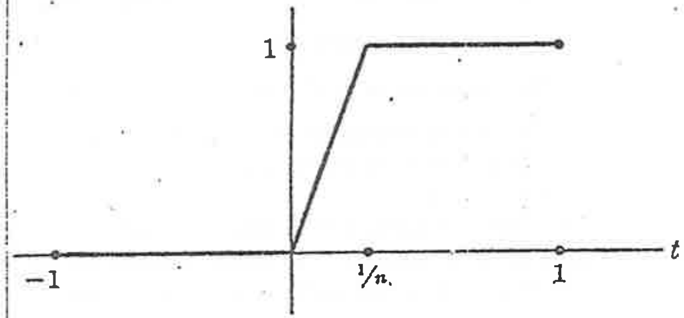
letting $n \to \infty$, $\sum_1^\infty \left| \frac{1}{k} - \lambda_k \right|^2 = 0$, hence $\lambda_k = 1/k$ for all k. This contradicts the assumption that x is finitely non-zero. ∎

Example 8. Let $\mathcal{P}$ be the pre-Hilbert space of continuous functions on a closed interval, say $[-1,1]$ for simplicity. Distances are defined as in *Example 1*:

$$d(x,y) = \| x - y \| = \left(\int_{-1}^{1} |x(t) - y(t)|^2 \, dt \right)^{1/2}.$$

pic

It will be shown that $\mathcal{P}$ is an incomplete metric space. The proposed Cauchy sequence without a limit is the sequence whose n'th term x_n has the following graph:



That is,

Function in pices

$$x_n(t) = \begin{cases} 0 & \text{for } -1 \le t \le 0 \\ nt & \text{for } 0 < t < 1/n \\ 1 & \text{for } 1/n \le t \le 1. \end{cases}$$

By elementary calculus, $\| x_m - x_n \| \to 0$ as $m,n \to \infty$; specifically, if $m > n$,

$$\int_{-1}^{1} |x_m(t) - x_n(t)|^2 \, dt = \frac{(m-n)^2}{3m^2 n}.$$

Assume to the contrary that $\mathcal{P}$ contains a (continuous) function x such that $x_n \to x$, that is,

$$\int_{-1}^{1} |x_n(t) - x(t)|^2 \, dt \to 0.$$

Since the integrands are ≥ 0,

$$\int_a^b |x_n(t) - x(t)|^2 \, dt \le \int_{-1}^{1} |x_n(t) - x(t)|^2 \, dt$$

for any sub-interval $[a,b]$ of $[-1,1]$, hence

$$\int_a^b |x_n(t) - x(t)|^2 \, dt \to 0.$$

In particular,

$$\int_{-1}^{0} |x_n(t) - x(t)|^2 \, dt \to 0;$$

since $x_n(t) = 0$ on $[-1,0]$, this reduces to

$$\int_{-1}^{0} |x(t)|^2 \, dt \to 0,$$

in other words,

$$\int_{-1}^{0} |x(t)|^2 \, dt = 0.$$

Since x is continuous, it follows that $x(t) = 0$ on $[-1,0]$.

Suppose $0 < \epsilon < 1$. One has

$$\int_{\epsilon}^{1} |x_n(t) - x(t)|^2 \, dt \to 0.$$

But, $x_n(t) = 1$ on $[\epsilon,1]$, provided $n > 1/\epsilon$, hence

$$\int_{\epsilon}^{1} |x_n(t) - x(t)|^2 \, dt = \int_{\epsilon}^{1} |1 - x(t)|^2 \, dt$$

for $n > 1/\epsilon$; letting $n \to \infty$, it follows that

$$\int_{\epsilon}^{1} |1 - x(t)|^2 \, dt = 0,$$

hence $x(t) = 1$ on $[\epsilon,1]$. Since $\epsilon > 0$ is arbitrary, $x(t) = 1$ on the semi-open interval $(0,1]$. Thus,

$$x(t) = \begin{cases} 0 & \text{for } -1 \le t \le 0 \\ 1 & \text{for } 0 < t \le 1. \end{cases}$$

This contradicts the assumption that x is continuous. ∎

Returning to general metric spaces,

Theorem 1. *In any metric space:*

(1) $|d(x,z) - d(y,z)| \le d(x,y)$

(2) *If $x_n \to x$ and $y_n \to y$, then $d(x_n,y_n) \to d(x,y)$.*

(3) *If x_n and y_n are Cauchy sequences, then $d(x_n,y_n)$ is a convergent sequence of real numbers.*

Proof.

(1): By the triangle inequality, $d(x,z) \le d(x,y) + d(y,z)$; transposing, $d(x,z) - d(y,z) \le d(x,y)$. Interchanging the role of x and y, $d(y,z) - d(x,z) \le d(y,x)$, that is, $-d(x,y) \le d(x,z) - d(y,z)$.

(2): Using part (1), and the triangle inequality for real numbers, $|d(x,y) - d(x_n,y_n)| \leq |d(x,y) - d(x_n,y)| + |d(x_n,y) - d(x_n,y_n)| \leq d(x,x_n) + d(y,y_n) \to 0 + 0$ as $n \to \infty$.

(3): Similarly, $|d(x_m,y_m) - d(x_n,y_n)| \leq d(x_m,x_n) + d(y_m,y_n) \to 0 + 0$ as $m,n \to \infty$. Since the real numbers are complete, the Cauchy sequence $d(x_n,y_n)$ converges. ∎

The concept of a bounded set of real numbers can be generalized as follows: a set $\mathcal{S}$ of points in a metric space is said to be *bounded* in case the distances $d(x,y)$ remain bounded as x and y vary over $\mathcal{S}$; that is, there is a constant $M \geq 0$ such that $d(x,y) \leq M$ whenever x and y are points of $\mathcal{S}$.

Examples

9. Let x_0 be a fixed point of the metric space $\mathfrak{X}$, and let ϵ be a fixed real number >0. The set $\mathcal{S}$, of all points x in $\mathfrak{X}$ such that $d(x,x_0) < \epsilon$, is bounded; for, if x and y are two such points, $d(x,y) \leq d(x,x_0) + d(x_0,y) < 2\epsilon$. $\mathcal{S}$ is called the *open ball* with *center* x_0 and *radius* ϵ. In particular, when $\mathfrak{X}$ is the metric space of real numbers, $\mathcal{S}$ is the open interval $(x_0 - \epsilon, x_0 + \epsilon)$.

10. Similarly, the set $\mathfrak{J}$ of all points x such that $d(x,x_0) \leq \epsilon$ is bounded; it is called the *closed ball* with center x_0 and radius ϵ. In particular, when $\mathfrak{X}$ is the metric space of real numbers, $\mathfrak{J}$ is the closed interval $[x_0 - \epsilon, x_0 + \epsilon]$.

11. Suppose x_0 is a fixed point of the metric space $\mathfrak{X}$, and $\mathcal{S}$ is a set of points of $\mathfrak{X}$. Then, $\mathcal{S}$ is bounded if and only if the numbers $d(x,x_0)$ are bounded as x varies over $\mathcal{S}$. For, suppose $d(x,x_0) \leq K$ for all x in $\mathcal{S}$; if x and y are points of $\mathcal{S}$, $d(x,y) \leq d(x,x_0) + d(x_0,y) \leq 2K$, thus $\mathcal{S}$ is bounded. Conversely, suppose $\mathcal{S}$ is bounded, say $d(x,y) \leq M$ for all x and y in $\mathcal{S}$; fix any point y_0 of $\mathcal{S}$; then $d(x,x_0) \leq d(x,y_0) + d(y_0,x_0) \leq M + d(y_0,x_0)$ for all x in $\mathcal{S}$.

Exercises

1. In a discrete metric space (*Example 3*), $x_n \to x$ if and only if there is an index N such that $x_n = x$ for all $n \geq N$.

2. In the metric space $\mathfrak{X}$ of *Example 5*, $x_n \to x$ means that $x_n(t) \to x(t)$ uniformly for t in $[a,b]$.

3. In the metric space $\mathbb{C}^m$ (see *Example 1*), $x_n \to x$ if and only if: for each $k = 1,\cdots,m$, the k'th component of x_n converges to the k'th component of x.

4. In the metric space $\mathbb{C}^m$, a sequence x_n is Cauchy if and only if: for each $k = 1,\cdots,m$, the sequence of k'th components is a Cauchy sequence of complex numbers. Deduce that $\mathbb{C}^m$ is complete.

5. Every discrete metric space is complete.

6. The metric space $\mathfrak{X}$ of *Example 5* is complete.

7. If $\mathcal{S}$ is a non-empty set of points in a metric space $\mathfrak{X}$, the following conditions on $\mathcal{S}$ are equivalent:
(a) $\mathcal{S}$ is bounded;
(b) given any point x of $\mathfrak{X}$, there is a ball centered at x (i.e. there exists a suitable radius $\epsilon > 0$) which includes all the points of $\mathcal{S}$;
(c) there exists a ball containing all the points of $\mathcal{S}$.

8. In a metric space, the following statements are equivalent:
(a) $x_n \to x$;
(b) each ball centered at x contains all but finitely many of the terms x_n;
(c) each open ball containing x contains all but finitely many of the terms x_n.
In part (c), can one replace "open ball" by "closed ball"?

§5. METRIC NOTIONS IN PRE-HILBERT SPACE; HILBERT SPACES

Let us translate the general metric space concepts of the preceding section into pre-Hilbert space terms; the dictionary is *Example 4.1*:

1. A sequence of vectors x_n *converges* to the *limit* vector x in case $\| x_n - x \| \to 0$ as $n \to \infty$; that is, given any $\epsilon > 0$, there is an index N such that $\| x_n - x \| \leq \epsilon$ whenever $n \geq N$. The vector x is then uniquely determined by the sequence x_n. Notations: $x_n \to x$, or $x_n \to x$ as $n \to \infty$, or $x = \lim x_n$, etc. . . .

2. A sequence of vectors x_n is *convergent* if there exists a vector x such that $x_n \to x$. Otherwise, the sequence is *divergent*.

3. A sequence of vectors x_n is *Cauchy* in case $\| x_m - x_n \| \to 0$ as $m,n \to \infty$; that is, given any $\epsilon > 0$, there is an index N such that $\| x_m - x_n \| \leq \epsilon$ whenever $m,n \geq N$. Every convergent sequence is Cauchy; the converse fails (see *Examples 4.7* and *4.8*).

4. A set S of vectors is *bounded* if there is a constant $M \geq 0$ such that $\| x \| \leq M$ for all x in S (see *Example 4.11*). In particular, a sequence of vectors x_n is bounded if there is a constant $M \geq 0$ such that $\| x_n \| \leq M$ for all n. If x_0 is a fixed vector, and $\epsilon > 0$, the set of all vectors x such that $\| x - x_0 \| < \epsilon$ is the *open ball* with *center* x_0 and *radius* ϵ; the *closed ball*, with center x_0 and radius ϵ, is the set of all vectors x such that $\| x - x_0 \| \leq \epsilon$; the set of all vectors x such that $\| x - x_0 \| = \epsilon$ is called the *sphere* with center x_0 and radius ϵ. In particular, the *open unit ball, closed unit ball*, and *unit sphere* are defined, respectively, by the conditions $\| x \| < 1$, $\| x \| \leq 1$, and $\| x \| = 1$.

5. A pre-Hilbert space is *complete* in case every Cauchy sequence converges. That is, if $\| x_m - x_n \| \to 0$, there exists a vector x such that $\| x_n - x \| \to 0$.

Definition 1. *A complete pre-Hilbert space is called a* **Hilbert space.**

In the sequel, the letters $\mathcal{H}$ and $\mathcal{K}$ will invariably denote Hilbert spaces. The most important example is the following:

Example 1. *The Hilbert space l^2.* Denote by $\mathcal{H}$ the set of all sequences $x = (\lambda_k)$ of complex numbers which are absolutely square-summable, that is, $\sum_1^\infty |\lambda_k|^2 < \infty$ (equivalently, the finite sums $\sum_1^n |\lambda_k|^2$ have a bound independent of n). For such an x, define $N(x) = \left(\sum_1^\infty |\lambda_k|^2\right)^{1/2}$.

If $x = (\lambda_k)$ and $y = (\mu_k)$, write $x = y$ in case $\lambda_k = \mu_k$ for all k.

Lemma 1. If $x = (\lambda_k)$ and $y = (\mu_k)$ are sequences belonging to $\mathcal{H}$, then so is the sequence $(\lambda_k + \mu_k)$, which is denoted $x + y$.

Proof:

By the parallelogram law for complex numbers, $|\lambda_k + \mu_k|^2 + |\lambda_k - \mu_k|^2 = 2|\lambda_k|^2 + 2|\mu_k|^2$, hence

$$\sum_1^n |\lambda_k + \mu_k|^2 \leq 2 \sum_1^n |\lambda_k|^2 + 2 \sum_1^n |\mu_k|^2$$

for all n. Clearly $\sum_1^\infty |\lambda_k + \mu_k|^2 < \infty$, by the "comparison test." ∎

If $x = (\lambda_k)$ belongs to $\mathcal{H}$, and λ is a complex number, $\sum_1^n |\lambda\lambda_k|^2 = |\lambda|^2 \sum_1^n |\lambda_k|^2$ shows that the sequence $(\lambda\lambda_k)$ is absolutely square-summable; it is denoted λx.

It is easy to see that $\mathcal{H}$ is a vector space, with respect to the operations $x + y$ and λx. Incidentally, $N(\lambda x) = |\lambda| N(x)$, and the proof of Lemma 1 shows that $N(x + y)^2 + N(x - y)^2 = 2N(x)^2 + 2N(y)^2$.

Lemma 2. If $x = (\lambda_k)$ and $y = (\mu_k)$ belong to $\mathcal{H}$, the series $\sum_1^\infty \lambda_k \mu_k{}^*$ converges absolutely.

Proof:

If a and b are real numbers, $(a - b)^2 \geq 0$ leads to $ab \leq \frac{1}{2}(a^2 + b^2)$; in particular, $|\lambda_k \mu_k{}^*| = |\lambda_k|\,|\mu_k| \leq \frac{1}{2}(|\lambda_k|^2 + |\mu_k|^2)$, thus $\sum_1^\infty |\lambda_k \mu_k{}^*|$ converges by the comparison test. ∎

Lemma 2 justifies the definition

$$(x|y) = \sum_1^\infty \lambda_k \mu_k{}^*.$$

The axioms for a pre-Hilbert space are easily verified. Incidentally, $\| x \| = N(x)$.

To show that $\mathcal{H}$ is a Hilbert space, it remains only to prove completeness. Suppose $x^1, x^2, x^3, \cdots$ is a Cauchy sequence in $\mathcal{H}$, that is, $\| x^m - x^n \| \to 0$ as $m,n \to \infty$. Say $x^n = (\lambda_k^n)$. For each k, $|\lambda_k^m - \lambda_k^n|^2 \leq \sum_{j=1}^\infty |\lambda_j^m - \lambda_j^n|^2 = \| x^m - x^n \|^2$ shows that the sequence $\lambda_k^1, \lambda_k^2, \lambda_k^3, \cdots$ of k'th components is Cauchy. Since the complex numbers are complete, $\lambda_k^n \to \lambda_k$ as $n \to \infty$, for suitable λ_k. It will be shown that $\sum_1^\infty |\lambda_k|^2 < \infty$, and that x^n converges to $x = (\lambda_k)$.

Given $\epsilon > 0$. Let p be an index such that $\| x^m - x^n \|^2 \leq \epsilon$ whenever $m,n \geq p$. Fix any positive integer r; one has

$$\sum_1^r |\lambda_k^m - \lambda_k^n|^2 \leq \| x^m - x^n \|^2 \leq \epsilon,$$

provided $m,n \geq p$; letting $m \to \infty$,

$$\sum_1^r |\lambda_k - \lambda_k^n|^2 \leq \epsilon$$

provided $n \geq p$; since r is arbitrary,

$$(*) \qquad \sum_1^\infty |\lambda_k - \lambda_k^n|^2 \leq \epsilon, \quad \text{whenever} \quad n \geq p.$$

In particular, $\sum_1^\infty |\lambda_k - \lambda_k^p|^2 \leq \epsilon$, hence the sequence $(\lambda_k - \lambda_k^p)$ belongs to $\mathcal{H}$; adding to it the sequence (λ_k^p) of $\mathcal{H}$, one obtains (λ_k), thus $x = (\lambda_k)$ belongs to $\mathcal{H}$. It follows from (*) that $\| x - x^n \|^2 \leq \epsilon$ whenever $n \geq p$. Thus, $x^n \to x$.

The Hilbert space $\mathcal{H}$ of absolutely square-summable sequences is denoted l^2. ∎

We resume the discussion of general pre-Hilbert spaces by noting some properties of Cauchy and convergent sequences of vectors:

Lemma. *Every Cauchy sequence is bounded.*

Proof.

Given a Cauchy sequence x_n, let N be an index such that $\| x_n - x_m \| \leq 1$ whenever $m,n \geq N$. If $n \geq N$, $\| x_n \| = \| (x_n - x_N) + x_N \| \leq \| x_n - x_N \| + \| x_N \| \leq 1 + \| x_N \|$. Thus, if M is the largest of the numbers $1 + \| x_N \|, \| x_1 \|, \| x_2 \|, \cdots, \| x_{N-1} \|$, one has $\| x_n \| \leq M$ for all n. ∎

Theorem 1. *In any pre-Hilbert space:*

(1) *If $x_n \to x$ and $y_n \to y$, then $(x_n|y_n) \to (x|y)$.*

(2) *If x_n and y_n are Cauchy sequences of vectors, then $(x_n|y_n)$ is a Cauchy (hence convergent) sequence of scalars.*

Proof.

(1): For all n, $(x_n|y_n) - (x|y) = (x_n - x|y_n - y) + (x|y_n - y) + (x_n - x|y)$. Using the triangle inequality for complex numbers, and the Cauchy-Schwarz inequality, one has $|(x_n|y_n) - (x|y)| \leq \| x_n - x \| \| y_n - y \| + \| x \| \| y_n - y \| + \| x_n - x \| \| y \|$; clearly the right hand side $\to 0$ as $n \to \infty$.

(2): Similarly, $|(x_n|y_n) - (x_m|y_m)| \leq \| x_n - x_m \| \| y_n - y_m \| + \| x_m \| \| y_n - y_m \| + \| x_n - x_m \| \| y_m \|$ for all m and n; since $\| x_m \|$ and $\| y_m \|$ are bounded by the *Lemma*, the right side $\to 0$ as $m,n \to \infty$. ∎

Corollary. *In any pre-Hilbert space:*

(1) *If $x_n \to x$, then $\| x_n \| \to \| x \|$.*

(2) *If x_n is Cauchy, then $\| x_n \|$ converges.*

what about the conv.

Exercises

1. Fill in the details in *Example 1*. $\|x_n\| \to \|x\|$

2. If $\mathfrak{M}$ is a linear subspace of a pre-Hilbert space $\mathcal{P}$, and $\mathfrak{M}$ contains a ball or a sphere, then $\mathfrak{M}$ contains every vector of $\mathcal{P}$.

3. In a metric space, every Cauchy sequence is bounded.

4. If $x_n \to \theta$ and y_n is bounded, then $(x_n|y_n) \to 0$. $x_n \to x$, $\|x_n - x\| \to 0$

5. Another proof of the Corollary to *Theorem 1* results from the inequality $| \| x \| - \| y \| | \leq \| x - y \|$.

6. If $\| x_n \| \to \| x \|$, and $(x_n|x) \to (x|x)$, then $x_n \to x$.

7. The argument in *Example 4.7* can be concluded as follows. Assume to the contrary that there is a vector $x = (\lambda_k)$ such that $x_n \to x$. Let $e_1 = (1,0,0,\cdots)$, $e_2 = (0,1,0,\cdots)$, $e_3 = (0,0,1,0,\cdots)$, $\cdots$. For each k, $(x_n|e_k) \to (x|e_k) = \lambda_k$. But $(x_n|e_k) = 1/k$ for $n \geq k$.

§6. ORTHOGONAL VECTORS, ORTHONORMAL VECTORS

Definition 1. *If x and y are vectors in a pre-Hilbert space, one says that x is* **orthogonal** *(or "perpendicular") to y in case $(x|y) = 0$. Notation: $x \perp y$.*

Suggested verbalization of $x \perp y$: "x perp y." The relation of orthogonality is symmetric: if $x \perp y$, then $y \perp x$; this results from $(y|x) = (x|y)^*$. If $x \perp x$, necessarily $x = \theta$. Every vector x is orthogonal to θ.

Theorem 1. *If x is orthogonal to each of $y_1, \cdots, y_n$, then x is orthogonal to every linear combination of the y_k.*

Proof.

If $x \perp y_k$ for all k, and $y = \sum_1^n \lambda_k y_k$, then $(x|y) = \sum_1^n \lambda_k^*(x|y_k) = \sum_1^n \lambda_k^* 0 = 0$. ∎

Definition 2. *A set $\mathcal{S}$ of vectors is said to be* **orthogonal** *in case $x \perp y$ whenever x and y are distinct vectors in $\mathcal{S}$. A sequence (finite or infinite) of vectors x_n is called an* **orthogonal sequence** *if $x_j \perp x_k$ whenever $j \neq k$.*

Examples

1. Let $\lambda_1, \lambda_2, \lambda_3, \cdots$ be any sequence of scalars. In the pre-Hilbert space of finitely non-zero sequences, define $x_1 = (\lambda_1,0,0,\cdots)$, $x_2 = (0,\lambda_2,0,\cdots)$, $x_3 = (0,0,\lambda_3,0,\cdots)$, $\cdots$. Then, x_n is an orthogonal sequence of vectors.

2. In the unitary space $\mathbb{C}^3$, the vectors $x_1 = (1,2,2)$, $x_2 = (2,1,-2)$, $x_3 = (2,-2,1)$ are orthogonal. Incidentally, $\| x_k \| = 3$ for all k.

3. In the pre-Hilbert space of continuous functions on $[-\pi,\pi]$, let the sequence of functions x_n be defined by the formulas $x_n(t) = \sin(nt)$ $(n = 1,2,3,\cdots)$. The sequence x_n is orthogonal, that is,

$$\int_{-\pi}^{\pi} \sin(mt)\sin(nt)\,dt = 0, \quad \text{when} \quad m \neq n.$$

Similarly the sequence $y_n(t) = \cos(nt)$ $(n = 0,1,2,3,\cdots)$ is orthogonal. Moreover, $x_m \perp y_n$ for all m and n.

Theorem 2. (Pythagorean relation) *If $x \perp y$, then*

$$\| x + y \|^2 = \| x \|^2 + \| y \|^2.$$

More generally, if $x_1,\cdots,x_n$ are orthogonal,

$$\| \textstyle\sum_1^n x_k \|^2 = \textstyle\sum_1^n \| x_k \|^2.$$

Proof.

If $x \perp y$, $\| x + y \|^2 = (x|x) + (x|y) + (y|x) + (y|y) = \| x \|^2 + 0 + 0 + \| y \|^2$. This is the case $n = 2$. Assume inductively that $\| \sum_1^{n-1} x_k \|^2 = \sum_1^{n-1} \| x_k \|^2$. Setting $x = \sum_1^{n-1} x_k$ and $y = x_n$, one has $x \perp y$ by *Theorem 1*; then, $\| \sum_1^n x_k \|^2 = \| x + y \|^2 = \| x \|^2 + \| y \|^2 = \sum_1^{n-1} \| x_k \|^2 + \| x_n \|^2$. ∎

Corollary. *If $x_1, x_2, x_3, \cdots$ is an orthogonal sequence (finite or infinite) of non-zero vectors, the x_k are linearly independent.*

Proof.

Suppose $\sum_1^n \lambda_k x_k = \theta$. The vectors $\lambda_1 x_1, \cdots, \lambda_n x_n$ are clearly orthogonal. By *Theorem 2*, $0 = \| \theta \|^2 = \sum_1^n \| \lambda_k x_k \|^2 = \sum_1^n |\lambda_k|^2 \| x_k \|^2$; since $\| x_k \| > 0$ for all k, necessarily $\lambda_k = 0$ for all k. ∎

Definition 3. *A set $\mathcal{S}$ of vectors in a pre-Hilbert space is said to be* **orthonormal** *in case* (i) *$\mathcal{S}$ is orthogonal in the sense of Definition 2, and* (ii) *$\| x \| = 1$ for every vector x in $\mathcal{S}$. A sequence (finite or infinite) of vectors x_n is called an* **orthonormal sequence** *if* (i) *$x_j \perp x_k$ whenever $j \neq k$, and* (ii) *$\| x_k \| = 1$ for all k.*

The condition for the orthonormality of a sequence can be expressed as follows: $(x_j|x_k) = \delta_{jk}$. In general, the "Kronecker delta" symbol δ_{st} can be defined for s and t varying over a set $\mathcal{J}$; it has the value 0 when $s \neq t$, and 1 when $s = t$. For example, in the notation of *Exercise I.7.4*, $x_s(t) = \delta_{st}$. A set $\mathcal{S}$ of vectors is orthonormal if and only if $(x|y) = \delta_{xy}$ for all x and y in $\mathcal{S}$.

Examples

4. If y_n is any orthogonal sequence of non-zero vectors, the sequence $x_n = \| y_n \|^{-1} y_n$ is orthonormal. For example:

5. In the unitary space $\mathbb{C}^3$, the vectors $(1/3,2/3,2/3)$, $(2/3,1/3,-2/3)$, $(2/3,-2/3,1/3)$ are orthonormal (see *Example 2*).

6. Notation as in *Example 3*. One has $\| y_0 \|^2 = 2\pi$, and $\| x_n \|^2 = \| y_n \|^2 = \pi$ for $n = 1,2,3,\cdots$. Define

$$u_n(t) = \frac{1}{\sqrt{\pi}} \sin(nt) \qquad (n = 1,2,3,\cdots)$$

$$v_0(t) = \frac{1}{\sqrt{2\pi}}$$

$$v_n(t) = \frac{1}{\sqrt{\pi}} \cos(nt) \qquad (n = 1,2,3,\cdots).$$

Then, the vectors u_m, v_n are orthonormal.

7. In the pre-Hilbert space of finitely non-zero sequences, let $e_1 = (1,0,0,\cdots)$, $e_2 = (0,1,0,\cdots)$, $e_3 = (0,0,1,0,\cdots),\cdots$. The sequence of vectors e_n is orthonormal.

8. In the Hilbert space l^2 (*Example 5.1*), let e_n be the orthonormal sequence described in *Example 7*. If $x = (\lambda_k)$, evidently $(x|e_k) = \lambda_k$. In particular, for every vector x, $\sum_1^\infty |(x|e_k)|^2 < \infty$; this result holds for any orthonormal sequence, as a consequence of "Bessel's inequality":

Theorem 3. (Bessel's equality and inequality) *Let $x_1,\cdots,x_n$ be orthonormal vectors in a pre-Hilbert space. For every vector x,*

(1) $\| x - \sum_1^n (x|x_k)x_k \|^2 = \| x \|^2 - \sum_1^n |(x|x_k)|^2,$

hence

(2) $\sum_1^n |(x|x_k)|^2 \leq \| x \|^2.$

Proof.

If $\lambda_1,\cdots,\lambda_n$ are arbitrary complex numbers, $\| \sum_1^n \lambda_k x_k \|^2 = \sum_1^n \| \lambda_k x_k \|^2 = \sum_1^n |\lambda_k|^2$ by *Theorem 2;* calculating as in the proof of *Theorem 3.4,*

$$\begin{aligned}\| x - \textstyle\sum_1^n \lambda_k x_k \|^2 &= \| x \|^2 - (\textstyle\sum_1^n \lambda_k x_k | x) \\ &\quad - (x | \textstyle\sum_1^n \lambda_k x_k) + \textstyle\sum_1^n |\lambda_k|^2 \\ &= \| x \|^2 - \textstyle\sum_1^n \lambda_k (x|x_k)^* \\ &\quad - \textstyle\sum_1^n (x|x_k)\lambda_k^* + \textstyle\sum_1^n \lambda_k \lambda_k^* \\ &= \| x \|^2 - \textstyle\sum_1^n |(x|x_k)|^2 + \textstyle\sum_1^n |(x|x_k) - \lambda_k|^2.\end{aligned}$$

In particular, setting $\lambda_k = (x|x_k)$, this reduces to the relation (1); the inequality (2) follows at once. ❙

Corollary. *If $x_1,x_2,x_3,\cdots$ is an orthonormal sequence, then for any vector x,*

$$\sum_1^\infty |(x|x_k)|^2 \le \| x \|^2.$$

In particular, $(x|x_k) \to 0$ as $k \to \infty$.

Proof.

Bessel's inequality holds for each n. ❙

Remarks

1. From the proof of *Theorem 3,* it is clear that the choice $\lambda_k = (x|x_k)$ *minimizes* $\| x - \sum_1^n \lambda_k x_k \|$, and thus provides a "best approximation" of x by a linear combination of $x_1,\cdots,x_n$. Moreover, only one set of coefficients gives best approximation, namely $\lambda_k = (x|x_k)$. Note that if $n > m$, then in the best approximation by $x_1,\cdots,x_n$, the first m coefficients are precisely those required for best approximation by $x_1,\cdots,x_m$.

2. Notation as in *Theorem 3.* Let $y = \sum_1^n (x|x_k)x_k$ and $z = x - y$. Clearly $(z|x_k) = 0$ for all k, hence $(z|y) = 0$. Thus, one has a decomposition $x = y + z$, where y is a linear combination of $x_1,\cdots,x_n$, and $z \perp x_k$ for all k. Such a decomposition is easily seen to be unique.

3. Bessel's inequality for $n = 1$ is essentially the Cauchy-Schwarz inequality; see the proof of *Theorem 3.4.*

4. If $y_k \to 0$, then $(x|y_k) \to (x|0) = 0$ for each vector x, by *Theorem 5.1.* The converse of this proposition fails; for, with notation as in the above *Corollary,* $(x|x_k) \to 0$ for each x, although $\| x_k \| = 1$ precludes $x_k \to 0$.

Example 9. Notation as in *Example 6.* The scalars $\lambda_k = (x|u_k)$ $(k = 1,2,3,\cdots)$ and $\mu_k = (x|v_k)$ $(k = 0,1,2,3,\cdots)$ are called the *Fourier coefficients* of the function x. By the above *Corollary,*

$$\sum_1^\infty |\lambda_k|^2 + \sum_0^\infty |\mu_k|^2 \le \int_{-\pi}^{\pi} |x(t)|^2 \, dt.$$

[It can be shown that the sum is actually *equal* to the integral.] ❙

Every orthonormal sequence of vectors is linearly independent, by the Corollary of *Theorem 2.* On the other hand, there is a systematic procedure for "orthonormalizing" any linearly independent sequence:

Theorem 4. (Gram-Schmidt orthonormalization procedure) *If $y_1,y_2,y_3,\cdots$ is a sequence of linearly independent vectors in a pre-Hilbert space, there exists an orthonormal sequence $x_1,x_2,x_3,\cdots$ such that $[x_1,\cdots,x_n] = [y_1,\cdots,y_n]$ for all n (that is, $x_1,\cdots,x_n$ generate the same linear subspace as $y_1,\cdots,y_n$).*

Proof.

The vectors x_n will be defined inductively. Let $x_1 = \| y_1 \|^{-1} y_1$. Assume inductively that orthonormal vectors $x_1,\cdots,x_{n-1}$ are already defined, in such a way that $[x_1,\cdots,x_k] = [y_1,\cdots,y_k]$ for $k = 1,\cdots,n-1$. The desired vector x_n must be a linear combination of $y_1,\cdots,y_n$, or equivalently of $x_1,\cdots,x_{n-1},y_n$; moreover, it must be orthogonal to each of $x_1,\cdots,x_{n-1}$. Guided by *Remark 2* following *Theorem 3,* set $z = y_n - \sum_1^{n-1} (y_n|x_k)x_k$; then, z is orthogonal to $x_1,\cdots,x_{n-1}$. Define $x_n = \| z \|^{-1} z$; this is permissible, since $z = \theta$ would imply that y_n is a linear combination of $x_1,\cdots,x_{n-1}$, hence of $y_1,\cdots,y_{n-1}$, contrary to the independence of the y's. The reader can easily verify that every linear combination of $x_1,\cdots,x_n$ is also a linear combination of $y_1,\cdots,y_n$, and vice versa. ❙

The Gram-Schmidt procedure applies equally well to a finite sequence $y_1,\cdots,y_n$ of independent vectors, and leads to orthonormal vectors $x_1, \cdots, x_n$ such that $[x_1, \cdots, x_k] = [y_1, \cdots, y_k]$ for $k = 1, \cdots, n$. In particular:

Corollary. *If $\mathcal{P}$ is a pre-Hilbert space of finite dimension n, $\mathcal{P}$ has a basis $x_1, \cdots, x_n$ of orthonormal vectors.*

Theorem 5. *Every finite-dimensional pre-Hilbert space is complete, hence is a Hilbert space.*

Proof.
By the above *Corollary*, there is a basis $x_1, \cdots, x_n$ of orthonormal vectors. If $x = \sum_1^n \lambda_k x_k$, then $\| x \|^2 = \sum_1^n |\lambda_k|^2$ by *Theorem 2;* completeness can now be proved using a simplification of the argument given for l^2 (see *Example 5.1*). ▌

Exercises

1. The relation $\| x + y \|^2 = \| x \|^2 + \| y \|^2$ is equivalent to $(x|y) + (y|x) = 0$; denote this relation by $x \mathrm{P} y$. Then $x \perp y$ if and only if $x \mathrm{P} \lambda y$ for all complex λ.

2. Let $\mathcal{P}$ be the pre-Hilbert space of finitely non-zero functions on a set $\mathfrak{J}$ (*Exercise 1.3*). For each point s of $\mathfrak{J}$, let x_s be defined as in *Exercise I.7.4*. Then, the x_s are an orthonormal set of vectors. That is, $(x_s|x_t) = \delta_{st}$ for all s and t in $\mathfrak{J}$.

3. In the pre-Hilbert space of continuous functions on a symmetric interval $[-a,a]$, every odd function x is orthogonal to every even function y (see *Exercise I.5.6*).

4. Starting with the vector $x_1 = (1,2,2,4)$ in the unitary space $\mathbb{C}^4$, construct an orthogonal set of vectors x_1, x_2, x_3, x_4 such that $\| x_k \| = 5$ for all k (integer components, preferably).

5. In the pre-Hilbert space of finitely non-zero sequences, orthonormalize the sequence of vectors $y_1 = (1,0,0,\cdots)$, $y_2 = (1,1,0,\cdots)$, $y_3 = (1,1,1,0,\cdots), \cdots$.

6. In the pre-Hilbert space of continuous functions on $[0,1]$, orthonormalize the first three terms of the sequence $y_n(t) = t^{n-1}$ $(n = 1,2,3,\cdots)$.

§7. INFINITE SUMS IN HILBERT SPACE

In the Hilbert space l^2, consider the orthonormal sequence e_n described in *Example 6.8*. If $x = (\lambda_1, \cdots, \lambda_m, 0, \cdots)$ is a finitely non-zero sequence, clearly $x = \sum_1^m \lambda_k e_k$; one could formally write $x = \sum_1^\infty \lambda_k e_k$, with the understanding that $\lambda_k = 0$ for all $k > m$.

Consider now an arbitrary vector $x = (\lambda_k)$ in l^2. What sense can be made of the expression $x = \sum_1^\infty \lambda_k e_k$? It is natural to define $\sum_1^\infty \lambda_k e_k$ to be the limit of the sequence of "partial sums" $y_n = \sum_1^n \lambda_k e_k$; this limit exists, in fact $y_n \to x$, since $\| x - y_n \|^2 = \| (0, \cdots, 0, \lambda_{n+1}, \lambda_{n+2}, \cdots) \|^2 = \sum_{n+1}^\infty |\lambda_k|^2 \to 0$ as $n \to \infty$.

These considerations will now be generalized for arbitrary orthonormal sequences in Hilbert space.

Lemma. *Let $x_1, x_2, x_3, \cdots$ be an orthonormal sequence of vectors in a pre-Hilbert space, and $\lambda_1, \lambda_2, \lambda_3, \cdots$ a sequence of scalars such that $\sum_1^\infty |\lambda_k|^2 < \infty$. Define $y_n = \sum_1^n \lambda_k x_k$. Then, the sequence y_n is Cauchy.*

Proof.

$$\| y_{n+p} - y_n \|^2 = \| \sum_{n+1}^{n+p} \lambda_k x_k \|^2 = \sum_{n+1}^{n+p} \| \lambda_k x_k \|^2$$

$$= \sum_{n+1}^{n+p} |\lambda_k|^2 \to 0 \quad \text{as} \quad n \to \infty. \quad ▌$$

Throwing in completeness, we have

Theorem 1. *If x_n is an orthonormal sequence of vectors in a Hilbert space, and λ_n is a sequence of scalars such that $\sum_1^\infty |\lambda_k|^2 < \infty$, then the sequence $y_n = \sum_1^n \lambda_k x_k$ converges to a limit x, denoted $x = \sum_1^\infty \lambda_k x_k$.*

More generally:

Definition 1. *If $x_1, x_2, x_3, \cdots$ is a sequence of vectors in a pre-Hilbert space, such that the sequence $y_n = \sum_1^n x_k$ converges to a limit x, one writes $x = \sum_1^\infty x_k$.*

Thus, $\| \sum_1^\infty x_k - \sum_1^n x_k \| \to 0$, by definition. The basic properties of the infinite sums described in *Theorem 1* are as follows:

Theorem 2. *Let x_n be an orthonormal sequence of vectors in a Hilbert space. Suppose $x = \sum_1^\infty \lambda_k x_k$ and $y = \sum_1^\infty \mu_k x_k$, in the sense of Theorem 1. Then:*

(1) $(x|y) = \sum_1^\infty \lambda_k \mu_k^*$, *the series converging absolutely.*

(2) $(x|x_k) = \lambda_k$

(3) $\| x \|^2 = \sum_1^\infty |\lambda_k|^2 = \sum_1^\infty |(x|x_k)|^2$.

Proof.

(1): Let $s_n = \sum_1^n \lambda_k x_k$ and $t_n = \sum_1^n \mu_k x_k$. By definition, $s_n \to x$ and $t_n \to y$, hence $(s_n|t_n) \to (x|y)$ by *Theorem 5.1*. Since $(s_n|t_n) = \sum_{j,k=1}^n \lambda_j \mu_k^* (x_j|x_k) = \sum_1^n \lambda_k \mu_k^*$, one has $(x|y) = \sum_1^\infty \lambda_k \mu_k^*$. Replacing (λ_k) by $(|\lambda_k|)$, and (μ_k) by $(|\mu_k|)$, it is clear that the convergence is absolute (see also *Example 5.1*).

(2): This is a special case of (1), with $\mu_k = 1$ and $\mu_j = 0$ for all $j \neq k$.

(3): Take $y = x$ in (1). ∎

Exercises

1. If x_n is an orthogonal sequence of vectors in a pre-Hilbert space, such that $\sum_1^\infty \| x_k \|^2 < \infty$, then the sequence $y_n = \sum_1^n x_k$ is Cauchy. What if $\sum_1^\infty \| x_k \| < \infty$?

2. Let x_n be a sequence of (not necessarily orthogonal) vectors in a pre-Hilbert space, such that $\sum_1^\infty \| x_k \| < \infty$. Then the sequence $y_n = \sum_1^n x_k$ is Cauchy. Hence, in Hilbert space, $\sum_1^\infty x_k$ exists.

3. If x_n is an orthonormal sequence, and $\sum_1^\infty \lambda_k x_k$ exists in the sense of *Definition 1*, necessarily $\sum_1^\infty |\lambda_k|^2 < \infty$.

4. Give an example of a sequence x_n such that $\sum_1^\infty \| x_k \|^2 < \infty$, but for which the sequence $y_n = \sum_1^n x_k$ is not Cauchy.

5. If $y = \sum_1^\infty y_k$, and $x \perp y_k$ for all k, then $x \perp y$.

6. (*Generalized Pythagorean relation*) Let x_n be an orthogonal sequence in Hilbert space, such that $\sum_1^\infty \| x_k \|^2 < \infty$, and form the vector $x = \sum_1^\infty x_k$ according to *Exercise 1*. Then $\| x \|^2 = \sum_1^\infty \| x_k \|^2$.

7. Suppose x_n is an orthonormal sequence in a pre-Hilbert space, and x is a vector such that $\| x \|^2 = \sum_1^\infty |(x|x_k)|^2$. Then $x = \sum_1^\infty (x|x_k)x_k$.

8. If $y_1, y_2, y_3, \cdots$ is a sequence of vectors in a Hilbert space $\mathcal{H}$, such that every vector x is a linear combination of finitely many y_k, then $\mathcal{H}$ is necessarily finite-dimensional.

§8. TOTAL SETS, SEPARABLE HILBERT SPACES, ORTHONORMAL BASES

Suppose x_n is an orthonormal sequence of vectors in a Hilbert space. Given any vector x, the scalars $\lambda_k = (x|x_k)$ satisfy $\sum_1^\infty |\lambda_k|^2 < \infty$, by the Corollary of *Theorem 6.3*. According to *Theorem 7.1*, one can form the vector $y = \sum_1^\infty \lambda_k x_k$, and by *Theorem 7.2*, $(y|x_k) = \lambda_k = (x|x_k)$ for all k. When can one conclude that $y = x$? In any case, $(y - x|x_k) = (y|x_k) - (x|x_k) = 0$ for all k. Thus, one could conclude $y = x$ if the vectors x_k had the following property: the only vector z which is orthogonal to every x_k is the vector $z = 0$. This leads to the following definitions:

Definition 1. *A set S of vectors in a pre-Hilbert space $\mathcal{P}$ is said to be* **total** *in case the only vector z of $\mathcal{P}$ which is orthogonal to every vector of S is the vector $z = 0$. A sequence (finite or infinite) of vectors x_n is called* **a total sequence** *in case: if $z \perp x_k$ for all k, necessarily $z = 0$.*

Examples

1. In any pre-Hilbert space $\mathcal{P}$, $\mathcal{P}$ is itself a total set of vectors. For, if $z \perp x$ for every vector x, in particular $z \perp z$.

2. If S is any system of generators (*Definition I.5.4*) for the pre-Hilbert space $\mathcal{P}$, S is total. For, if z is orthogonal to every vector in S, it is orthogonal to every linear combination of vectors in S; in particular, $z \perp z$.

3. In the Hilbert space l^2 (or in the pre-Hilbert space of finitely non-zero sequences), the sequence of vectors $e_1 = (1,0,0,\cdots)$, $e_2 = (0,1,0,\cdots)$, $e_3 = (0,0,1,0,\cdots)$, $\cdots$ is total. So is the sequence $x_1 = (1,0,0,\cdots)$, $x_2 = (1,1,0,\cdots)$, $x_3 = (1,1,1,0,\cdots)$,

*4. In the pre-Hilbert space of continuous functions on $[-\pi,\pi]$, it can be shown that the functions $1,t,t^2,\cdots$ form a total sequence. So do the functions $u_1,u_2,u_3,\cdots,v_0,v_1,v_2,v_3,\cdots$ described in *Example 6.6*.

Definition 2. *A sequence (finite or infinite) of vectors x_n is called an* **orthonormal basis** *for a Hilbert space $\mathcal{H}$, if it is* (i) *orthonormal, and* (ii) *total.*

Not every Hilbert space possesses such a sequence; those which do are characterized in *Theorem 3* below.

Example 5. The sequence e_n described in *Example 3* is an orthonormal basis for the Hilbert space l^2. This will be referred to as the *canonical orthonormal basis* of l^2. ▌

If a Hilbert space $\mathcal{H}$ has an orthonormal basis x_n consisting of infinitely many vectors, $\mathcal{H}$ is an infinite-dimensional vector space, by the Corollary of *Theorem 6.2*. It follows that the x_n cannot form a *basis*, in the sense of *Definition I.7.2*, for the vector space $\mathcal{H}$; that is, not every vector x can be expressed as a linear combination of the x_k (see *Exercise 7.8*).

On the other hand, suppose $\mathcal{P}$ is a pre-Hilbert space possessing a finite sequence $x_1,\cdots,x_n$ which is orthonormal and total. If x is any vector in $\mathcal{P}$, the vector $x - \sum_1^n (x|x_k)x_k$ is orthogonal to every x_k, hence is θ. Thus, $x = \sum_1^n (x|x_k)x_k$, $\mathcal{P}$ is finite-dimensional, and $x_1,\cdots,x_n$ is a basis for $\mathcal{P}$ in the sense of *Definition I.7.2*. Clearly, in a finite-dimensional space, the concepts "orthonormal basis" and "basis consisting of orthonormal vectors" coincide.

By the remarks at the beginning of the section,

Theorem 1. *If x_n is an orthonormal basis for the infinite-dimensional Hilbert space $\mathcal{H}$, then for each vector x one has $x = \sum_1^\infty (x|x_k)x_k$.*

Alternative descriptions of an orthonormal basis are contained in the following:

Theorem 2. *If x_n is an orthonormal infinite sequence in a Hilbert space $\mathcal{H}$, the following conditions are equivalent:*

(a) *The x_n are an orthonormal basis.*

(b) $\sum_1^\infty |(x|x_k)|^2 = \|x\|^2$, *for each vector x.*

(c) $\sum_1^\infty (x|x_k)x_k = x$, *for each vector x.*

Proof.

(a) *implies* (c) by *Theorem 1*.

(c) *implies* (b) by *Theorem 7.2*.

(b) *implies* (a): For, if $(x|x_k) = 0$ for all k, clearly $x = \theta$ by (b). ▌

The obvious finite-dimensional analogs of *Theorems 1* and *2* are true; the proofs are elementary.

Which Hilbert spaces possess an orthonormal basis? Such a space necessarily contains a total *sequence;* it will be shown in *Theorem 3* that this condition is also sufficient for the existence of an orthonormal basis.

Definition 3. *A Hilbert space is said to be* **separable** *if it possesses a total sequence (finite or infinite).*

Examples

6. If $\mathcal{H}$ is a finite-dimensional Hilbert space, every basis $y_1,\cdots,y_n$ is total, hence $\mathcal{H}$ is separable.

7. The Hilbert space l^2 is separable (see *Example 5*).

Theorem 3. *The following conditions on a Hilbert space $\mathcal{H}$ are equivalent:*

(a) *$\mathcal{H}$ is separable;*

(b) *$\mathcal{H}$ has an orthonormal basis x_n.*

Proof.

(b) *implies* (a): This is clear from *Definitions 2* and *3*.

(a) *implies* (b): Suppose $z_1,z_2,z_3,\cdots$ is a total sequence in $\mathcal{H}$. By

P. 19 ← *Theorem I.6.2*, there is a linearly independent subsequence $y_1,y_2,y_3,\cdots$ of the z_k, generating the same linear subspace as the z_k. The y_k are also total; for, if a vector z is orthogonal to every y_k, it is orthogonal to every linear combination of the y_k, hence to every z_k (hence $z = \theta$). By *Theorem 6.4*, there is an orthonormal sequence $x_1,x_2,x_3,\cdots$ generating the same linear subspace as the y_k. The x_k are total, by the above reasoning; thus, the x_k are an orthonormal basis.

Incidentally, if $\mathcal{H}$ is finite-dimensional, the independent sequence y_k must be finite, say $y_1,\cdots,y_n$; then $x_1,\cdots,x_n$ is an orthonormal basis. Both the y_k and the x_k are bases in the sense of Chapter I (see the remarks preceding *Theorem 1*). ▌

There is another frequently used definition of "separable" Hilbert space, equivalent to the one given in *Definition 3*; this material is sketched in *Exercises 3–5*.

Exercises

1. In the pre-Hilbert space of finitely non-zero sequences $x = (\lambda_k)$, let $\mathcal{S}$ be the set of all vectors x such that $\sum_1^\infty (1/k)\lambda_k = 0$. Then $\mathcal{S}$ is a total set (actually a linear subspace).

2. If a pre-Hilbert space $\mathcal{P}$ possesses a finite total set $x_1, \cdots, x_m$, then $\mathcal{P}$ is finite-dimensional (hence is a Hilbert space). An independent set $y_1, \cdots, y_n$ is a basis if and only if it is total.

3. A metric space is said to be *separable* if it contains a sequence x_n with the following property: given any point x in the space, there is a subsequence x_{n_k} converging to x. Such a sequence x_n is called a *dense sequence*. Show that if x_n is a dense sequence in a Hilbert space $\mathcal{H}$, then x_n is a total sequence, hence $\mathcal{H}$ is a separable Hilbert space in the sense of *Definition 3*.

*4. If $\mathcal{H}$ is a Hilbert space of finite dimension n, and $x_1, \cdots, x_n$ is an orthonormal basis, then the vectors of the form $x = \sum_1^n \gamma_k x_k$, where the γ_k are Gaussian-rational (see Chapter I, §8), can be enumerated in a sequence, and this sequence is dense. Thus, $\mathcal{H}$ is a separable metric space in the sense of *Exercise 3*. [The sophisticated point is the enumeration.]

*5. Suppose $\mathcal{H}$ is an infinite-dimensional Hilbert space, separable in the sense of *Definition 3*. Then $\mathcal{H}$ is a separable metric space in the sense of *Exercise 3*.

*6. Verify *Example 4*.

*7. It can be shown that every Hilbert space contains a total orthonormal set $\mathcal{J}$; such a set is called an *orthonormal basis* for the space. [However, it may not be possible to enumerate $\mathcal{J}$ in a sequence.]

§9. ISOMORPHIC HILBERT SPACES; CLASSICAL HILBERT SPACE

Definition 1. *A Hilbert space $\mathcal{H}$ is said to be* **isomorphic** *with a Hilbert space $\mathcal{K}$ if there exists a function T which assigns, to each vector x in $\mathcal{H}$, one and only one vector Tx in $\mathcal{K}$, in such a way that the following conditions hold:*

(i) *If x and y are vectors of $\mathcal{H}$ such that $x \neq y$, then $Tx \neq Ty$.*

(ii) *If z is any vector in $\mathcal{K}$, there is a vector x in $\mathcal{H}$ such that $Tx = z$.*

(iii) *$T(x + y) = Tx + Ty$, for all x and y in $\mathcal{H}$.*

(iv) *$T(\lambda x) = \lambda(Tx)$, for all x in $\mathcal{H}$, and all scalars λ.*

(v) *$(Tx|Ty) = (x|y)$, for all x and y in $\mathcal{H}$.*

Such a function T is called a **Hilbert space isomorphism** *of $\mathcal{H}$ onto $\mathcal{K}$.*

Isomorphic Hilbert spaces are in a sense "equal," for an isomorphism T distinguishes between distinct points, and "preserves" sums, scalar multiples, and scalar products. One can think of $\mathcal{K}$ as being essentially the space $\mathcal{H}$ with a "tag" T attached to each vector x.

Two types of separable Hilbert spaces were noted in *Examples 8.6* and *8.7*: (1) for each n, the unitary space $\mathbb{C}^n$ is a separable Hilbert space, and (2) the Hilbert space l^2 is separable. Up to isomorphism, these are the *only* separable Hilbert spaces:

Theorem 1. *Let $\mathcal{H}$ be a separable Hilbert space:*

(1) *If $\mathcal{H}$ has finite dimension n, it is isomorphic with $\mathbb{C}^n$.*

(2) *If $\mathcal{H}$ is infinite-dimensional, it is isomorphic with l^2.*

Proof.

Let us show (2); the proof of (1) is much simpler, and is left to the reader.

Suppose $\mathcal{H}$ is infinite-dimensional and separable. An isomorphism T of $\mathcal{H}$ onto l^2 will be constructed. By *Theorem 8.3*, $\mathcal{H}$ has an orthonormal basis x_n.

Given any vector x in $\mathcal{H}$; the problem is to define Tx in l^2. By the Corollary of *Theorem 6.3*, $\sum_1^\infty |(x|x_k)|^2 < \infty$, hence we may define $Tx = ((x|x_k))$; that is, Tx is the sequence whose k'th term is $(x|x_k)$.

Incidentally, $x = \sum_1^\infty (x|x_k)x_k$, by *Theorem 8.1*. Let us verify the properties (i)–(v) of *Definition 1:*

(i): If $x \neq y$, then $x - y \neq \theta$; since the x_k are total, $(x - y|x_j) \neq 0$ for some index j, thus $(x|x_j) \neq (y|x_j)$. It follows that $Tx \neq Ty$, by the definition of equality in l^2 (see *Example 5.1*).

(ii): Given any vector (λ_k) in l^2. Define $x = \sum_1^\infty \lambda_k x_k$ as in *Theorem 7.1*. Since $(x|x_k) = \lambda_k$ by *Theorem 7.2*, one has $Tx = (\lambda_k)$.

(iii): If x and y are vectors in $\mathcal{H}$, $(x + y|x_k) = (x|x_k) + (y|x_k)$ for all k; this shows that $T(x + y) = Tx + Ty$.

(iv): Similarly, $T(\lambda x) = \lambda(Tx)$ results from $(\lambda x|x_k) = \lambda(x|x_k)$.

(v): If x and y are vectors in $\mathcal{H}$, then $x = \sum_1^\infty (x|x_k)x_k$ and $y = \sum_1^\infty (y|x_k)x_k$, hence $(x|y) = \sum_1^\infty (x|x_k)(y|x_k)^*$ by *Theorem 7.2*. In other words, $(x|y) = (Tx|Ty)$. ∎

Thus, up to isomorphism, there is just one infinite-dimensional separable Hilbert space:

Definition 2. *Any infinite-dimensional separable Hilbert space will be referred to as* **classical Hilbert space.**

Exercises

1. If $\mathcal{H}$ and $\mathcal{K}$ are classical Hilbert spaces, there exists an isomorphism of $\mathcal{H}$ onto $\mathcal{K}$.

2. If T is an isomorphism of a classical Hilbert space $\mathcal{H}$ onto a Hilbert space $\mathcal{K}$, then $\mathcal{K}$ is also classical.

3. If $\mathcal{H}$ is isomorphic with $\mathcal{K}$, then $\mathcal{K}$ is isomorphic with $\mathcal{H}$.

Chapter III

Closed Linear Subspaces

§1. SOME NOTATIONS FROM SET THEORY

A *set* $\mathfrak{X}$ is composed of objects $x, y, \cdots$ called the *elements* or *members* of the set $\mathfrak{X}$. As in the first two chapters, I assume that these terms are meaningful to the reader, without further elaboration.

The statement "x is a member of the set $\mathfrak{X}$" is symbolized "$x \in \mathfrak{X}$," which may be read "x belongs to $\mathfrak{X}$." For example: if $\mathfrak{V}$ is a vector space, the symbol $x \in \mathfrak{V}$ means that x is a vector in the space $\mathfrak{V}$; if $\mathfrak{X}$ is a metric space, $x \in \mathfrak{X}$ means that x is a point of the space $\mathfrak{X}$; $\lambda \in \mathfrak{C}$ means that λ is a complex number.

If $x \in \mathfrak{X}$ and $y \in \mathfrak{X}$, the statement "$x = y$" means that x and y are symbols representing the *same* element of $\mathfrak{X}$. One assumes the following properties of equality: (1) $x = x$ for every $x \in \mathfrak{X}$; (2) if $x = y$, then $y = x$; and (3) if $x = y$ and $y = z$, then $x = z$. These are known as the *reflexive, symmetric,* and *transitive* properties of equality.

In addition, the following set-theoretic concepts and notations will be employed in the sequel:

1. Let $\mathfrak{X}$ and $\mathfrak{Y}$ be sets. If every element of $\mathfrak{X}$ is also an element of $\mathfrak{Y}$ (that is, if $x \in \mathfrak{X}$ implies $x \in \mathfrak{Y}$), one says that $\mathfrak{X}$ is a *subset* of $\mathfrak{Y}$. Other terms for this: $\mathfrak{X}$ is *contained* in $\mathfrak{Y}$, $\mathfrak{X}$ is *included* in $\mathfrak{Y}$, $\mathfrak{Y}$ *includes*

$\mathcal{X}$, $\mathcal{Y}$ *contains* $\mathcal{X}$, $\mathcal{Y}$ is a *superset* of $\mathcal{X}$. Notations: $\mathcal{X} \subset \mathcal{Y}$, or $\mathcal{Y} \supset \mathcal{X}$. For example, if $\mathcal{R}$ is the set of all real numbers, and $\mathcal{S}$ is the set of real numbers in the interval $[a,b]$, then $\mathcal{S} \subset \mathcal{R}$.

2. If both $\mathcal{X} \subset \mathcal{Y}$ and $\mathcal{Y} \subset \mathcal{X}$, the sets $\mathcal{X}$ and $\mathcal{Y}$ are said to be *equal*. This means: $x \in \mathcal{X}$ if and only if $x \in \mathcal{Y}$. Notation: $\mathcal{X} = \mathcal{Y}$. For example, if $\mathcal{X}$ is the set of all real numbers ≥ 0, and $\mathcal{Y}$ is the set of all real numbers which can be expressed as the square of a real number, then $\mathcal{X} = \mathcal{Y}$.

3. Let $\mathcal{X}$ be a set, and suppose that for each $x \in \mathcal{X}$ there is given a statement involving x (denoted, say, by $s(x)$), which may or may not be true. The symbol

$$\{x \in \mathcal{X}: s(x)\}$$

stands for the set of all $x \in \mathcal{X}$ for which $s(x)$ is true. For example, if $\mathcal{R}$ is the set of all real numbers, and $\mathcal{S}$ is the interval $[a,b]$, then

$$\mathcal{S} = \{x \in \mathcal{R}: a \leq x \leq b\}.$$

4. A set $\mathcal{X}$ is *empty* if it has no members (that is, the relation $x \in \mathcal{X}$ does not hold for any x). Otherwise, $\mathcal{X}$ is said to be *non-empty*. The symbol $\varnothing$ denotes an empty set. For example,

$$\{x \in \mathcal{R}: x > 2 \text{ and } x \leq 1\} = \varnothing,$$

$$\{x \in \mathcal{R}: x^2 < 0\} = \varnothing.$$

5. The set whose only element is x is denoted $\{x\}$. The set whose only elements are $x_1, \cdots, x_n$ (not necessarily distinct) is denoted $\{x_1, \cdots, x_n\}$. Thus, $x \in \{x_1, \cdots, x_n\}$ if and only if $x = x_k$ for some k.

Exercises

1. If $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are sets,
   (i) $\mathcal{X} \subset \mathcal{X}$;
   (ii) if $\mathcal{X} \subset \mathcal{Y}$ and $\mathcal{Y} \subset \mathcal{Z}$, then $\mathcal{X} \subset \mathcal{Z}$;
   (iii) $\mathcal{X} = \mathcal{Y}$ if and only if both $\mathcal{X} \subset \mathcal{Y}$ and $\mathcal{Y} \subset \mathcal{X}$.
2. If $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are sets,
   (i) $\mathcal{X} = \mathcal{X}$;
   (ii) if $\mathcal{X} = \mathcal{Y}$, then $\mathcal{Y} = \mathcal{X}$;
   (iii) if $\mathcal{X} = \mathcal{Y}$ and $\mathcal{Y} = \mathcal{Z}$, then $\mathcal{X} = \mathcal{Z}$.

3. If $\mathcal{X}$ and $\mathcal{Y}$ are both empty sets, then $\mathcal{X} = \mathcal{Y}$. Thus one can speak of *the* empty set.

§ 2. ANNIHILATORS

Definition 1. *Let $\mathcal{S}$ be a subset of a pre-Hilbert space $\mathcal{P}$. A vector $x \in \mathcal{P}$ is said to be* **orthogonal** *to $\mathcal{S}$ in case $x \perp s$ for all $s \in \mathcal{S}$; notation: $x \perp \mathcal{S}$. The set of all such vectors x is called the* **annihilator** *of $\mathcal{S}$, and is denoted $\mathcal{S}^{\perp}$. Thus:*

$$\mathcal{S}^{\perp} = \{x \in \mathcal{P}: (x|s) = 0 \quad \textit{for all} \quad s \in \mathcal{S}\}.$$

Examples

1. Recall that $\{0\}$ is the set whose only element is the vector 0. Then, $\mathcal{P}^{\perp} = \{0\}$. In fact, a subset $\mathcal{S}$ is total if and only if $\mathcal{S}^{\perp} = \{0\}$.

2. Clearly $\{0\}^{\perp} = \mathcal{P}$.

3. The convention is that $\varnothing^{\perp} = \mathcal{P}$, where $\varnothing$ is the empty subset of $\mathcal{P}$. The rationalization for this is as follows: given any vector $x \in \mathcal{P}$, one has $(x|s) = 0$ whenever $s \in \varnothing$ (which is never).

Theorem 1. *If $\mathcal{S}$ is any subset of a pre-Hilbert space $\mathcal{P}$, then $\mathcal{S}^{\perp}$ is a linear subspace of $\mathcal{P}$. Moreover, if $x_n \in \mathcal{S}^{\perp}$ for all n, $x \in \mathcal{P}$, and $x_n \to x$, then $x \in \mathcal{S}^{\perp}$.* is adherent to $S^{\perp}$

Proof.
Clearly $0 \in \mathcal{S}^{\perp}$, and $\mathcal{S}^{\perp}$ is a linear subspace by *Theorem II.6.1*. Suppose $x_n \perp \mathcal{S}$ for all n, and $x_n \to x$. For any $s \in \mathcal{S}$, $(x|s) = \lim (x_n|s)$ by *Theorem II.5.1*; since $(x_n|s) = 0$ for all n, $(x|s) = 0$. ∎

Definition 2. *Let $\mathcal{S}$ be a subset of a metric space $\mathcal{X}$. A point $x \in \mathcal{X}$ is said to be* **adherent** *to $\mathcal{S}$ in case there exists a sequence $s_n \in \mathcal{S}$ such that $s_n \to x$.*

In particular, if $\mathcal{S}$ is a subset of a pre-Hilbert space $\mathcal{P}$, a vector $x \in \mathcal{P}$ is adherent to $\mathcal{S}$ if and only if there is a sequence $s_n \in \mathcal{S}$ such that $\| s_n - x \| \to 0$ as $n \to \infty$.

Definition 3. *Let $\mathcal{X}$ be a metric space, $\mathcal{S} \subset \mathcal{X}$. $\mathcal{S}$ is said to be a* **closed** *subset of $\mathcal{X}$ if it contains every point adherent to it. That is, the relations $s_n \in \mathcal{S}$, $x \in \mathcal{X}$, $s_n \to x$, imply $x \in \mathcal{S}$.*